Lecture notes for Math 272, Winter 2023

Lenya Ryzhik[1]

March 8, 2023

[1]Department of Mathematics, Stanford University, Stanford, CA 94305, USA. Email: ryzhik@stanford.edu

These notes will be continuously updated during the course.

The plan for this class is to cover the following topics:

I. Basic theory of Hamilton-Jacobi equations.
II. An introduction to mean-field games.
III. Some applications to macroeconomics.

Contents

Chapter 1

Inviscid Hamilton-Jacobi equations

1.1 Introduction

We will consider in this chapter the Hamilton-Jacobi equations

$$u_t + H(x, \nabla u) = 0 \tag{1.1.1}$$

on the unit torus $\mathbb{T}^n \subset \mathbb{R}^n$, or, sometimes, in all of $\mathbb{R}^n$. As we will see, a physically reasonable class of solutions to (1.1.1) behave very much like the solutions to a regularized problem

$$u_t^\varepsilon + H(x, \nabla u^\varepsilon) = \varepsilon \Delta u^\varepsilon, \tag{1.1.2}$$

with a small diffusivity $\varepsilon > 0$. Most of the techniques for the analysis of the solutions to such nonlinear diffusive equations rely on the positivity of the diffusion coefficient and deteriorate badly when the diffusion coefficient is small. However, we will see that some of the bounds may survive even as the diffusion term vanishes, because they are helped by the nonlinear Hamiltonian $H(x, \nabla u)$. Obviously, not every nonlinearity is beneficial: for example, solutions to the linear advection equation

$$u_t + b(x) \cdot \nabla u(x) = 0, \tag{1.1.3}$$

are typically no more regular than the initial condition $u_0(x) = u(0, x)$, no matter how smooth the drift $b(x)$ is. Therefore, we will have to restrict ourselves to a class of Hamiltonians $H(x, p)$ that do help to regularize the problem. This nonlinear regularization effect is one of the main points of this chapter.

1.2 An informal derivation of the Hamilton-Jacobi equations

We begin by providing an informal derivation of the Hamilton-Jacobi equations. The material of this section will reappear later in the form of the Lax-Oleinik formula for the solutions to the Hamilton-Jacobi equations.

We start with a random walk on a lattice of size h in $\mathbb{R}^n$, and a time step τ. The walker evolves as follows. If the walker is located at a position $X(t) \in h\mathbb{Z}^n$ at a time $t = m\tau$, $m \in \mathbb{N}$, then at a time $t + \tau$ it finds itself at a position

$$X(t+\tau) = X(t) + v(t)\tau + h\xi(t). \tag{1.2.1}$$

Here, $\xi(t) \in \mathbb{R}^n$ is an $\mathbb{R}^n$-valued random variable such that each of the coordinates $\xi_k(t)$, with $k = 1, \dots, n$, are independent and take the values ± 1 with probabilities equal to $1/2$, so that

$$\mathbb{E}(\xi_k(t)) = 0, \quad \mathbb{E}(\xi_k(t)\xi_m(t')) = \delta_{km}\delta_{t,t'}, \tag{1.2.2}$$

for all $1 \leq k, m \leq n$ and all t, t'. The velocity $v(t)$ is known as a control, that the walker can choose from a set $\mathcal{A}$ of admissible velocities. The choice of the velocity v on the time interval $[t, t+\tau]$ comes with a cost $L(v)\tau$, where $L(v)$ is a prescribed cost function. At the terminal time $T = N\tau$ the walker finds itself at a position $X(T)$ and pays the terminal cost $f(X(T))$, where $f(x)$ is also a given function. The total cost of the trajectory that starts at a time $t = m\tau$ at a position x and continues until the time $T = N\tau$ is

$$w(t, x; V) = \sum_{k=m}^{N} L(v(k\tau))\tau + f(X(N\tau)). \tag{1.2.3}$$

Note that the total cost involves both the running cost and the terminal cost. We have denoted here by $V = (v(t), v(t+\tau), \dots, v((N-1)\tau))$ the whole sequence of the controls (velocities) chosen by the walker between the times $t = m\tau$ and $T = N\tau$.

The quantity of interest is the least possible average cost, optimized over all choices of the velocities:

$$u(t, x) = \inf_{V \in \mathcal{A}_{t,T}} \mathbb{E}\big[w(t, x; V)\big] = \inf_{V \in \mathcal{A}_{t,T}} \mathbb{E}\Big(\sum_{k=m}^{N} L(v(k\tau))\tau + f(X(N\tau))\Big). \tag{1.2.4}$$

Here, the expectation $\mathbb{E}$ is taken with respect to the random variables $\xi(s)$, for all $s = k\tau$ with $m \leq k < N$ that describe the random contribution at each of the time steps between t and T. The set $\mathcal{A}_{t,T}$ is the set of all possible controls chosen between the times $t = m\tau$ and $T = N\tau$. The velocities v are viewed as not random, as they can be chosen by the walker. The function $u(t, x)$ is known as the value function and is the basic object of study in the control theory.

As the velocities $v(s)$ are chosen separately by the walker at each time s between t and T, and the random variables $\xi(s)$ and $\xi(s')$ are independent for $s \neq s'$, the function $u(t, x)$ satisfies the following relation:

$$u(t, x) = \inf_{v \in \mathcal{A}} \mathbb{E}\big[L(v)\tau + u(t+\tau, x + v\tau + h\xi(t))\big]. \tag{1.2.5}$$

This is the simplest version of a dynamic programming principle, a fundamental notion of the control theory. Here, v is the velocity chosen at the initial time t and the expectation is taken solely with respect to the random variable $\xi(t)$.

A version of the dynamic programming principle, such as (1.2.5), is a very common starting point for the derivation of the Hamilton-Jacobi and other related types of equations that come from the optimal control theory. To illustrate this idea, let us assume that $u(t,x)$ is a sufficiently smooth function and that the time step τ and the spatial step h are sufficiently small. Expanding the right side of (1.2.5) in $h \ll 1$ and $\tau \ll 1$ gives

$$\begin{aligned} u(t,x) &= \inf_{v\in\mathcal{A}} \mathbb{E}\big[L(v)\tau + u(t+\tau, x+v\tau+h\xi(t))\big] = u(t,x) + \tau u_t + \frac{\tau^2}{2} u_{tt}(t,x) \\ &+ \inf_{v\in\mathcal{A}} \mathbb{E}\Big[L(v)\tau + (v\tau + h\xi(t))\cdot\nabla u(t,x) + \tau(v\tau+h\xi(t))\cdot\nabla u_t(t,x) \\ &+ \frac{1}{2}\sum_{i,j=1}^{n}(v_i\tau + h\xi_i(t))(v_j\tau + h\xi_j(t))\frac{\partial^2 u(t,x)}{\partial x_i \partial x_j}\Big] + l.o.t. \end{aligned} \tag{1.2.6}$$

Note that the terms of the order $O(1)$ in the left and the right sides of (1.2.6) cancel automatically. In addition, the terms that are linear in $\xi(t)$ vanish after taking the expectation because of (1.2.2). An interesting choice of the relation between the temporal and spatial steps τ and h is

$$h^2 = 2D\tau, \tag{1.2.7}$$

with a diffusion coefficient $D > 0$ (the true diffusion coefficient is, of course, $2D$ but we make this normalization for convenience, as is common in the PDE literature). Then, after taking into account the aforementioned cancellations, the leading order terms in (1.2.6) are of the order $O(\tau) = O(h^2)$. Still, keeping in mind (1.2.2), we see that they combine to give the following equation for $u(t,x)$:

$$u_t(t,x) + \inf_{v\in\mathcal{A}}\big[L(v) + v\cdot\nabla u(t,x)\big] + D\Delta u(t,x) = 0. \tag{1.2.8}$$

Let us introduce the function

$$H(p) = \inf_{v\in\mathcal{A}}\big[L(v) + v\cdot p\big], \tag{1.2.9}$$

defined for $p \in \mathbb{R}^n$. It is usually called the Hamiltonian and is the Legendre transform of the Lagrangian $L(v)$. Then, (1.2.8) can be written as

$$u_t + H(\nabla u) + D\Delta u = 0. \tag{1.2.10}$$

This equation should be supplemented by the terminal condition $u(T,x) = f(x)$ that comes simply from the definition of the value function. Recall that $f(x)$ is the terminal cost function.

Equation (1.2.10) is backward in time. It is convenient to define the function

$$\bar{u}(t,x) = u(T-t, x),$$

which satisfies the forward in time Cauchy problem:

$$\begin{aligned} &\bar{u}_t = H(\nabla\bar{u}) + D\Delta\bar{u}, \quad t > 0. \\ &\bar{u}(0,x) = f(x), \end{aligned} \tag{1.2.11}$$

For the sake of convenience we will focus on this forward in time Cauchy problem.

This is how the viscous Hamilton-Jacobi equations can be derived informally. Their rigorous derivation starting with a continuous in space and time stochastic control problem is not very different but requires the use of the stochastic calculus and the Ito formula. The inviscid equations of the form

$$u_t = H(\nabla u), \tag{1.2.12}$$

are derived in a very similar way but the walk is taken to be purely deterministic, driven solely by the control v, with $\xi(t) \equiv 0$.

Exercise 1.2.1 Generalize the above derivation to obtain a spatially inhomogeneous Hamilton-Jacobi equation of the form

$$u_t = H(x, \nabla u) + D\Delta u. \tag{1.2.13}$$

Exercise 1.2.2 Show that the function $H(p)$ defined in (1.2.9) is concave.

This exercise explains why we will often consider below the Hamilton-Jacobi equations of the form

$$u_t + H(x, \nabla u) = D\Delta u, \tag{1.2.14}$$

with a convex Hamiltonian $H(p)$, either with $D > 0$ or $D = 0$.

1.3 The simple world of viscous Hamilton-Jacobi equations

As a warm-up, we are going to study the long time behavior of the solutions to the Cauchy problem for viscous Hamilton-Jacobi equations

$$u_t - \Delta u = H(x, \nabla u), \quad t > 0, \quad x \in \mathbb{T}^n, \tag{1.3.1}$$

with a given initial condition $u(0,x) = u_0(x)$. We now make some assumptions on the nonlinearity $H(x,p)$. First, we assume that H is smooth and 1-periodic in x. We also make the uniformly Lipschitz assumption on the function $H(x,p)$: there exists $C_L > 0$ so that

$$|H(x, p_1) - H(x, p_2)| \leq C_L |p_1 - p_2|, \text{ for all } x, p_1, p_2 \in \mathbb{R}^n. \tag{1.3.2}$$

In addition, we assume that H is growing linearly in p at infinity: there exist $\alpha > 0$ and $\beta > 0$ so that

$$0 < \alpha \leq \liminf_{|p|\to+\infty} \frac{H(x,p)}{|p|} \leq \limsup_{|p|\to+\infty} \frac{H(x,p)}{|p|} \leq \beta < +\infty, \text{ uniformly in } x \in \mathbb{T}^n. \tag{1.3.3}$$

One consequence of (1.3.3) is that $H(x,p)$ is uniformly bounded from below. Note also that if $u(t,x)$ solves (1.3.1) then $u(t,x) + Kt$ solves (1.3.1) with the Hamiltonian $H(x,p)$ replaced by $H(x,p) + K$. Therefore, we may assume without loss of generality that there exist $C_{1,2} > 0$ so that

$$C_1(1 + |p|) \leq H(x,p) \leq C_2(1 + |p|), \text{ for all } x \in \mathbb{T}^n \text{ and } p \in \mathbb{R}^n, \tag{1.3.4}$$

so that, in particular,

$$H(x,p) > C_1 \text{ for all } x \in \mathbb{T}^n \text{ and } p \in \mathbb{R}^n. \tag{1.3.5}$$

Relatively standard theory for nonlinear diffusion equations implies that these assumptions ensure the existence of a unique smooth 1-periodic solution $u(t,x)$ to (1.3.1) supplemented by a continuous, 1-periodic initial condition $u_0(x)$. In order to discuss its long time behavior, we need to introduce a special class of solutions of (1.3.1).

Theorem 1.3.1 *Under the above assumptions, there exists a unique $c \in \mathbb{R}$ so that (1.3.1) has solutions (that we will call the wave solutions) of the form*

$$w(t,x) = ct + \phi(x), \tag{1.3.6}$$

with a 1-periodic function $\phi(x)$. The profile $\phi(x)$ is unique up to an additive constant: if $w_1(t,x)$ and $w_2(t,x)$ are two such solutions then there exists $k \in \mathbb{R}$ so that $\phi_1(x) - \phi_2(x) \equiv k$ for all $x \in \mathbb{T}^n$.

The constant c is often referred to as the speed of the plane wave. The reason is that the solutions to the Hamilton-Jacobi equations, apart from the optimal control theory context that we have discussed above, also often describe the height of an interface, so that c may be thought of as the speed at which the height of the interface is moving up, and $\phi(x)$ as the fixed profile of that interface as it moves up at a constant speed.

Exercise 1.3.2 Give an interpretation to the profile $\phi(x)$ and the speed c in the context of the optimal control formulation for the solutions to the Hamilton-Jacobi equations.

Exercise 1.3.3 Consider the following discrete growing interface model, defined on the lattice $h\mathbb{Z}$, with a time step τ. The interface height $u(t,x)$ at the time t and the position x evolves as follows:

$$\begin{aligned} u(t+\tau,x) &= \frac{1}{2}\big[u(t,x-h) + u(t,x+h)\big] \\ &\quad + \frac{1}{2}\Big[F(u(t,x+h) - u(t,x)) + F(u(t,x) - u(t,x-h))\Big] + \delta V(t,x), \end{aligned} \tag{1.3.7}$$

with a given function $F(p)$, and a prescribed source $V(t,x)$. The terms in the right side of (1.3.7) can be interpreted as follows: (1) the first term has an equilibrating effect, leveling the interface out, (2) the second term says that the rate of the interface growth depends on its slope – things falling from above can stick to the interface, and (3) the last term is an outside source of the interface growth (things falling from above). Find a scaling limit that relates τ, h and δ so that in the limit you get a Hamilton-Jacobi equation of the form

$$u_t = \Delta u + H(x, \nabla u) + V(t,x). \tag{1.3.8}$$

The large time behavior of the solution to (1.3.1) is summarized in the next theorem.

Theorem 1.3.4 *Let $u(t,x)$ be the solution to the Cauchy problem for (1.3.1) with a continuous 1-periodic initial condition $u(0,x) = u_0(x)$. There is a wave solution $w(t,x)$ to (1.3.1) of the form (1.3.6), a constant $\omega > 0$ that does not depend on u_0 and $C_0 > 0$ that depends on u_0 such that*

$$|u(t,x) - w(t,x)| \le C_0 e^{-\omega t}, \tag{1.3.9}$$

for all $t \ge 0$ and $x \in \mathbb{T}^n$.

We will first prove the existence part of Theorem 1.3.1, and that will occupy most of the rest of this section, while its uniqueness part and the convergence claim of Theorem 1.3.4 will be proved together rather quickly in the end.

1.3.1 The wave solutions

Outline of the existence proof

We first present an outline of the existence proof, before going into the details of the argument. Plugging the ansatz (1.3.6) into (1.3.1) and integrating over $\mathbb{T}^n$ gives a relation

$$c = \int_{\mathbb{T}^n} H(x, \nabla\phi(x))dx. \tag{1.3.10}$$

The equation for ϕ can, therefore, be written as

$$-\Delta\phi = H(x, \nabla\phi(x)) - \int_{\mathbb{T}^n} H(z, \nabla\phi(z))dz, \tag{1.3.11}$$

and this will be the starting point of our analysis.

We are going to use a continuation method. As this strategy is typical for the existence proofs for many nonlinear PDEs, it is worth sketching out the general plan. Instead of just looking at (1.3.11) with a given Hamiltonian $H(x,p)$, we consider a family of equations

$$-\Delta\phi_\sigma = H_\sigma(x, \nabla\phi_\sigma) - \int_{\mathbb{T}^n} H_\sigma(z, \nabla\phi_\sigma)dz, \tag{1.3.12}$$

with the Hamiltonians

$$H_\sigma(x,p) = (1-\sigma)H_0(x,p) + \sigma H(x,p), \tag{1.3.13}$$

parametrized by $\sigma \in [0,1]$. At $\sigma = 0$, we start with a particular choice of $H_0(x,p)$ for which we know that (1.3.12) has a solution. Here, we take

$$H_0(x,p) = \sqrt{1+|p|^2}.$$

Note that $\phi_0(x) \equiv 0$ is an explicit solution to (1.3.12) with $\sigma = 0$. At $\sigma = 1$, we end with

$$H_1(x,p) = H(x,p). \tag{1.3.14}$$

We are interested in the existence of a solution to (1.3.13) when $\sigma = 1$. However, we will actually show that (1.3.12) has a solution for all $\sigma \in [0,1]$ and not just for $\sigma = 0$ by showing that the set Σ of σ for which (1.3.12) has a solution is both open and closed in $[0,1]$. This is the continuation method – you start with a problem at $\sigma = 0$ for which you know that a solution exists and then extend the existence to the case $\sigma = 1$, which is the only one you are really interested in, using a continuity argument.

Showing that Σ is closed requires a priori bounds on the solution ϕ_σ of (1.3.12) that would both be uniform in $\sigma \in [0,1]$ and ensure the compactness of the sequence ϕ_{σ_n} of solutions to (1.3.12) as $\sigma_n \to \sigma$ in a suitable function space. The estimates should be strong enough to ensure that the limit ϕ_σ is a solution to (1.3.12).

In order to show that Σ is open, one relies on the implicit function theorem. Let us assume that (1.3.12) has a solution $\phi_\sigma(x)$ for some $\sigma \in [0,1]$. In order to show that (1.3.12) has a solution for $\sigma + \varepsilon$, with a sufficiently small ε, we are led to consider the linearized problem

$$-\Delta h - \frac{\partial H_\sigma(x, \nabla\phi_\sigma)}{\partial p_j}\frac{\partial h}{\partial x_j} + \int_{\mathbb{T}^n} \frac{\partial H_\sigma(z, \nabla\phi_\sigma)}{\partial p_j}\frac{\partial h(z)}{\partial z_j}dz = f, \tag{1.3.15}$$

with

$$f(x) = H(x, \nabla\phi_\sigma) - H_0(x, \nabla\phi_\sigma) - \int_{\mathbb{T}^n} H(z, \nabla\phi_\sigma(z))dz + \int_{\mathbb{T}^n} H_0(z, \nabla\phi_\sigma(z))dz. \quad (1.3.16)$$

The implicit function theorem guarantees existence of the solution $\phi_{\sigma+\varepsilon}$, provided that the linearized operator in the left side of (1.3.15) is invertible, with the norm of the inverse a priori bounded in σ. This will show that the set Σ of $\sigma \in [0,1]$ for which the solution to (1.3.12) exists is open.

The bounds on the operator that maps $f \to h$ in (1.3.15) also require the a priori bounds on ϕ_σ. Thus, both proving that Σ is open and that it is closed require us to prove the a priori uniform bounds on ϕ_σ. Therefore, our first step will be to assume that a solution $\phi_\sigma(x)$ to (1.3.12) exists and obtain a priori bounds on ϕ_σ. Note that if ϕ_σ is a solution to (1.3.12), then $k + \phi_\sigma$ is also a solution for any $k \in \mathbb{R}$. Thus, it is more natural to obtain a priori bounds on $\nabla\phi_\sigma$ than on ϕ_σ itself, and then normalize the solution so that $\phi_\sigma(0) = 0$ to ensure that ϕ_σ is bounded.

It is important to observe that the Hamiltonians $H_\sigma(x,p)$ obey the Lipschitz bound (1.3.2), with a Lipschitz constant C_L that does not depend on $\sigma \in [0,1]$, and estimate (1.3.4) also holds for H_σ with the same $C_{1,2} > 0$ for all $\sigma \in [0,1]$. The key bound to prove will be to show that there exists a constant $K > 0$ that depends only on the Lipschitz constant of H in (1.3.2) and the two constants in the linear growth estimate (1.3.4) such that any solution to (1.3.12) satisfies

$$\|\nabla\phi_\sigma\|_{L^\infty(\mathbb{T}^n)} \leq K. \quad (1.3.17)$$

We stress that this bound will be obtained not just for one Hamiltonian but for all Hamiltonians with the same Lipschitz constant C_L in (1.3.2) that also satisfy (1.3.4) with the same constants $C_{1,2} > 0$. The estimate (1.3.17) will turn out to be sufficient to apply the argument we have outlined above.

An a priori L^1-bound on the gradient

Before establishing the L^∞-bound (1.3.17), let us first prove that there exists a constant $C > 0$ that only depends on C_L in (1.3.2) and $C_{1,2}$ in (1.3.4) such that any solution $\phi_\sigma(x)$ of (1.3.12) satisfies

$$\int_{\mathbb{T}^n} H_\sigma(x, \nabla\phi_\sigma)dx \leq C. \quad (1.3.18)$$

Because of the lower bound in (1.3.4), this is equivalent to an a priori L^1-bound on $|\nabla\phi_\sigma|$:

$$\int_{\mathbb{T}^n} |\nabla\phi_\sigma(x)|dx \leq C, \quad (1.3.19)$$

with a possibly different $C > 0$ that still depends only on C_L and $C_{1,2}$. Note that here already the coercivity (growth at infinity) of the Hamiltonian plays an important role in the bound on the gradient. To prove (1.3.18), we recall the following result.

Proposition 1.3.5 *Let $b(x)$ be a smooth vector field over $\mathbb{T}^n$. The linear equation*

$$-\Delta e + \nabla\cdot(eb) = 0, \qquad x \in \mathbb{T}^n, \quad (1.3.20)$$

has a unique solution $e_1^*(x)$ normalized so that

$$\|e_1^*\|_{L^\infty} = 1, \tag{1.3.21}$$

and such that $e_1^* > 0$ on $\mathbb{T}^n$. Moreover, for all $\alpha \in (0,1)$, the function e_1^* is α-Hölder continuous, with the α-Hölder norm bounded by a universal constant depending only on $\|b\|_{L^\infty(\mathbb{T}^n)}$.

Exercise 1.3.6 Prove this proposition. As a first step, consider the Laplace equation

$$-\Delta u = \nabla \cdot g(x), \quad x \in \mathbb{T}^n, \tag{1.3.22}$$

with a smooth function $g(x)$. Show that there exists a constant $C > 0$ that does not depend on the function $g(x)$ such that $[u]_{C^\alpha(\mathbb{T}^n)} \le C\|g\|_{L^\infty(\mathbb{T}^n)}$. Here, we use the notation $[u]_{C^\alpha(\mathbb{T}^n)}$ for the α-Hölder constant of the function $u(x)$.

Let us first see why Proposition 1.3.5 implies (1.3.18). An immediate consequence of the normalization (1.3.21) and the claim about the Hölder norm of e_1^*, together with the positivity of e_1^* is that

$$\int_{\mathbb{T}^n} e_1^*(x)dx \ge K_1 > 0, \tag{1.3.23}$$

with a constant $K_1 > 0$ that depends only on $\|b\|_{L^\infty}$. Now, given a solution $\phi_\sigma(x)$ of (1.3.12), set

$$b_j(x) = \int_0^1 \partial_{p_j} H_\sigma(x, r\nabla\phi_\sigma(x))dr, \tag{1.3.24}$$

so that

$$b(x) \cdot \nabla\phi_\sigma(x) = \sum_{j=1}^n b_j(x)\frac{\partial \phi_\sigma}{\partial x_j} = H_\sigma(x, \nabla\phi_\sigma) - H_\sigma(x, 0), \tag{1.3.25}$$

and (1.3.12) can be re-stated as

$$-\Delta\phi_\sigma - b(x)\cdot\nabla\phi_\sigma = H_\sigma(x,0) - \int_{\mathbb{T}^n} H_\sigma(z, \nabla\phi_\sigma)dz. \tag{1.3.26}$$

Note that while $b(x)$ does depend on $\nabla\phi_\sigma$, the L^∞-norm of $b(x)$ depends only on the Lipschitz constant C_L of the function $H_\sigma(x,p)$ in the p-variable. Let now e_1^* be the solution to (1.3.20) given by Proposition 1.3.5, with the above $b(x)$. Multiplying (1.3.26) by e_1^* and integrating over $\mathbb{T}^n$ yields

$$0 = \int_{\mathbb{T}^n} e_1^*(x)H_\sigma(x,0)dx - \Big(\int_{\mathbb{T}^n} e_1^*(x)dx\Big)\Big(\int_{\mathbb{T}^n} H_\sigma(z,\nabla\phi_\sigma)dz\Big), \tag{1.3.27}$$

hence

$$\int_{\mathbb{T}^n} H_\sigma(x,\nabla\phi_\sigma)dx = \Big(\int_{\mathbb{T}^n} e_1^*(x)dx\Big)^{-1}\int_{\mathbb{T}^n} e_1^*(x)H_\sigma(x,0)dx, \tag{1.3.28}$$

and (1.3.19) follows from (1.3.23) and (1.3.4). As the constant K_1 in (1.3.23) depends only on the L^∞-norm of $b(x)$ that, in turn, depends only on C_L, the constant C in the right side of (1.3.18), indeed, depends only on C_L and $C_{1,2}$.

An a priori L^∞ bound on the gradient

So far, we have obtained an a priori L^1-bound (1.3.19) for the gradient of any solution ϕ_σ to (1.3.12). Now, we improve this estimate to an L^∞ bound.

Proposition 1.3.7 *There is a constant $C > 0$ that depends only on the constants C_L and $C_{1,2}$, such that any solution ϕ_σ to*

$$-\Delta\phi_\sigma = H_\sigma(x, \nabla\phi_\sigma) - \int_{\mathbb{T}^n} H_\sigma(z, \nabla\phi_\sigma)dz, \tag{1.3.29}$$

satisfies

$$\|\nabla\phi_\sigma\|_{L^\infty(\mathbb{T}^n)} \leq C. \tag{1.3.30}$$

As a consequence, if ϕ_σ is normalized such that $\phi_\sigma(0) = 0$, then we also have $\|\phi_\sigma\|_{L^\infty(\mathbb{T}^n)} \leq C$.

Proof. Let ϕ_σ be a solution to (1.3.29) such that $\phi_\sigma(0) = 0$. The only estimate we have so far is the L^1-bound (1.3.19) for $\nabla\phi_\sigma$ – the idea is to estimate the L^∞-norm $\|\nabla\phi_\sigma\|_{L^\infty(\mathbb{T})}$ solely from the L^1-norm of $\nabla\phi_\sigma$ and the equation.

Let $\Gamma(x)$ be a nonnegative smooth function equal to 1 in the cube $[-2, 2]^n$ and to zero outside of the cube $(-3, 3)^n$, and set $\psi(x) = \Gamma(x)\phi_\sigma(x)$. The function $\psi(x)$ satisfies an equation of the form

$$-\Delta\psi = -2\nabla\Gamma \cdot \nabla\phi_\sigma - \phi_\sigma\Delta\Gamma + F(x), \quad x \in \mathbb{R}^n, \tag{1.3.31}$$

with

$$F(x) = \Gamma(x)\Big[H_\sigma(x, \nabla\phi_\sigma(x)) - \int_{\mathbb{T}^n} H_\sigma(z, \nabla\phi_\sigma(z))dz\Big]. \tag{1.3.32}$$

The only a priori information we have about $F(x)$ and the term $\nabla\Gamma\cdot\nabla\phi_\sigma(x)$ so far is that they are supported inside $[-3, 3]^n$ and are uniformly bounded in $L^1(\mathbb{R}^n)$ via (1.3.18) and (1.3.19). Here, we again use the assumption (1.3.4) that the Hamiltonian $H(x, p)$ is uniformly positive. It helps to combine these two terms:

$$G(x) = F(x) - 2\nabla\Gamma(x) \cdot \nabla\phi_\sigma(x), \tag{1.3.33}$$

with $G(x)$ supported inside $[-3, 3]^n$, and

$$\int_{\mathbb{R}^n} |G(x)|dx \leq C, \tag{1.3.34}$$

with a constant $C > 0$ that depends only on C_L and $C_{1,2}$, due to (1.3.18) and (1.3.19). We also know that

$$|G(x)| \leq C(1 + |\nabla\phi_\sigma(x)|, \tag{1.3.35}$$

because of (1.3.4).

Next, we use the fundamental solution $E(x)$ to the Laplace equation in $\mathbb{R}^n$ to write

$$\psi(x) = \int_{\mathbb{R}^n} E(x - y)[G(y) - \phi_\sigma(y)\Delta\Gamma(y)]dy. \tag{1.3.36}$$

Differentiating (1.3.36) in x gives

$$\nabla\psi(x) = \int_{\mathbb{R}^n} \nabla E(x - y)\big[G(y) - \phi_\sigma(y)\Delta\Gamma(y)\big]dy. \tag{1.3.37}$$

Exercise 1.3.8 Note that the function $E(x-y)$ has a singularity at $y=x$. Show that nevertheless one can differentiate in (1.3.36) under the integral sign to obtain (1.3.37).

The function $\nabla E(x-y)$ has an integrable singularity at $y=x$, of the order $|x-y|^{-n+1}$, and is bounded everywhere else. Thus, for all $\varepsilon>0$ we have, with the help of (1.3.34) and (1.3.35):

$$\begin{aligned}
\Big|\int_{\mathbb{R}^n} G(y)\nabla E(x-y)dy\Big| &\le \Big|\int_{|x-y|\le\varepsilon} G(y)\nabla E(x-y)dy\Big|+\Big|\int_{|x-y|\ge\varepsilon} G(y)\nabla E(x-y)dy\Big|\\
&\le C(1+\|\nabla\phi_\sigma\|_{L^\infty})\int_{|x-y|\le\varepsilon}\frac{dy}{|x-y|^{n-1}}+C\varepsilon^{-n+1}\int_{|x-y|\ge\varepsilon}|G(y)|dy\\
&\le C\varepsilon(1+\|\nabla\phi_\sigma\|_{L^\infty})+C\varepsilon^{1-n}. \qquad (1.3.38)
\end{aligned}$$

The integral in (1.3.37) also contains a factor of ϕ_σ, whereas our bounds so far deal with $\nabla\phi_\sigma$. However, we have assumed without loss of generality that $\phi_\sigma(0)=0$, hence for any $\delta>0$ we may write

$$\phi_\sigma(y)=\int_0^1 y\cdot\nabla\phi_\sigma(sy)ds=\int_0^\delta y\cdot\nabla\phi_\sigma(sy)ds+\int_\delta^1 y\cdot\nabla\phi_\sigma(sy)ds,$$

so that both, as $|y|\le 1$, we have

$$|\phi_\sigma(y)|\le\|\nabla\phi_\sigma\|_{L^\infty}, \qquad (1.3.39)$$

and

$$\begin{aligned}
\int_{\mathbb{T}^n}|\phi_\sigma(y)|dy &\le C\delta\|\nabla\phi_\sigma\|_{L^\infty}+\int_\delta^1\int_{\mathbb{T}^n}|y||\nabla\phi_\sigma(sy)|dyds\\
&\le C\delta\|\nabla\phi_\sigma\|_{L^\infty}+C\int_\delta^1\int_{s\mathbb{T}^n}|y||\nabla\phi_\sigma(y)|dy\frac{ds}{s^{n+1}}\le C\delta\|\nabla\phi_\sigma\|_{L^\infty}+C\int_\delta^1\frac{ds}{s^{1+n}}\\
&\le C\delta\|\nabla\phi_\sigma\|_{L^\infty}+C\delta^{-n}. \qquad (1.3.40)
\end{aligned}$$

We used above the a priori bound (1.3.19) on $\|\nabla\phi\|_{L^1(\mathbb{T}^n)}$. Combining (1.3.39) and (1.3.40), we obtain, as in (1.3.38):

$$\begin{aligned}
&\Big|\int_{\mathbb{R}^n}\phi_\sigma(y)\Delta\Gamma(y)\nabla E(x-y)dy\Big|\le\int_{|x-y|\le\varepsilon}|\phi_\sigma(y)||\Delta\Gamma(y)|\nabla E(x-y)|dy\\
&+\int_{|x-y|\ge\varepsilon}|\phi_\sigma(y)||\Delta\Gamma(y)||\nabla E(x-y)|dy\le C\varepsilon\|\phi_\sigma\|_{L^\infty}+C\varepsilon^{1-n}\int_{\mathbb{T}^n}|\phi_\sigma(y)|dy\\
&\le C\varepsilon\|\nabla\phi_\sigma\|_{L^\infty}+C\varepsilon^{1-n}\delta\|\nabla\phi_\sigma\|_{L^\infty}+C\varepsilon^{1-n}\delta^{-n}. \qquad (1.3.41)
\end{aligned}$$

Together, (1.3.38) and (1.3.41) tell us that

$$\|\nabla\psi\|_{L^\infty}\le C\varepsilon(1+\|\nabla\phi_\sigma\|_{L^\infty})+C\varepsilon^{1-n}+C\varepsilon\|\nabla\phi_\sigma\|_{L^\infty}+C\varepsilon^{1-n}\delta\|\nabla\phi_\sigma\|_{L^\infty}+C\varepsilon^{1-n}\delta^{-n}. \quad (1.3.42)$$

Next, observe that, because $\Gamma\equiv 1$ in $[-2,2]^n$ and ϕ_σ is 1-periodic, we have

$$\|\nabla\phi_\sigma\|_{L^\infty(\mathbb{T}^n)}=\|\nabla(\Gamma\phi_\sigma)\|_{L^\infty([-1,1]^n)}\le\|\nabla(\Gamma\phi_\sigma)\|_{L^\infty([-3,3]^n)}=\|\nabla\psi\|_{L^\infty}. \qquad (1.3.43)$$

Thus, if we take $\delta = \varepsilon^n$ in (1.3.42), we would obtain

$$\|\nabla\phi_\sigma\|_{L^\infty} \leq C\varepsilon\|\nabla\phi_\sigma\|_{L^\infty} + C_\varepsilon, \tag{1.3.44}$$

with a universal constant $C > 0$ and C_ε that does depend on ε. Now, the proof of (1.3.30) is concluded by taking $\varepsilon > 0$ small enough. □

Going back to equation (1.3.11) for ϕ:

$$-\Delta\phi = H(x, \nabla\phi) - \int_{\mathbb{T}^n} H(x, \nabla\phi)dx, \tag{1.3.45}$$

the reader should do the following exercise.

Exercise 1.3.9 Use the L^∞-bound on $\nabla\phi$ in Proposition 1.3.7 to deduce from (1.3.45) that, under the assumption that $H(x,p)$ is smooth (infinitely differentiable) in both variables x and p, the function $\phi(x)$ is, actually, infinitely differentiable, with all its derivatives of order n bounded by a priori constants C_n that do not depend on ϕ.

The linearized problem

We need one more ingredient to finish the proof of the existence part of Theorem 1.3.1: to set-up an application of the implicit function theorem. Let ϕ_σ be a solution to (1.3.12) and let us consider the linearized problem, with an unknown h:

$$-\Delta h - \partial_{p_j} H_\sigma(x, \nabla\phi_\sigma)\partial_{x_j} h + \int_{\mathbb{T}^n} \partial_{p_j} H_\sigma(y, \nabla\phi_\sigma)\partial_{x_j} h(y)dy = f \quad x \in \mathbb{T}^n. \tag{1.3.46}$$

We assume that $f \in C^{1,\alpha}(\mathbb{T}^n)$ for some $\alpha \in (0,1)$, and f has zero mean over $\mathbb{T}^n$:

$$\int_{\mathbb{T}^n} f(x)dx = 0,$$

and require that the solution h to (1.3.46) also has zero mean:

$$\int_{\mathbb{T}^n} h(x)dx = 0. \tag{1.3.47}$$

Proposition 1.3.10 *Given $f \in C^{1,\alpha}(\mathbb{T}^n)$ with zero mean, equation (1.3.46) has a unique solution $h \in C^{3,\alpha}(\mathbb{T}^n)$ with zero mean. The mapping $f \mapsto h$ is continuous from the set of $C^{1,\alpha}$ functions with zero mean to the set of $C^{3,\alpha}(\mathbb{T}^n)$ functions with zero mean.*

Proof. The Laplacian is a one-to-one map between the set of $C^{m+2,\alpha}$ functions with zero mean and the set of $C^{m,\alpha}(\mathbb{T}^n)$ functions with zero mean, for any $m \in \mathbb{N}$. Thus, we may talk about its inverse that we denote by $(-\Delta)^{-1}$. Equation (1.3.46) is thus equivalent to

$$(I + K)h = (-\Delta)^{-1}f, \tag{1.3.48}$$

with the operator

$$Kh = (-\Delta)^{-1}\Big(-\partial_{p_j} H_\sigma(x, \nabla\phi_\sigma)\partial_{x_j} h + \int_{\mathbb{T}^n} \partial_{p_j} H(y, \nabla\phi_\sigma)\partial_{x_j} h(y)dy\Big). \tag{1.3.49}$$

Exercise 1.3.11 Show that K is a compact operator on the set of functions in $C^{3,\alpha}(\mathbb{T}^n)$ with zero mean.

The problem has been now reduced to showing that the only solution of

$$(I+K)h = 0 \tag{1.3.50}$$

with $h \in C^{3,\alpha}(\mathbb{T}^n)$ with zero mean is $h \equiv 0$. Note that (1.3.50) simply says that h is a solution of (1.3.46) with $f \equiv 0$. Let $e_1^* > 0$ be given by Proposition 1.3.5, with

$$b_j(x) = -\partial_{p_j} H_\sigma(x, \nabla\phi_\sigma). \tag{1.3.51}$$

That is, e_1^* is the positive solution of the equation

$$-\Delta e_1^* + \nabla \cdot (e_1^* b) = 0, \tag{1.3.52}$$

normalized so that $\|e_1^*\|_{L^\infty(\mathbb{T}^n)} = 1$. The uniform Lipschitz bound on $H_\sigma(x,p)$ in the p-variable implies that $b(x)$ is in $L^\infty(\mathbb{T}^n)$, and thus Proposition 1.3.5 can be applied. Multiplying (1.3.46) with $f = 0$ by e_1^* and integrating gives, as $e_1^* > 0$:

$$\int_{\mathbb{T}^n} \partial_{p_j} H_\sigma(y, \nabla\phi_\sigma)\partial_{x_j} h(y)dy = 0.$$

But then, the equation for h becomes simply

$$-\Delta h + b_j(x)\partial_{x_j} h = 0, \quad x \in \mathbb{T}^n,$$

which entails that h is constant, by the Krein-Rutman theorem. Because h has zero mean, we get $h \equiv 0$. $\Box$

Exercise 1.3.12 Let $H_0(x,p)$ satisfy the assumptions of Theorem 1.3.4, and assume that equation (1.3.11), with $H = H_0$,:

$$-\Delta\phi_0 = H_0(x, \nabla\phi_0) - \int_{\mathbb{T}^n} H_0(z, \nabla\phi_0)dz, \tag{1.3.53}$$

has a solution $\phi_0 \in C(\mathbb{T}^n)$. Consider $H_1(x,p) \in C^\infty(\mathbb{T} \times \mathbb{R}^n)$. Prove, with the aid of Propositions 1.3.7 and 1.3.10, and the implicit function theorem that there exist $R_0 > 0$ and $\varepsilon_0 > 0$ such that if

$$|H_1(x,p)| \le \varepsilon_0, \quad \text{for } x \in \mathbb{T}^n \text{ and } |p| \le R_0, \tag{1.3.54}$$

then equation (1.3.11) with $H = H_0 + H_1$:

$$-\Delta\phi = H(x, \nabla\phi) - \int_{\mathbb{T}^n} H(z, \nabla\phi)dz, \tag{1.3.55}$$

has a solution ϕ.

Existence of the solution

We finally prove the existence part of Theorem 1.3.1. Consider $H(x,p)$ satisfying the assumptions of the theorem. As before, we set

$$H_0(x,p) = \sqrt{1+|p|^2} - 1,$$

and

$$H_\sigma(x,p) = (1-\sigma)H_0(x,p) + \sigma H(x,p),$$

so that $H_1(x,p) = H(x,p)$, and consider existence of a solution to (1.3.12):

$$-\Delta\phi_\sigma = H_\sigma(x,\nabla\phi_\sigma) - \int_{\mathbb{T}^n} H_\sigma(z,\nabla\phi_\sigma)dz, \tag{1.3.56}$$

Consider the set

$$\Sigma = \{\sigma \in [0,1] : \quad \text{equation (1.3.56) has a solution}\}.$$

Our goal is to show that $\Sigma = [0,1]$. We know that Σ is non empty, because $0 \in \Sigma$: indeed, $\phi_0(x) \equiv 0$ is a solution to (1.3.56) at $\sigma = 0$. Thus, if we show that Σ is both open and closed in $[0,1]$, this will imply that $\Sigma = [0,1]$, and, in particular, establish the existence of a solution to (1.3.56) for $H_1(x,p) = H(x,p)$.

Now that we know that the linearized problem is invertible, the openness of Σ is a direct consequence of the inverse function theorem, as explained in Exercise 1.3.12. Closedness of Σ is not too difficult to see either: consider a sequence $\sigma_n \in [0,1]$ converging to $\bar{\sigma} \in [0,1]$, and let ϕ_n be a solution to (1.3.56) with $H(x,p) = H_{\sigma_n}(x,p)$, normalized so that

$$\phi_n(0) = 0. \tag{1.3.57}$$

Proposition 1.3.7 implies that

$$\|\nabla\phi_n\|_{L^\infty(\mathbb{T}^n)} \le C,$$

and thus

$$\|H(x,\nabla\phi_n)\|_{L^\infty} \le C.$$

However, this means that ϕ_n solve an equation of the form

$$-\Delta\phi_n = F_n(x), \quad x \in \mathbb{T}^n, \tag{1.3.58}$$

with a uniformly bounded function

$$F_n(x) = H_{\sigma_n}(x,\nabla\phi_n) - \int_{\mathbb{T}^n} H_{\sigma_n}(z,\nabla\phi_n(z))dz. \tag{1.3.59}$$

It follows that that ϕ_n is bounded in $C^{1,\alpha}(\mathbb{T}^n)$, for all $\alpha \in [0,1)$:

$$\|\phi_n\|_{C^{1,\alpha}(\mathbb{T}^n)} \le C. \tag{1.3.60}$$

But this implies, in turn, that the functions $F_n(x)$ in (1.3.59) are also uniformly bounded in $C^\alpha(\mathbb{T}^n)$, hence ϕ_n are uniformly bounded in $C^{2,\alpha}(\mathbb{T}^n)$:

$$\|\phi_n\|_{C^{2,\alpha}(\mathbb{T}^n)} \le C. \tag{1.3.61}$$

Now, the Arzela-Ascoli theorem implies that a subsequence ϕ_{n_k} will converge in $C^2(\mathbb{T}^n)$ to a function $\bar{\phi}$, which is a solution to (1.3.19) with $H = H_{\bar{\sigma}}$. Thus, $\bar{\sigma} \in \Sigma$, and Σ is closed. This finishes the proof of the existence part of the theorem.

1.3.2 Long time convergence and uniqueness of the wave solutions

We will now prove simultaneously the claim of the uniqueness of the speed c and of the profile $\phi(x)$ in Theorem 1.3.1, and the long time convergence for the solutions to the Cauchy problem stated in Theorem 1.3.4.

Let $u(t,x)$ be the solution to (1.3.1)

$$u_t = \Delta u + H(x,\nabla u), \quad t>0,\ x\in\mathbb{T}^n, \tag{1.3.62}$$

with $u(0,x)=u_0(x)\in C(\mathbb{T}^n)$. We also take a speed $c\in\mathbb{R}$ and a solution $\phi(x)$ to

$$\Delta\phi + H(x,\nabla\phi) = c, \tag{1.3.63}$$

without assuming that either c or ϕ is unique.

We wish to prove that there exists $\bar{k}\in\mathbb{R}$ so that $u(t,x)-ct$ is attracted exponentially fast in time to $\phi(x)+\bar{k}$:

$$|u(t,x)-ct-\bar{k}-\phi(x)| \le C_0 e^{-\omega t}, \tag{1.3.64}$$

with some $C_0>0$ and $\omega>0$, such that C_0 depends on the initial condition u_0 but ω does not. The idea is to squeeze the solution between two different wave solutions, and show that the difference between the squeezers tends to zero as $t\to+\infty$.

As a simple remark, we may assume that $c=0$, just by setting

$$\widetilde{H}(x,p) = H(x,p) - c,$$

and dropping the tilde, and this is what we will do. In other words, $\phi(x)$ is the solution to

$$\Delta\phi + H(x,\nabla\phi) = 0. \tag{1.3.65}$$

Let ϕ be any solution to (1.3.65), and set

$$k_0^- = \sup\{k: \quad u(0,x)\ge\phi(x)+k \text{ for all } x\in\mathbb{T}^n\},$$

and

$$k_0^+ = \inf\{k: \quad u(0,x)\le\phi(x)+k \text{ for all } x\in\mathbb{T}^n.\}$$

Because $\phi(x)+k_0^\pm$ solve (1.3.65) with $c=0$, and $u(t,x)$ solves (1.3.62), we have, by the maximum principle:

$$\phi(x)+k_0^- \le u(t,x) \le \phi(x)+k_0^+, \text{ for all } t\ge 0 \text{ and } x\in\mathbb{T}^n. \tag{1.3.66}$$

Now, for all $q\in\mathbb{N}$, let us set

$$k_q^- = \sup\{k: \quad u(t=q,x)\ge\phi(x)+k \text{ for all } x\in\mathbb{T}^n\} = \inf_{x\in\mathbb{T}^n}[u(t=q,x)-\phi(x)], \tag{1.3.67}$$

and

$$k_q^+ = \inf\{k: \quad u(t=q,x)\le\phi(x)+k \text{ for all } x\in\mathbb{T}^n\} = \sup_{x\in\mathbb{T}^n}[u(t=q,x)-\phi(x)]. \tag{1.3.68}$$

The strong maximum principle implies that the sequence k_q^- is increasing, whereas k_q^+ is decreasing, and that, as in (1.3.66), we have

$$\phi(x) + k_q^- \leq u(t,x) \leq \phi(x) + k_q^+, \text{ for all } t \geq q \text{ and } x \in \mathbb{T}^n. \tag{1.3.69}$$

Hence, the theorem will be proved if we manage to show that

$$0 \leq k_q^+ - k_q^- \leq Ca^q, \quad \text{for all } q \geq 0, \tag{1.3.70}$$

with some $C \in \mathbb{R}$ that may depend on the initial condition u_0 and $a \in (0,1)$ that does not depend on u_0. In order to prove (1.3.70), it suffices to show that

$$k_{q+1}^+ - k_{q+1}^- \leq (1 - r_0)(k_q^+ - k_q^-), \tag{1.3.71}$$

with some $r_0 \in (0,1)$. This is a quantification of the strong maximum principle: by the time $t = q+1$ $u(x)$ has to detach "by a fixed amount" from the respective lower and upper bounds $\phi(x) + k_q^\pm$ that hold at $t = q$. Such estimates typically rely on the Harnack inequality, and this is what we will use.

To bring the Harnack inequality in, note that the function

$$w(t,x) = u(t,x) - \phi(x) - k_q^-$$

is nonnegative for $t \geq q$, and solves an equation of the form

$$\partial_t w - \Delta w + b_j(t,x)\partial_{x_j} w = 0, \qquad t > q, \quad x \in \mathbb{T}^n, \tag{1.3.72}$$

with a bounded drift $b(t,x)$ given by

$$b(t,x) = \int_0^1 \nabla_p H(x, (1-s)\nabla\phi(x) + s\nabla u(t,x))ds, \tag{1.3.73}$$

so that

$$b(t,x) \cdot [\nabla u(t,x) - \nabla\phi(x)] = H(x, \nabla u(t,x)) - H(x, \nabla\phi(x)),$$

and

$$|b_j(t,x)| \leq C_L, \quad \text{for all } t \geq q \text{ and } x \in \mathbb{T}^n. \tag{1.3.74}$$

The Harnack inequality and (1.3.74) imply that there exists $r_0 > 0$ that depends only on C_L such that

$$\inf_{x\in\mathbb{T}^n} w(q+1,x) \geq r_0 \sup_{x\in\mathbb{T}^n} w(q,x). \tag{1.3.75}$$

Exercise 1.3.13 Explain on the intuitive level why the constant in the Harnack inequality should depend only on the L^∞-norm of the vector field $b(x)$.

Using (1.3.67) and (1.3.68), together with (1.3.75), we may write

$$\begin{aligned} r_0 \sup_{x\in\mathbb{T}^n} w(q,x) &= r_0 \sup_{x\in\mathbb{T}^n}[u(q,x) - \phi(x) - k_q^-] = r_0[k_q^+ - k_q^-] \leq \inf_{x\in\mathbb{T}^n} w(q+1,x) \\ &= \inf_{x\in\mathbb{T}^n}[u(q+1,x) - \phi(x) - k_q^-] = k_{q+1}^- - k_q^-, \end{aligned} \tag{1.3.76}$$

so that

$$k_{q+1}^- \geq k_q^- + r_0[k_q^+ - k_q^-]. \tag{1.3.77}$$

As $k_{q+1}^+ \leq k_q^+$, it follows that

$$k_{q+1}^+ - k_{q+1}^- \leq k_q^+ - k_q^- - r_0(k_q^+ - k_q^-) \leq (1 - r_0)(k_q^+ - k_q^-), \tag{1.3.78}$$

which is (1.3.71). This implies the geometric decay as in (1.3.70), hence the theorem, because of (1.3.69) and (1.3.70). Note that the constant

$$a = 1 - r_0$$

comes from the Harnack inequality and does not depend on the initial condition u_0 but only on the Lipschitz constant C_L of $H(x, p)$. □

Exercise 1.3.14 (i) Why does the uniqueness of c and of the profile $\phi(x)$ follow?
(ii) How is the constant ω in Theorem 1.3.4 related to the constant a in the above proof?

Exercise 1.3.15 Consider a modified equation, not quite of the Hamilton-Jacobi form:

$$u_t - \Delta u = R(x, u)\sqrt{1 + |\nabla u|^2}, \tag{1.3.79}$$

where $R(x, u)$ is a smooth, positive function, that is 1-periodic in x and 1-periodic in u. You may either assume that the Cauchy problem for (1.3.79) with $u(0, x) = u_0(x)$ is well posed for $u_0 \in C(\mathbb{T}^n)$ or prove that.
(i) Prove the existence of a unique $T > 0$ such that equation (1.3.79) has solutions of the form

$$u(t, x) = \frac{t}{T} + \phi(t, x), \tag{1.3.80}$$

where ϕ is T-periodic in t and 1-periodic in x. We will call such a solution a pulsating wave solution. Why is it not reasonable to expect that under the above assumptions (1.3.79) has a wave solution of the form $u(t, x) = ct + \psi(x)$ with a 1-periodic function $\psi(x)$?
(ii) Show that every solution of the Cauchy problem which is initially 1-periodic in x converges, exponentially fast in time, to a wave solution of the form (1.3.80). If in doubt, [lease consult [111]. Note that the topological degree argument used in that reference can be replaced by a more elementary implicit function theorem argument we have used in the existence proof here.

1.4 A glimpse of the classical solutions to the Hamilton-Jacobi equations

1.4.1 Smooth solutions and their limitations

We now turn our attention to first order inviscid Hamilton-Jacobi equations of the form

$$u_t + H(x, \nabla u) = 0. \tag{1.4.1}$$

The standard philosophy of the construction of a solution to a first order equation is to find its values on special curves, known as characteristics, that will, hopefully, fill the whole space. This is the strategy that is also classically used to solve (1.4.1). Consider a time $t > 0$ and a point $x \in \mathbb{R}^n$. In order to assign a value to $u(t,x)$ we consider a curve $\gamma(s)$, with $s \in [0,t]$, such that $\gamma(t) = x$, and set

$$p(s) = \nabla u(s, \gamma(s)).$$

Here, $u(t,x)$ is the sought for solution to (1.4.1). Assuming that everything is smooth we have, using the dot to denote the differentiation in s:

$$\begin{aligned}\dot{p}_k(s) &= \partial_{x_k} u_t(s,\gamma(s)) + \frac{\partial^2 u(s,\gamma(s))}{\partial x_k \partial x_m}\dot{\gamma}_m(s)\\ &= -\frac{\partial H(\gamma(s),p(s))}{\partial x_k} - \frac{\partial H(\gamma(s),p(s))}{\partial p_m}\frac{\partial^2 u(s,\gamma(s))}{\partial x_k \partial x_m} + \frac{\partial^2 u(s,\gamma(s))}{\partial x_k \partial x_m}\dot{\gamma}_m(s).\end{aligned} \tag{1.4.2}$$

We see that it is convenient to choose $\gamma(s)$ that satisfies the following system of ODEs:

$$\begin{aligned}\dot{\gamma}(s) &= \nabla_p H(\gamma(s), p(s))\\ \dot{p}(s) &= -\nabla_x H(\gamma(s), p(s)),\end{aligned} \tag{1.4.3}$$

for $0 \le s \le t$. This dynamical system is to be complemented by the boundary conditions at $s = 0$ and $s = t$:

$$p(0) = \nabla u_0(\gamma(0)), \qquad \gamma(t) = x. \tag{1.4.4}$$

The system (1.4.3) has the form of a Hamiltonian system with the Hamiltonian $H(x,p)$, and the curves $(\gamma(s), p(s))$ are called the characteristic curves. In order to solve (1.1.1), we need to find a solution to (1.4.3)-(1.4.4), and it would be excellent to prove that such solution is unique. The trouble is that there is no good reason, in general, for existence and uniqueness of a solution to this boundary value problem.

Exercise 1.4.1 Consider $x_0 \in \mathbb{R}^n$ and $t > 0$ and assume that $u(t,x)$ is smooth in a ball around x_0. Prove, for instance, with the help of the implicit function theorem, that the boundary value problem (1.4.3)-(1.4.4) has a unique solution $(\gamma(s), p(s))$ as soon as t is small enough and x is in the vicinity of x_0, and that this solution is smooth in t and x.

Once $\gamma(s)$ and $p(s)$ are constructed, we may assign a value to $u(t,x)$ as follows. The function

$$\varphi(s) = u(s, \gamma(s))$$

satisfies

$$\begin{aligned}\dot{\varphi}(s) &= u_t(s,\gamma(s)) + \dot{\gamma}(s)\nabla u(s,\gamma(s)) = u_t(s,\gamma(s)) + \dot{\gamma}(s)\cdot p(s)\\ &= -H(\gamma(s),p(s)) + \dot{\gamma}(s)\cdot p(s).\end{aligned} \tag{1.4.5}$$

Integrating (1.4.5) from $s = 0$ to $s = t$ gives an expression for $u(t,x)$ in terms of the curves $\gamma(s)$ and $p(s)$, $0 \le s \le t$:

$$u(t,x) = u_0(\gamma(0)) + \int_0^t \big(-H(\gamma(s),p(s)) + \dot{\gamma}(s)\cdot p(s)\big)ds. \tag{1.4.6}$$

Exercise 1.4.2 Check that (1.4.6) indeed gives a solution to (1.4.1) such that $u(0,x) = u_0(x)$.

To see that this strategy can not always lead to smooth solutions for all times, consider what may be the simplest nonlinear equation in one space dimension

$$u_t + \frac{u_x^2}{2} = 0 \text{ for } t > 0 \text{ and } x \in \mathbb{R}, \quad u(0,x) = u_0(x). \tag{1.4.7}$$

In that case, the system (1.4.3) becomes

$$\begin{aligned} \dot{\gamma}(s) &= p(s) \\ \dot{p}(s) &= 0. \end{aligned} \tag{1.4.8}$$

This system is supplemented by the boundary conditions (1.4.4)

$$p(0) = \nabla u_0(\gamma(0)), \qquad \gamma(t) = x. \tag{1.4.9}$$

The solution to the boundary value problem (1.4.8)-(1.4.9) amounts to finding $\gamma(0)$ solving the equation

$$x = \gamma(0) + t u_0'(\gamma(0)), \tag{1.4.10}$$

for a given $t > 0$ and $x \in \mathbb{R}$. The issue is that this equation may, or may not have a unique solution $\gamma(0)$. If $u_0'' > 0$, the the right side of (1.4.10) is increasing in $\gamma(0)$. Thus, the solution is unique and we are on the safe side. However, if $u_0''(x_0) < 0$ at some point x_0, the right side of this equation is not increasing in $\gamma(0)$ as soon as

$$t \geq t_c = \frac{1}{\sup(-u_0'')},$$

and uniqueness of the solution fails for $t > t_c$.

Thus, we need a more elaborate theory to construct solutions to (1.4.1) in general. Nevertheless, the characteristic curves will turn out to be extremely important, and in the rest of this section, we wish to show the reader one interesting situation where smooth solutions can be constructed.

Before we end this short section, let us mention, in the form of an exercise (this will be revisited in the context of viscosity solutions), a very strong form of uniqueness.

Exercise 1.4.3 *(Finite speed of propagation).* Let $H(x,p)$ be smooth and uniformly Lipschitz with respect to its second variable. Let u_0 and v_0 be two smooth, compactly supported initial conditions, and assume that each generates a smooth solution to the Cauchy problem for (1.4.1), on a common time interval $[0,T]$. Compute, in terms of $\nabla_p H$, a constant K such that, if

$$\text{dist}\big(x, \text{supp}(u_0 - v_0))\big) > Kt,$$

then $u(t,x) = v(t,x)$. Hint: it may be helpful to solve, first, the following question: let $b(t,x)$ be smooth and uniformly Lipschitz in its second variable. Let u_0 be a smooth compactly supported function, and $u(t,x)$ the solution to

$$\begin{aligned} &u_t + b(t,x) \cdot \nabla u = 0, \quad t > 0, x \in \mathbb{R}^n \\ &u(0,x) = u_0(x). \end{aligned} \tag{1.4.11}$$

Show that if

$$\mathrm{dist}(x, \mathrm{supp}(u_0)) > t\|b\|_\infty,$$

then $u(t,x) = 0$. Note that, since (1.4.11) is a linear equation, it can be solved by the method of characteristics.

1.4.2 An example of classical global solutions

We now discuss a situation when classical smooth solutions do exist. Consider solutions to the equation

$$u_t + \frac{1}{2}|\nabla u|^2 - R(x) = 0, \tag{1.4.12}$$

with an initial condition $u(0,x) = u_0(x)$. The Hamiltonian

$$H(x,p) = \frac{|p|^2}{2} - R(x), \tag{1.4.13}$$

comes from the classical mechanics: the first term above corresponds to the kinetic energy, and the second is the potential energy.

We assume that both $u_0(x)$ and $R(x)$ are strictly convex smooth functions on $\mathbb{R}^n$. That is, there is $\alpha \in (0,1)$ so that, for all $x \in \mathbb{R}^n$ and all $\xi \in \mathbb{R}^n$ we have:

$$\alpha|\xi|^2 \le (D^2u_0(x)\xi\cdot\xi) \le \alpha^{-1}|\xi|^2, \quad \alpha|\xi|^2 \le (D^2R(x)\xi\cdot\xi) \le \alpha^{-1}|\xi|^2. \tag{1.4.14}$$

Exercise 1.4.4 First, consider the case $R = 0$. Argue informally, just by looking at the equation and using pictures that if $u_0(x)$ is strictly convex but its Hessian is uniformly bounded then the graph of $u(t,x)$ should not form a corner, and if $u_0(x)$ is strictly concave but its Hessian is uniformly bounded then it is plausible that the graph of $u(t,x)$ will form a corner. It may be helpful to start by looking at $u(0,x) = |x|^2$ and $u(0,x) = -|x|^2$.

We now use the approach via the characteristic curves to show that a smooth solution exists for all $t > 0$ under the above assumptions. For the Hamiltonian (1.4.13), the characteristic system (1.4.3)-(1.4.4) reduces to

$$\dot\gamma(s) = p(s), \quad \dot p(s) = \nabla R(\gamma(s)),$$

which can be written as a second order equation

$$-\ddot\gamma + \nabla R(\gamma) = 0, \tag{1.4.15}$$

with the boundary conditions

$$\dot\gamma(0) - \nabla u_0(\gamma(0)) = 0, \quad \gamma(t) = x. \tag{1.4.16}$$

To establish uniqueness and smoothness of the solution $u(t,x)$ to (1.4.12) with the initial condition $u(0,x) = u_0(x)$, we need to prove that (1.4.15)-(1.4.16) has a unique solution $\gamma(s)$ that depends smoothly on t and x. Then, $u(t,x)$ will be given by (1.4.6), which, in the present case takes the form

$$\begin{aligned} u(t,x) &= u_0(\gamma(0)) + \int_0^t \big(-H(\gamma(s),p(s)) + \dot\gamma(s)\cdot p(s)\big)ds \\ &= u_0(\gamma(0)) + \int_0^t \Big(\frac{|\dot\gamma(s)|^2}{2} + R(\gamma(s))\Big)ds. \end{aligned} \tag{1.4.17}$$

Existence of the characteristic curves

To construct a solution to (1.4.15)-(1.4.16), we observe that this system is the Euler-Lagrange equation for the energy functional

$$J_{t,x}(\gamma) = u_0(\gamma(0)) + \int_0^t \Big(\frac{|\dot\gamma(s)|^2}{2} + R(\gamma(s))\Big) ds, \tag{1.4.18}$$

over $H^1([0,t])$, with the constraint $\gamma(t) = x$. To see this, let us consider a minimizer $\gamma(s)$ of the functional $J_{t,x}$ over all such γ. Consider a small perturbation $\gamma(s) + \delta\psi(s)$ with a smooth $\psi(s)$ such that $\psi(t) = 0$. Then, we have

$$Q(\delta) := J_{t,x}(\gamma(s) + \delta\psi(s)) = u_0(\gamma(0) + \delta\psi(0)) + \int_0^t \Big(\frac{|\dot\gamma(s) + \delta\dot\psi(s)|^2}{2} + R(\gamma(s) + \delta\psi(s))\Big) ds. \tag{1.4.19}$$

As $\gamma(s)$ is a minimizer for $J_{t,x}(\gamma)$, we have $Q'(0) = 0$. We compute

$$\begin{aligned} Q'(0) &= \nabla u_0(\gamma(s)) \cdot \psi(0) + \int_0^t \big(\dot\gamma(s) \cdot \dot\psi(s) + \nabla R(\gamma(s))\psi(s)\big) ds \\ &= \nabla u_0(\gamma(0)) \cdot \psi(0) + \dot\gamma(t)\psi(t) - \dot\gamma(0)\psi(0) + \int_0^t [-\ddot\gamma(s) + \nabla R(\gamma(s))]\psi(s) ds. \end{aligned} \tag{1.4.20}$$

As $\psi(t) = 0$ and $\psi(s)$ is arbitrary, we deduce from the integral term in (1.4.20) that $\gamma(s)$ must satisfy

$$-\ddot\gamma(s) + \nabla R(\gamma(s)) = 0, \tag{1.4.21}$$

which is (1.4.15). In addition, we see from the boundary term at $s = 0$ in (1.4.20) that $\gamma(s)$ satisfies the boundary condition

$$\dot\gamma(0) = \nabla u_0(\gamma(0)), \tag{1.4.22}$$

which is (1.4.16).

The reader will surely have noticed the striking similarity between the function $J_{t,x}(\gamma)$ and the form of the solution in (1.4.17). The difference is, of course, that the curve $\gamma(s)$ in (1.4.17) is the solution to (1.4.15)-(1.4.16) while in (1.4.18) the curve $\gamma(s)$ is any element of $H^1([0,t])$ such that $\gamma(t) = x$. This observation is strongly connected to the optimal control formulation for the Hamilton-Jacobi equation that we have discussed earlier and that we will revisit soon again.

Exercise 1.4.5 Verify that claim: show that if the minimizer of $J_{t,x}(\gamma)$ over the set

$$S = \{\gamma \in H^1[0,t] : \ \gamma(t) = x\}$$

exists and is smooth then it satisfies both (1.4.15) and the boundary condition at $s = 0$ in (1.4.16). Next, define what it means for $\gamma \in H^1[0,t]$ (without assuming γ is smooth) to be a weak solution to (1.4.15)-(1.4.16) and show that a minimizer of $J_{t,x}$ over S (if it exists) is a weak solution.

As both $u_0(x)$ and $R(x)$ are strictly convex, they are bounded from below, and it is easy to see that the functional $J_{t,x}$ is bounded from below over S. Let us define

$$\bar{J}_{t,x} = \inf_{\gamma \in S} J_{t,x}(\gamma), \tag{1.4.23}$$

and let $\gamma_n \in S$ be a minimizing sequence, so that $J_{t,x}(\gamma_n)$ decreases to $\bar{J}_{t,x}$. Once again, as u_0 and R are bounded from below, there exists $C > 0$ so that

$$\int_0^t |\gamma_n'(s)|^2 ds \le C,$$

for all n. As, in addition, $\gamma_n(t) = x$ for all n, there is a subsequence, that we will still denote by γ_n that converges uniformly over $[0,t]$, and weakly in $H^1([0,t])$ to a limit $\bar{\gamma}_{t,x} \in S$.

To prove that $J_{t,x}(\bar{\gamma}_{t,x}) = \bar{J}_{t,x}$ we simply observe that by the weak convergence we have

$$\|\bar{\gamma}_{t,x}'\|_{L^2}^2 \le \liminf_{n \to +\infty} \|\gamma_n'\|_{L^2}^2,$$

which, combined with the uniform convergence of γ_n to $\bar{\gamma}_{t.x}$ on $[0,t]$ implies that

$$J_{t,x}(\bar{\gamma}_{t,x}) \le \lim_{n \to +\infty} J_{t,x}(\gamma_n) = \bar{J}_{t,x},$$

and thus

$$J_{t,x}(\bar{\gamma}_{t,x}) = \bar{J}_{t,x}.$$

Thus, $\bar{\gamma}_{t,x}$ is a minimizer in (1.4.23).

Uniqueness of the characteristic curve

To prove the uniqueness of the minimizer, we will use the convexity of $u_0(x)$ and $R(x)$ and not just their boundedness from below. Let γ_1 and γ_2 be two solutions to (1.4.15)-(1.4.16). The difference

$$\widetilde{\gamma} = \gamma_2 - \gamma_1.$$

satisfies

$$-\widetilde{\gamma}_k'' + A_{kj}(s)\widetilde{\gamma}_j = 0, \quad 1 \le k \le n, \tag{1.4.24}$$

with the boundary conditions

$$\widetilde{\gamma}_k'(0) - B_{kj}\widetilde{\gamma}_j(0) = 0, \quad \widetilde{\gamma}_k(t) = 0, \quad 1 \le k \le n. \tag{1.4.25}$$

The matrices A and B are given by

$$A_{kj}(s) = \int_0^1 \frac{\partial^2 R(\gamma_1(s) + \sigma(\gamma_2(s) - \gamma_1(s)))}{\partial x_k \partial x_j} d\sigma,$$

and

$$B_{kj} = \int_0^1 \frac{\partial^2 u_0(\gamma_1(0) + \sigma(\gamma_2(0) - \gamma_1(0)))}{\partial x_k \partial x_j} d\sigma.$$

Let us take the inner product of (1.4.24) with $\widetilde{\gamma}$, and integrate in time. This gives

$$\int_0^t \big(|\widetilde{\gamma}'(s)|^2 + (A(s)\widetilde{\gamma}(s)\cdot\widetilde{\gamma}(s))\big)ds + (B\widetilde{\gamma}(0)\cdot\widetilde{\gamma}(0)) = 0. \tag{1.4.26}$$

Using assumptions (1.4.14) on the convexity of the functions $R(x)$ and $u_0(x)$, we deduce that the matrices $A(s)$ an B are strictly positive definite. Thus, (1.4.26) implies that $\widetilde{\gamma}(s)\equiv 0$, so that the minimizer is unique. Hence, $u(t,x)$ is well-defined by (1.4.17):

$$\begin{aligned} u(t,x) &= u_0(\gamma(0)) + \int_0^t \big(-\frac{|p(s)|^2}{2} + R(\gamma(s)) + \dot{\gamma}(s)\cdot p(s)\big)ds \\ &= u_0(\gamma(0)) + \int_0^t \big(\frac{|\dot{\gamma}(s)|^2}{2} + R(\gamma(s))\big)ds. \end{aligned} \tag{1.4.27}$$

This may be rephrased as

$$u(t,x) = \inf_{\gamma(t)=x}\Big(u_0(\gamma(0)) + \int_0^t \big(\frac{|\gamma'(s)|^2}{2} + R(\gamma(s))\big)ds\Big). \tag{1.4.28}$$

This formula, known as the Lax-Oleinik formula, is the starting point of the Lagrangian theory of Hamilton-Jacobi equations, and has immense implications. We will spend some time with this aspect of Hamilton-Jacobi equations later in this chapter. We will see that we can take it as a good definition of a solution to the Cauchy problem, at least when the Hamiltonian is strictly convex in p.

Smoothness of the solution

Let us quickly examine the smoothness of $u(t,x)$ in x in the set-up of the present section. We see from (1.4.27) that it is equivalent to the smoothness of the minimizer γ in x. If $h\in\mathbb{R}$ and $i\in\{1,...,n\}$, consider the partial difference

$$\gamma_h^i(s) = \frac{\gamma_{t,x+he_i}(s) - \gamma_{t,x}(s)}{h}.$$

It solves a system similar to (1.4.24), except for the boundary condition at $s=t$ that is now $\gamma_h^i(t)=e_i$. The exact same integration by parts argument yields the uniform boundedness of $\|\gamma_h^i\|_{H^1}$, hence the uniform boundedness of γ_h^i. Sending h to 0 and repeating the analysis shows that γ_h^i converges to the unique solution of an equation of the type (1.4.24), with

$$A(s) = D^2R(\gamma_{t,x}(s)), \qquad B = D^2u_0(\gamma_{t,x}(0)).$$

This argument may be repeated over and over again, to yield the C^∞ smoothness of $\gamma_{t,x}$ in t and x, as long as u_0 and $R(x)$ are infinitely differentiable. Finally, using (1.4.6) we can conclude that

$$u(t,x) = \bar{J}_{t,x},$$

is infinitely differentiable as well.

Exercise 1.4.6 Show that u is convex in x, for all $t > 0$, in two ways. First, fix $\xi \in \mathbb{R}^n$ and get a differential equation for $Q(t,x) = (D^2u(t,x)\xi \cdot \xi)$. Use a maximum principle type argument to conclude that $Q(t,x) > 0$ for all $t > 0$ and $x \in \mathbb{R}^n$. An alternative and more elegant way is to proceed as follows.

(i) Assume the existence of $\kappa > 0$ such that, for all $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$ and $\sigma \in [0,1]$, we have:

$$u(t, \sigma x + (1-\sigma)y) \leq \sigma u(t,x) + (1-\sigma)u(t,y) - \kappa\sigma(1-\sigma)|x-y|^2. \tag{1.4.29}$$

Show that then the function $u(t,x)$ is strictly convex.

(ii) Show that there exists $\lambda > 0$ such that if $\gamma_{t,x}$ and $\gamma_{t,y}$ are, respectively, the minimizing curves for $u(t,x)$ and $u(t,y)$, then

$$\begin{aligned} u(t, \sigma x + (1-\sigma)y) \leq{}& \sigma u(t,x) + (1-\sigma)u(t,y) \\ &- \lambda\sigma(1-\sigma)\Big(|\gamma_{t,x}(0) - \gamma_{t,y}(0)|^2 + \|\gamma_{t,x} - \gamma_{t,y}\|^2_{H^1([0,t])}\Big). \end{aligned} \tag{1.4.30}$$

Hint: use the test curve $\gamma_\sigma = \sigma\gamma_{t,x} + (1-\sigma)\gamma_{t,y}$ in the Lax-Oleinik formula (1.4.28) for $u(t, \sigma x + (1-\sigma)y)$, together with the convexity of the functions $u_0(x)$ and $R(x)$.

(iii) Finish the proof of (1.4.29), by noticing that

$$|\gamma_{t,x}(0) - \gamma_{t,y}(0)|^2 = |x-y|^2 - \int_0^t \frac{d}{ds}|\gamma_{t,x}(s) - \gamma_{t,y}(s)|^2 ds.$$

The qualitative behavior of $u(t,x)$ can be investigated further, implying the large time stabilization of the whole solution. We will come back to this class of questions later, when we study the large time behavior of viscosity solutions on the torus. For the time being, we leave the classical theory.

1.5 Viscosity solutions

We have just seen that, in order to find reasonable solutions to an inviscid Hamilton-Jacobi equation

$$u_t + H(x, \nabla u) = 0, \tag{1.5.1}$$

we should relax the constraint that "u is continuously differentiable". The first idea would be to replace it by "u is Lipschitz", and require (1.5.1) to hold almost everywhere. Alas, there are, in general, several such solutions to the Cauchy problem for (1.5.1) with a Lipschitz (or even smooth) initial condition. This parallels the fact that the weak solutions to the conservation laws are not unique – for uniqueness, one must require that the weak solution satisfies the entropy condition. See, for instance, [93] for a discussion of these issues. A simple illustration of this phenomenon is to consider the Hamilton-Jacobi equation

$$u_t + \frac{1}{2}u_x^2 = 0, \tag{1.5.2}$$

in one dimension, with the Lipschitz continuous initial condition

$$u_0(x) = 0 \text{ for } x \leq 0 \text{ and } u_0(x) = x \text{ for } x > 0. \tag{1.5.3}$$

It is easy to check that one Lipschitz solution to (1.5.2) that satisfies this equation almost everywhere and obeys the initial condition (1.5.3) is

$$u^{(1)}(t,x) = 0 \text{ for } x < t/2 \text{ and } u^{(1)}(t,x) = x - t/2 \text{ for } x > t/2.$$

However, another solution to (1.5.4)-(1.5.3) is given by

$$u^{(2)}(t,x) = 0 \text{ for } x < 0,\ u^{(2)}(t,x) = \frac{x^2}{2t} \text{ for } 0 < x < t \text{ and } u(t,x) = x - \frac{t}{2} \text{ for } x > t.$$

Exercise 1.5.1 Consider the solution $u^\varepsilon(t,x)$ to a viscous version of (1.5.4):

$$u_t^\varepsilon + \frac{1}{2}(u_x^\varepsilon)^2 = \varepsilon u_{xx}^\varepsilon, \tag{1.5.4}$$

also with the initial condition $u^\varepsilon(0,x) = u_0(x)$, as in (1.5.3). Use the Hopf-Cole transform

$$v^\varepsilon(t,x) = \exp\Big(-\frac{u^\varepsilon(t/\varepsilon,x)}{2\varepsilon}\Big),$$

to show that v^ε satisfies the standard heat equation

$$v_t^\varepsilon = v_{xx}^\varepsilon.$$

Find $v^\varepsilon(t,x)$ explicitly and use this to show that

$$u^\varepsilon(t,x) \to u^{(2)}(t,x) \text{ as } \varepsilon \to 0.$$

A natural question is, therefore, to know if an additional condition, less stringent than the C^1-regularity, but stronger than the mere Lipschitz regularity, enables us to select a unique solution to the Cauchy problem – as the notion of the entropy solutions does for the conservation laws. Exercise 1.5.1 suggests that regularizing the inviscid Hamilton-Jacobi equation with a small diffusion can provide one such approach, but for more general Hamilton-Jacobi equations than (1.5.4), for which the Hopf-Cole transform is not available, this procedure would be much less explicit.

The above considerations have motivated the introduction, by Crandall and Lions [44], at the beginning of the 1980's, of the notion of a *viscosity solution* to (1.1.1). The idea is to select, among all the solutions of (1.1.1), "the one that has a physical meaning", intrinsically, without directly appealing to the small diffusion regularization, – though understanding the connection to physics may require some additional thought. Being weaker than the notion of a classical solution, it introduces new difficulties to the existence, regularity and uniqueness issues, as well as into getting insight into the qualitative properties of solutions.

As a concluding remark to this introduction, we must mention that we will by no means do justice to a very rich subject in this short section and provide just a brief glance of a still developing subject. The reader interested to learn more may enjoy reading Barles [158], or Lions [93] as a starting point.

1.5.1 The definition and the basic properties of the viscosity solutions

The definition of a viscosity solution

Let us begin with more general equations than (1.1.1) – we will restrict the assumptions as the theory develops. Consider the Cauchy problem

$$u_t + F(x, u, \nabla u) = 0, \quad t > 0, \ x \in \mathbb{T}^n, \tag{1.5.5}$$

with a continuous initial condition $u(0, x) = u_0(x)$, and $F \in C(\mathbb{T}^n \times \mathbb{R} \times \mathbb{R}^n; \mathbb{R})$.

In order to motivate the notion of a viscosity solution, one takes the point of view that the smooth solutions to the regularized problem

$$u^\varepsilon_t + F(x, u^\varepsilon, \nabla u^\varepsilon) = \varepsilon \Delta u^\varepsilon \tag{1.5.6}$$

are a good approximation to $u(t, x)$. Existence of the solution to the Cauchy problem for (1.5.6) for $\varepsilon > 0$ is not really an issue since the diffusivity $\varepsilon > 0$ is positive, Hence, a natural attempt would be to pass to the limit $\varepsilon \downarrow 0$ in (1.5.6). It is possible to prove that there is a unique limiting solution and that one actually ends up with a nonlinear semigroup. In particular, one may show that, if we take this notion of solution as a definition, there are uniqueness and contraction properties analogous to what we will see below – see [93] for further details. Taking this limit as a definition, however, raises an important issue: there is always the danger that the solution depends on the underlying regularization – why regularize with the Laplacian? What if we were to regularize differently? For instance, what if we would consider a dispersive regularization in one dimension

$$u^\varepsilon_t + F(x, u^\varepsilon, u^\varepsilon_x) = \varepsilon u^\varepsilon_{xxx}, \quad x \in \mathbb{R}, \tag{1.5.7}$$

which is a generalized Korteweg-de Vries equation, and let $\varepsilon \to 0$ in (1.5.7) instead?

We now describe an alternative and more intrinsic approach, instead of using (1.5.6) in this very direct fashion of passing to the limit $\varepsilon \downarrow 0$. The idea is that the key property that should be inherited from the diffusive regularization is the maximum principle, as it is usually inherent in the origins of such models in the corresponding applications, be it physics, such as motion of interfaces, or optimal control problems. There is an interesting separate question of what happens as $\varepsilon \to 0$ to the solutions coming from regularizations that do not admit the maximum principle, such as (1.5.7). The situation is not quite trivial, especially for non-convex fluxes F – we refer an interested reader to [92].

Our approach will be to use the comparison principle idea to extend the notions of a sub-solution and a super-solution to (1.5.5) and then simply say that a function $u(t, x)$ is a solution to (1.5.5) if it is both a sub-solution and a super-solution. To understand the upcoming definition of a viscosity sub-solution to (1.5.5), consider first a smooth sub-solution $u(t, x)$ to the regularized problem (1.5.6):

$$u_t + F(x, u, \nabla u) \le \varepsilon \Delta u. \tag{1.5.8}$$

Let us take a smooth function $\phi(t, x)$ such that the difference $\phi - u$ attains its minimum at a point (t_0, x_0). One may simply think of the case when $\phi(t_0, x_0) = u(t_0, x_0)$ and $\phi(t, x) \ge u(t, x)$ elsewhere. Then, at this point we have

$$u_t(t_0, x_0) = \phi_t(t_0, x_0), \quad \nabla\phi(t_0, x_0) = \nabla u(t_0, x_0),$$

and

$$D^2\phi(t_0,x_0) \ge D^2u(t_0,x_0),$$

in the sense of the quadratic forms. It follows that

$$\begin{aligned}&\phi_t(t_0,x_0)+F(x_0,u(t_0,x_0),\nabla\phi(t_0,x_0))-\varepsilon\Delta\phi(t_0,x_0)\\&\le u_t(t_0,x_0)+F(x_0,u(t_0,x_0),\nabla u(t_0,x_0))-\varepsilon\Delta u(t_0,x_0)\le 0.\end{aligned} \tag{1.5.9}$$

In other words, if u is a smooth sub-solution to (1.5.6), and ϕ is a smooth function that touches u at the point (t_0,x_0) from above, then ϕ is also a sub-solution to (1.5.6) at this point.

In a similar vein, if $u(t,x)$ is a smooth super-solution to the regularized problem:

$$u_t + F(x,u,\nabla u) \ge \varepsilon\Delta u, \tag{1.5.10}$$

we consider a smooth function $\phi(t,x)$ such that the difference $\phi-u$ attains its maximum at a point (t_0,x_0). Again, we may assume without loss of generality that $\phi(t_0,x_0)=u(t_0,x_0)$ and $\phi(t,x)\le u(t,x)$ elsewhere. Then, at this point we have

$$\phi_t(t_0,x_0)+F(x_0,u(t_0,x_0),\nabla\phi(t_0,x_0))-\varepsilon\Delta\phi(t_0,x_0)\ge 0, \tag{1.5.11}$$

by a computation similar to (1.5.9). That is, if u is a smooth super-solution to (1.5.6), and ϕ is a smooth function that touches u at (t_0,x_0) from below, then ϕ is also a super-solution to (1.5.6) at this point.

These two observations lead to the following definition, where we simply drop the requirement that u is smooth, only use the regularity of the test function that touches it from above or below, and send $\varepsilon\to 0$ in (1.5.9) and (1.5.11).

Definition 1.5.2 *A continuous function $u(t,x)$ is a viscosity sub-solution to*

$$u_t + F(x,u,\nabla u) = 0, \tag{1.5.12}$$

if, for all test functions $\phi\in C^1([0,+\infty)\times\mathbb{T}^n)$ and all $(t_0,x_0)\in(0,+\infty)\times\mathbb{T}^n$ such that (t_0,x_0) is a local minimum for $\phi-u$, we have:

$$\phi_t(t_0,x_0)+F(x_0,u(t_0,x_0),\nabla\phi(t_0,x_0))\le 0. \tag{1.5.13}$$

Furthermore, a continuous function $u(t,x)$ is a viscosity super-solution to (1.5.12) if, for all test functions $\phi\in C^1((0,+\infty)\times\mathbb{T}^n)$ and all $(t_0,x_0)\in(0,+\infty)\times\mathbb{T}^n$ such that the point (t_0,x_0) is a local maximum for the difference $\phi-u$, we have:

$$\phi_t(t_0,x_0)+F(x_0,u(t_0,x_0),\nabla\phi(t_0,x_0))\ge 0. \tag{1.5.14}$$

Finally, $u(t,x)$ is a viscosity solution to (1.5.12) if it is both a viscosity sub-solution and a viscosity super-solution to (1.5.12).

Definition 1.5.2 extends to steady equations of the type

$$F(x,u,\nabla u) = 0 \text{ on } \mathbb{T}^n,$$

by requiring that $u(x)$ is a viscosity sub-solution (respectively, super-solution) to

$$u_t + F(x, u, \nabla u) = 0,$$

that happens to be time-independent.

This definition was introduced by Crandall and Lions in their seminal paper [44]. The name "viscosity solution" comes out of the diffusive regularization we have discussed above. Definition 1.5.2 is intrinsic and bypasses the philosophical question we have mentioned above: "Why regularize with the Laplacian?" much like the notion of an entropy solution does this for the conservation laws. We stress, however, that it does make the assumption that the underlying model must respect the comparison principle. Let us also note that the notion of a viscosity solution has turned out to be also very much relevant to the second order elliptic and parabolic equations – especially those fully nonlinear with respect to the Hessian of the solution. There have been spectacular developments, which are out of the scope of these notes.

The main issue we will need to face soon is whether such a seemingly weak definition has any selective power – can it possibly ensure uniqueness of the solution? The expectation is that it should, due to the general principle that "the comparison principle implies uniqueness".

First, the following exercises may help the reader gain some intuition.

Exercise 1.5.3 Show that a C^1 solution to

$$u_t + F(x, u, \nabla u) = 0, \quad t > 0, \ x \in \mathbb{T}^n, \tag{1.5.15}$$

is a viscosity solution.

Exercise 1.5.4 Consider the Hamilton-Jacobi equation

$$u_t + u_x^2 = 0, \quad x \in \mathbb{R}. \tag{1.5.16}$$

(i) Which of the following two functions is a viscosity solution to (1.5.16): $v(t, x) = |x| - t$ or $w(t, x) = -t - |x|$? Hint: pay attention to the fact that at the point $x = 0$ a smooth function $\phi(t, x)$ can only touch $v(t, x)$ from the bottom, and $w(t, x)$ from the top. This will tell you something about $|\phi_x(t, 0)|$ and determine the answer to this question.
(ii) Consider (1.5.16) with a zigzag initial condition $u_0(x) = u(0, x)$:

$$u_0(x) = \begin{cases} x, & 0 \le x \le 1/2, \\ 1 - x, & 1/2 \le x \le 1, \end{cases} \tag{1.5.17}$$

extended periodically to $\mathbb{R}$. How will the viscosity solution $u(t, x)$ to the Cauchy problem look like? Where will it be smooth, and where will it be just Lipschitz? Hint: it may help to do this in at least two ways: (1) use the definition of the viscosity solution, (2) use the notion of the entropy solution for the Burgers' equation for $v(t, x) = u_x(t, x)$ if you are familiar with the basic theory of one-dimensional conservation laws.

Exercise 1.5.5 *(Intermezzo: a Laplace asymptotics of integrals).* Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a real-valued smooth function such that there are two positive constants α and β such that

$$\varphi(x) \ge \alpha |x|^2 - \beta.$$

For $\varepsilon > 0$, consider the integral

$$I_\varepsilon = \int_{\mathbb{R}^n} e^{-\varphi(x)/\varepsilon} dx.$$

The goal of this exercise is to show that

$$\lim_{\varepsilon \to 0} \big(-\varepsilon \log I_\varepsilon \big) = \min_{x \in \mathbb{R}^n} \varphi(x). \tag{1.5.18}$$

Note that it suffices to assume that

$$\min_{x \in \mathbb{R}^n} \varphi(x) = 0, \tag{1.5.19}$$

and show that

$$\lim_{\varepsilon \to 0} \big(-\varepsilon \log I_\varepsilon \big) = 0. \tag{1.5.20}$$

Exercise 1.5.6 Let us add the term εu_{xx} to the right side of (1.5.16), which produces a solution $u_\varepsilon(t,x)$. Use the Hopf-Cole transformation $z_\varepsilon(t,x) = \exp(u_\varepsilon(t,x)/\varepsilon)$, solve the linear problem for $z(t,x)$ and then pass to the limit $\varepsilon \to 0$ using Exercise 1.5.5. Study what happens when $u_0'(x)$ has limits at $\pm\infty$.

Basic properties of the viscosity solutions

We now describe some basic corollaries of the definition of a viscosity solution.

Exercise 1.5.7 Show that the maximum of two viscosity subsolutions to (1.5.15) is a viscosity subsolution, and the minimum of two viscosity supersolutions is a viscosity supersolution.

Exercise 1.5.8 (Stability) Let $F_j(x,u,p)$ be a sequence of functions in $C(\mathbb{T}^n \times \mathbb{R} \times \mathbb{R}^n)$, which converges locally uniformly to $F \in C(\mathbb{T}^n \times \mathbb{R} \times \mathbb{R}^n)$. Let $u_j(t,x)$ be a sequence of viscosity solutions to (1.5.5) with $F = F_j$:

$$\partial_t u_j + F_j(x, u_j, \nabla u_j) = 0, \quad t > 0, \; x \in \mathbb{T}^n, \tag{1.5.21}$$

and assume that u_j converges locally uniformly to $u \in C([0,+\infty), \mathbb{T}^n)$. Show that then u is a viscosity solution to the limiting problem

$$u_t + F(x, u, \nabla u) = 0, \quad t > 0, \; x \in \mathbb{T}^n, \tag{1.5.22}$$

Hint: if (t_0, x_0) is, for instance, a local minimum of the difference $\phi - u$, one can turn it into a strict minimum by changing $\phi(t,x)$ into $\phi(x) + M((t-t_0)^2 + |x-x_0|^2)$, without changing $\phi_t(t_0,x_0)$ and $\nabla\phi(t_0,x_0)$. In this situation, show that there is a sequence (t_j, x_j) of minima of $\phi - u_j$ converging to (t_0, x_0), and use the fact that each u_j is a viscosity solution to (1.5.21) to conclude.

The above exercise is extremely important: it shows that, somewhat similar to the weak solutions, we do not need to check the convergence of the derivatives of u_j to the derivatives of u to know that u is a viscosity solution – this is an extremely useful property to have. Exercise 1.5.8 asserts that one may safely "pass to the limit" in equation (1.5.5), as soon as estimates on the moduli of continuity of the solutions are available rather than on the derivatives.

The next proposition says that viscosity solutions that are Lispchitz continuous do satisfy the equation in the classical sense almost everywhere.

Proposition 1.5.9 *Let u be a locally Lipschitz viscosity solution to*

$$u_t + F(x, u, \nabla u) = 0, \quad t > 0, \ x \in \mathbb{T}^n. \tag{1.5.23}$$

Then it satisfies (1.5.23) almost everywhere.

This property relies on the following lemma [44].

Lemma 1.5.10 *At a point of differentiability (t_0, x_0) of $u(t, x)$, one may construct a C^1 test function $\phi(t, x)$ such that (t_0, x_0) is a local maximum (respectively, a local minimum) of $\phi - u$.*

Proof. For simplicity, we do not take the t-dependence into account, leaving this to the reader, so we assume that $u(x)$ is a function of x that is differentiable at x_0. Without loss of generality, we assume that $x_0 = 0$, $u(0) = 0$, and that $\nabla u(0) = 0$, so that $u(x)$ satisfies, in the vicinity of $x = 0$:

$$u(x) = |x|\varepsilon(x), \qquad \lim_{|x|\to 0} \varepsilon(x) = 0. \tag{1.5.24}$$

Our goal is to construct a C^1 function $\phi(x)$ such that $\phi(x) \leq u(x)$ and $\phi(0) = 0$. Note that this could be impossible for $u(x)$ that is merely Lipschitz and not differentiable – the simple counterexample is $u(x) = -|x|$. We look for a radially symmetric function $\phi(x)$ in the form $\phi(x) = |x|\zeta(|x|)$ with a C^1-function $\zeta(r)$ such that

$$\zeta(|x|) \leq \varepsilon(x), \quad \lim_{r\to 0} \zeta(r) = 0. \tag{1.5.25}$$

To this end, consider the sequence

$$\varepsilon_n = \inf_{2^{-n-1} \leq |r| < 2^{-n}} \varepsilon(r),$$

and produce the function $\zeta(r) \leq \varepsilon(r)$ by smoothing the piecewise constant function

$$\sum_{n=0}^{+\infty} \varepsilon_n \mathbf{1}_{2^{-n-1} \leq r < 2^{-n}}.$$

We can do this while ensuring that

$$|\zeta'(r)| \leq \frac{|\varepsilon_n - \varepsilon_{n+1}| + |\varepsilon_n - \varepsilon_{n-1}|}{2^n}, \quad \text{if } 2^{-n-1} \leq r \leq 2^{-n}. \tag{1.5.26}$$

As the sequence $\varepsilon_n \to 0$ as $n \to +\infty$, this will ensure that

$$|\zeta'(r)| \leq \frac{\alpha(r)}{r},$$

with $\alpha(r) \to 0$ as $r \to 0$. It follows that $\phi(x) = |x|\zeta(|x|)$ is the sought C^1-function. □

Proof of Proposition 1.5.9. The conclusion of this proposition follows essentially immediately from Lemma 1.5.10 and the Rademacher theorem. The latter says that a Lipschitz function is differentiable a.e., see for instance [57]. At any differentiability point we can construct a C^1-function $\phi(t, x)$ such that the difference $\phi - u$ attains its minimum at (t_0, x_0), so that

$$\phi_t(t_0, x_0) = u_t(t_0, x_0) \text{ and } \nabla\phi(t_0, x_0) = \nabla u(t_0, x_0). \tag{1.5.27}$$

The definition of a viscosity sub-solution together with (1.5.27) implies that

$$u_t(t_0,x_0)+H(x,u(t_0,x_0),\nabla u(t_0,x_0))=\phi_t(t_0,x_0)+H(x,u(t_0,x_0),\nabla\phi(t_0,x_0))\le 0.$$

Similarly, we can show that

$$u_t(t_0,x_0)+H(x,u(t_0,x_0),\nabla u(t_0,x_0))\ge 0,$$

using the fact that $u(t,x)$ is a viscosity super-solution. This finishes the proof. □

Warning. For the rest of this section, a solution to (1.1.1) will always be meant in the viscosity sense.

1.5.2 Uniqueness of the viscosity solutions

Let us first give the simplest uniqueness result, that we will prove by the method of doubling of variables. This argument appears in almost all uniqueness proofs, in more or less elaborate forms.

Proposition 1.5.11 *Assume that the Hamiltonian $H(x,p)$ is continuous in all its variables, and satisfies the coercivity assumption*

$$\lim_{|p|\to+\infty} H(x,p)=+\infty,\quad \text{uniformly in } x\in\mathbb{T}^n. \tag{1.5.28}$$

Consider the equation

$$H(x,\nabla u)+u=0,\qquad x\in\mathbb{T}^n. \tag{1.5.29}$$

Let $\underline{u}$ and $\overline{u}$ be, respectively, a viscosity sub- and a super-solution to (1.5.29), then $\underline{u}\le\overline{u}$.

Proof. Assume for a moment that both $\underline{u}$ and $\overline{u}$ are C^1-functions, so that we can use each of them as a test function in the definition of the viscosity sub- and super-solutions. First, we use the fact that $\overline{u}$ is a super-solution to (1.5.29) and take $\underline{u}$ as a test function. Let x_0 be a maximum of $\underline{u}-\overline{u}$, then we deduce from the definition of a viscosity super-solution to (1.5.29) that

$$H(x_0,\nabla\underline{u}(x_0))+\overline{u}(x_0)\ge 0. \tag{1.5.30}$$

On the other hand, $\overline{u}-\underline{u}$ attains its minimum at the same point x_0, and, as $\underline{u}$ is a viscosity sub-solution to (1.5.29), and $\overline{u}$ can serve as a test function, we have

$$H(x_0,\nabla\overline{u}(x_0))+\underline{u}(x_0)\le 0. \tag{1.5.31}$$

As x_0 is a minimum of $\overline{u}-\underline{u}$, and $\underline{u}$ and $\overline{u}$ are differentiable, we have $\nabla\overline{u}(x_0)=\nabla\underline{u}(x_0)$, whence (1.5.30) and (1.5.31) imply

$$\underline{u}(x_0)\le\overline{u}(x_0).$$

Once again, as $\overline{u}-\underline{u}$ attains its minimum at x_0, we conclude that $\overline{u}(x)\ge\underline{u}(x)$ for all $x\in\mathbb{T}^n$ if both of these functions are in $C^1(\mathbb{T}^n)$. Unfortunately, we only know that $\underline{u}$ and $\overline{u}$ are continuous, so we can not use the elegant argument above unless we know, in addition, that they are both C^1-functions.

In the general case, the method of doubling the variables gives a way around the technical difficulty that $\underline{u}(x)$ and $\overline{u}(x)$ are merely continuous and not necessarily differentiable. Let us define, for $\varepsilon > 0$, the penalization

$$u_\varepsilon(x,y) = \overline{u}(x) - \underline{u}(y) + \frac{|x-y|^2}{2\varepsilon^2}$$

and let $(x_\varepsilon, y_\varepsilon) \in \mathbb{T}^{2n}$ be a minimum for $u_\varepsilon(x,y)$ over $\mathbb{T}^{2n}$.

Exercise 1.5.12 Show that

$$\lim_{\varepsilon\to 0} |x_\varepsilon - y_\varepsilon| = 0. \tag{1.5.32}$$

and that the family $(x_\varepsilon, y_\varepsilon)$ converges, as $\varepsilon \to 0$, up to a subsequence, to a point (x_0, x_0), where x_0 is a minimum for $\overline{u}(x) - \underline{u}(x)$.

Consider the function

$$\phi(x) = \underline{u}(y_\varepsilon) - \frac{|x-y_\varepsilon|^2}{2\varepsilon^2},$$

as a (smooth) quadratic function of the variable x. The difference

$$\phi(x) - \overline{u}(x) = -u_\varepsilon(x, y_\varepsilon)$$

attains its maximum, as a function of x, at the point $x = x_\varepsilon$. As $\overline{u}(x)$ is a viscosity super-solution to (1.5.29), we have

$$H(x_\varepsilon, \frac{y_\varepsilon - x_\varepsilon}{\varepsilon^2}) + \overline{u}(x_\varepsilon) \geq 0. \tag{1.5.33}$$

Next, we apply the viscosity sub-solution part of Definition 1.5.13 to the quadratic test function

$$\psi(y) = \overline{u}(x_\varepsilon) + \frac{|x_\varepsilon - y|^2}{2\varepsilon^2}.$$

The difference

$$\psi(y) - \underline{u}(y) = \overline{u}(x_\varepsilon) + \frac{|x_\varepsilon - y|^2}{2\varepsilon^2} - \underline{u}(y) = u_\varepsilon(x_\varepsilon, y)$$

attains its minimum at $y = y_\varepsilon$. As $\underline{u}(y)$ is a viscosity sub-solution to (1.5.29), we obtain

$$H(y_\varepsilon, \frac{y_\varepsilon - x_\varepsilon}{\varepsilon^2}) + \underline{u}(y_\varepsilon) \leq 0. \tag{1.5.34}$$

The coercivity of the Hamiltonian and (1.5.34), together with the boundedness of $\underline{u}_\varepsilon$, imply that $|x_\varepsilon - y_\varepsilon|/\varepsilon^2$ is bounded, uniformly in ε: there exists R so that

$$\frac{|x_\varepsilon - y_\varepsilon|}{\varepsilon^2} \leq R.$$

The uniform continuity of $H(x,p)$ on the set $\{(x,p) : \; x \in \mathbb{T}^n, p \in B(0,R)\}$ implies that, as consequence, we have

$$H(y_\varepsilon, \frac{y_\varepsilon - x_\varepsilon}{\varepsilon^2}) = H(x_\varepsilon, \frac{y_\varepsilon - x_\varepsilon}{\varepsilon^2}) + o(1), \text{ as } \varepsilon \to 0.$$

Subtracting (1.5.34) from (1.5.33), we obtain

$$\overline{u}(x_\varepsilon) - \underline{u}(y_\varepsilon) \geq o(1), \text{ as } \varepsilon \to 0.$$

Sending $\varepsilon \to 0$ with the help of the result of Exercise 1.5.12 implies

$$\overline{u}(x_0) - \underline{u}(x_0) \geq 0,$$

and, as x_0 is the minimum of $\overline{u} - \underline{u}$, the proof is complete. □

An immediate consequence of Proposition 1.5.11 is that (1.5.29) has at most one solution.

The comparison principle and weak contraction

The proof of Proposition 1.5.11 can be adapted to establish two fundamental properties for the viscosity solutions to the Cauchy problem: the comparison principle and the weak contraction property.

Exercise 1.5.13 (The comparison principle) Assume that $H(x,p)$, is a continuous function that satisfies the coercivity property (1.5.28). Let $u_1(t,x)$ and $u_2(t,x)$ be, respectively, a viscosity sub-solution, and a viscosity super-solution to

$$u_t + H(x, \nabla u) = 0, \quad t > 0, \ x \in \mathbb{T}^n, \tag{1.5.35}$$

with the initial conditions u_{10} and u_{20} such that $u_{10}(x) \leq u_{20}(x)$ for all $x \in \mathbb{T}^n$. Modify the proof of Proposition 1.5.11 to show that then $u_1(t,x) \leq u_2(t,x)$ for all $t \geq 0$ and $x \in \mathbb{T}^n$. This proves the uniqueness of the viscosity solutions.

Exercise 1.5.14 (Weak contraction) Let $H(x,p)$ be a continuous function that satisfies the coercivity property (1.5.28), and u_1 and u_2 be two solutions to (1.5.35) with continuous initial conditions u_{10} and u_{20}, respectively. Show that then we have

$$\|u_1(t,\cdot) - u_2(t,\cdot)\|_{L^\infty} \leq \|u_{10} - u_{20}\|_{L^\infty}.$$

Hint: notice that if $u(t,x)$ solves (1.5.35) then so does $u(t,x) + k$ for any $k \in \mathbb{R}$, and use Exercise 1.5.13.

1.5.3 Finite speed of propagation

We are now going to prove a finite speed of propagation property, partly to acquire some further familiarity with the notion of a viscosity solution, and partly to emphasize the sharp contrast with viscous models: if the equation carried a Laplacian, an initially nonnegative solution would instantly become positive everywhere. As this property makes better sense in $\mathbb{R}^n$ and not on the torus, this is the case we will consider.

Proposition 1.5.15 *Let H be uniformly Lipschitz with respect to its second variable:*

$$|H(x,p_1) - H(x,p_2)| \leq C_L|p_1 - p_2| \quad \textit{for all } x \in \mathbb{R}^n \textit{ and } p_1, p_2 \in \mathbb{R}^n. \tag{1.5.36}$$

Let u_0 and v_0 be two continuous, compactly supported initial conditions, and assume that each generates a globally Lipschitz solution, respectively denoted by $u(t,x)$ and $v(t,x)$ to the Cauchy problem

$$u_t + H(x, \nabla u) = 0, \quad v_t + H(x, \nabla v) = 0, \quad 0 < t \le T, \quad x \in \mathbb{R}^n, \tag{1.5.37}$$

with $u(0,x) = u_0(x)$ and $v(0,x) = v_0(x)$ for all $x \in \mathbb{R}^n$. Then, if $x_0 \in \mathbb{R}^n$ and $t_0 \in [0,T]$ satisfy

$$\operatorname{dist}\big(x_0, \operatorname{supp}(u_0 - v_0))\big) > t_0 C_L,$$

then $u(t_0, x_0) = v(t_0, x_0)$.

Proof. The idea is simple: assuming that everything is smooth, the function $w = u - v$ satisfies the inequalities

$$w_t \le C_L |\nabla w|, \tag{1.5.38}$$

and

$$w_t \ge -C_L |\nabla w|. \tag{1.5.39}$$

Exercise 1.5.16 Use the method of characteristics to show that if w is a smooth function that satisfies (1.5.38) and

$$\operatorname{dist}\big(x_0, \operatorname{supp}(w(0,\cdot))\big) > C_L t_0, \tag{1.5.40}$$

then $w(t_0, x_0) \le 0$, and if a smooth function w satisfies (1.5.39)-(1.5.40), then $w(t_0, x_0) \ge 0$.

Thus, the conclusion of this proposition follows from Exercise 1.5.16 if u and v are smooth. Unfortunately, if u and v are not smooth, then we can not use the characteristics but only the definition of a viscosity solution. Let us fix a point $x_0 \in \mathbb{R}^n$ and $t_0 > 0$ so that

$$\operatorname{dist}\big(x_0, \operatorname{supp}(u_0 - v_0))\big) > C_L t_0, \tag{1.5.41}$$

take $\varepsilon > 0$ sufficiently small, so that

$$\varepsilon < \frac{1}{2}\big(\operatorname{dist}\big(x_0, \operatorname{supp}(u_0 - v_0))\big) - C_L t_0\big), \tag{1.5.42}$$

and let $\phi_0(r)$ be a C^1-function equal to 1 outside of the the ball $B_{C_L t_0 + \varepsilon}(0)$, and to 0 in the ball $B_{C_L t_0}(0)$. The function

$$\overline{w}(t,x) = \|u_0 - v_0\|_{L^\infty} \phi_0(|x - x_0| + C_L t) \tag{1.5.43}$$

is a smooth solution to

$$\partial_t \overline{w} - C_L |\nabla \overline{w}| = 0, \quad t > 0, \ x \in \mathbb{R}^n, \tag{1.5.44}$$

such that $\overline{w}(t, x_0) = 0$ for $t \le t_0$. Moreover, because of (1.5.42), at $t = 0$ we have

$$\overline{w}(0,x) = \|u_0 - v_0\|_{L^\infty} \phi_0(|x - x_0|) \ge |u_0(x) - v_0(x)| \text{ for all } x \in \mathbb{R}^n. \tag{1.5.45}$$

Our goal is to show this inequality persists until the time t_0:

$$|u(t,x) - v(t,x)| \le \overline{w}(t,x) \text{ for all } 0 \le t \le t_0 \text{ and } x \in \mathbb{R}^n. \tag{1.5.46}$$

Indeed, using (1.5.46) at $x = x_0$ and $t = t_0$ would give

$$|u(t_0, x_0) - v(t_0, x_0)| \leq \|u_0 - v_0\|_{L^\infty} \phi_0(C_L t_0) = 0, \tag{1.5.47}$$

which is what we need.

The comparison principle in Exercise 1.5.13 together with (1.5.45) implies that (1.5.46) would follow if we show that $\overline{v}(t, x) = u(t, x) + \overline{w}(t, x)$ is a viscosity super-solution to (1.5.37):

$$\partial_t \overline{v} + H(x, \nabla \overline{v}) \geq 0. \tag{1.5.48}$$

Observe that (1.5.48) and (1.5.45) together would imply

$$v(t, x) \leq \overline{v}(t, x) = u(t, x) + \overline{w}(t, x) \text{ for all } 0 \leq t \leq t_0 \text{ and } x \in \mathbb{R}^n. \tag{1.5.49}$$

As the roles of u and v can be reversed, we would deduce (1.5.46).

Thus, we need to prove the viscosity super-solution property for $\overline{v}(t, x)$. Let $\varphi(t, x)$ be a smooth test function, and (t_1, x_1) be a minimum point for

$$\overline{v}(t, x) - \varphi(t, x) = u(t, x) + \overline{w}(t, x) - \varphi(t, x) = u(t, x) - \psi(t, x), \tag{1.5.50}$$

with a C^1-function

$$\psi(t, x) = \varphi(t, x) - \overline{w}(t, x).$$

In other words, (t_1, x_1) is a minimum point for $u(t, x) - \psi(t, x)$. As u is a viscosity solution to (1.5.37), it follows that

$$\partial_t \psi(t_1, x_1) + H(x_1, \nabla \psi(t_1, x_1)) \geq 0, \tag{1.5.51}$$

which is nothing but

$$\partial_t \varphi(t_1, x_1) - \partial_t \overline{w}(t_1, x_1) + H\big(x_1, \nabla \varphi(t_1, x_1) - \nabla \overline{w}(t_1, x_1)\big) \geq 0, \tag{1.5.52}$$

Using the inequality

$$H(\bar{x}, \nabla \varphi - \nabla \overline{w}) \leq H(\bar{x}, \nabla \varphi) + C_L |\nabla \overline{w}|.$$

in (1.5.52) gives

$$\partial_t \varphi(t_1, x_1) - \partial_t \overline{w}(t_1, x_1) + H\big(x_1, \nabla \varphi(t_1, x_1)\big) + C_L |\nabla \overline{w}(t_1, x_1)| \geq 0. \tag{1.5.53}$$

Recalling (1.5.44), we obtain

$$\partial_t \varphi(t_1, x_1) + H\big(x_1, \nabla \varphi(t_1, x_1)\big) \geq 0. \tag{1.5.54}$$

We conclude that $\overline{v}(t, x)$ is a viscosity super-solution to (1.5.37), finishing the proof. □

Exercise 1.5.17 *(Hole filling property).* Let $u(t, x)$ be a viscosity solution to

$$u_t = R(t, x)|\nabla u|, \quad t > 0, \ x \in \mathbb{R}^n,$$

with $R(t, x) \geq R_0 > 0$. Assume that (i) $u(0, x) = u_0(x) \geq \delta_0 > 0$ outside a ball $B(0, R)$, and (ii) the set $\mathbb{R}^n \backslash \big(\mathrm{supp}(u_0)\big)$ is compact. Prove that there exists $T_0 > 0$ such that $u(t, x) > 0$ for all $t \geq T_0$, and all $x \in \mathbb{R}^n$.

1.6 Construction of solutions

So far, we have set up a beautiful notion of viscosity solutions, and we have also proved that this works well in settling our worries about uniqueness, distinguishing them from the Lipschitz solutions. Now, we have to prove that, as far as existence is concerned, this new notion does better than the classical solutions, in the sense that solutions to the Cauchy problem exist for all $t > 0$ under reasonable assumptions. In this section, we will show how these solutions can be constructed. First, we will produce wave solutions to the time-dependent problem

$$\partial_t u + H(x, \nabla u) = 0, \quad x \in \mathbb{T}^n. \tag{1.6.1}$$

Next, we are going to prove that the Cauchy problem for (1.6.1) is well-posed as soon as a continuous initial condition is specified. Eventually, we will show that the wave solutions describe the long time behavior of the solutions to the Cauchy problem.

1.6.1 Existence of waves, and the Lions-Papanicolaou-Varadhan theorem

Wave solutions for (1.6.1) will be sought in the same form as viscous waves, that is

$$v(t, x) = -ct + u(x), \tag{1.6.2}$$

with a constant $c \in \mathbb{R}$. A function $v(t, x)$ of this form is a solution to (1.6.1) if $u(x)$ solves a time-independent problem

$$H(x, \nabla u) = c, \quad x \in \mathbb{T}^n. \tag{1.6.3}$$

Exercise 1.6.1 Show that a function $v(t, x)$ of the form (1.6.2) is a viscosity solution to (1.6.1) if and only if $u(x)$ is a viscosity solution to (1.6.3).

We will think of $v(t, x)$ as the height of an interface, and refer to the constant c as the speed of the wave, and to $u(x)$ as its shape. Let us point out that the speed is unique: (1.6.3) may have viscosity solutions for at most one c. Indeed, assume there exist $c_1 \neq c_2$, such that (1.6.3) has a viscosity solution u_1 for $c = c_1$ and another viscosity solution u_2 for $c = c_2$. Let $K > 0$ be such that

$$u_1(x) - K \leq u_2(x) \leq u_1(x) + K, \quad \text{for all } x \in \mathbb{T}^n.$$

By Exercise 1.6.1 the functions

$$v_{1,\pm}(t, x) = -c_1 t + u_1(x) \pm K$$

and

$$v_2(t, x) = -c_2 t + u_2(x)$$

are the viscosity solutions to the Cauchy problem (1.1.1) with the respective initial conditions

$$v_{1,\pm}(x) = u_1(x) \pm K, \quad v_2(0, x) = u_2(x).$$

By the comparison principle (Exercise 1.5.13), we have

$$-c_1 t + u_1(x) - K \le -c_2 t + u_2(x) \le -c_1 t + u_1(x) + K, \quad \text{for all } t \ge 0 \text{ and } x \in \mathbb{T}^n.$$

This is a contradiction since $c_1 \neq c_2$, and the functions u_1 and u_2 are bounded. Therefore, the wave speed c is unique. Note that this does not address the question of uniqueness of the shape $u(x)$ – we leave this question for later.

The main result of this section is the following theorem, due to Lions, Papanicolaou, Varadhan [94], that asserts the existence of a constant c for which (1.6.3) has a solution.

Theorem 1.6.2 *Assume that $H(x,p)$ is continuous, uniformly Lipschitz:*

$$|H(x,p_1) - H(x,p_2)| \le C_L |p_1 - p_2|, \text{ for all } x \in \mathbb{T}^n, \text{ and } p_1, p_2 \in \mathbb{R}^n, \tag{1.6.4}$$

the coercivity condition

$$\lim_{|p| \to +\infty} H(x,p) = +\infty, \quad \text{uniformly in } x \in \mathbb{T}^n. \tag{1.6.5}$$

holds, and

$$|\nabla_x H(x,p)| \le K_0(1 + |p|), \text{ for all } x \in \mathbb{T}^n, \text{ and } p \in \mathbb{R}^n. \tag{1.6.6}$$

Then there is a unique $c \in \mathbb{R}$ for which

$$H(x, \nabla u) = c, \quad x \in \mathbb{T}^n. \tag{1.6.7}$$

has a solution.

It is important to point out that the periodicity assumption in x on the Hamiltonian is indispensable – for instance, when $H(x,p)$ is a random function (in x) on $\mathbb{R}^n \times \mathbb{R}^n$, the situation is much more complicated – an interested reader should consult the literature on stochastic homogenization of the Hamilton-Jacobi equations, a research area that is active and evolving at the moment of this writing. We also mention that the only assumption made in [94] is that $H(x,p)$ is continuous and coercive. The Lipschitz condition (1.6.4) in p and (1.6.6) in x have been added here for convenience.

The homogenization connection

Before proceeding with the proof of the Lions-Papanicolaou-Varadhan theorem, let us explain how the steady equation (1.6.7) appears in the context of periodic homogenization, which was probably the main motivation behind this theorem. We can not possibly do justice to the area of homogenization here – an interested reader should explore the huge literature on the subject, with the book [115] by G. Pavliotis and A. Stuart providing a good starting point. Let us just briefly illustrate the general setting on the example of the periodic Hamilton-Jacobi equations. Consider the Cauchy problem

$$u^\varepsilon_t + H(x, \nabla u^\varepsilon) = 0, \quad t > 0, \ x \in \mathbb{R}^n, \tag{1.6.8}$$

in the whole space $\mathbb{R}^n$ (and not on the torus), with the Hamiltonian $H(x,p)$ that is 1-periodic in all coordinates x_j, $j = 1, \dots, n$. We are interested in the regime where the initial condition is slowly varying and large:

$$u^\varepsilon(0,x) = \varepsilon^{-1} u_0(\varepsilon x). \tag{1.6.9}$$

Let us note that if one thinks of the solution to (1.6.8) as the height of some interface at the position $x \in \mathbb{R}^n$ at a time $t > 0$, then it is very natural that if $u^\varepsilon(0,x)$ varies on a scale ε^{-1} in the x-variable, then its height should also be of the order ε^{-1}, which is exactly what we see in (1.6.9).

The general question of homogenization is how the "microscopic" variations in the Hamiltonian that varies on the scale $O(1)$ affect the evolution of the initial condition that varies on the "macroscopic" scale $O(\varepsilon^{-1})$. The goal is to describe the evolution in purely "macroscopic" terms, and extract an effective macroscopic problem that approximates the full microscopic problem well. This allows to avoid, say, in numerical simulations, modeling the microscopic variations of the Hamiltonian, and do the simulations on the macroscopic scale – a huge advantage in engineering problems. It also happens that from the purely mathematical view point, homogenization is also an extremely rich subject.

This general philosophy translates into the following strategy. As the initial condition in (1.6.9) is slowly varying, one should observe the solution on a macroscopic spatial scale, in the "slow" variable $y = \varepsilon x$. Since $u^\varepsilon(0,x)$ is also very large itself, of the size $O(\varepsilon^{-1})$, it is appropriate to rescale it down. In other words, instead of looking at $u^\varepsilon(t,x)$ directly, we would represent it as

$$u^\varepsilon(t,x) = \varepsilon^{-1} w^\varepsilon(t,\varepsilon x),$$

and consider the evolution of $w^\varepsilon(t,y)$, which satisfies

$$w^\varepsilon_t + \varepsilon H(\frac{y}{\varepsilon}, \nabla w^\varepsilon) = 0, \tag{1.6.10}$$

with the initial condition $w^\varepsilon(0,y) = u_0(y)$ that is now independent of ε. However, we see that w^ε evolves very slowly in t – its time derivative is of the size $O(\varepsilon)$. Hence, we need to wait a long time until it changes. To remedy this, we introduce a long time scale of the size $t = O(\varepsilon^{-1})$. In other words, we write

$$w^\varepsilon(t,y) = v^\varepsilon(\varepsilon t, y).$$

In the new variables the problem takes the form

$$v^\varepsilon_s + H\Big(\frac{y}{\varepsilon}, \nabla v^\varepsilon\Big) = 0, \quad y \in \mathbb{R}^n, \quad s > 0, \tag{1.6.11}$$

with the initial condition $v^\varepsilon(0,y) = u_0(y)$.

It seems that we have merely shifted the difficulty – we used to have ε in the initial condition in (1.6.9) while now we have it appear in the equation itself – the Hamiltonian depends on y/ε. However, it turns out that we may now find an ε-independent problem that has a spatially uniform Hamiltonian that provides a good approximation to (1.6.11). The reason this is possible is that we have chosen the correct temporal and spatial scales to track the evolution of the solution.

Here is an informal way to find the approximating problem. Let us seek the solution to (1.6.11) in the form of an asymptotic expansion

$$v^\varepsilon(s,y) = \bar{v}(s,y) + \varepsilon v_1(s,y,\frac{y}{\varepsilon}) + \varepsilon^2 v_2(s,y,\frac{y}{\varepsilon}) + \dots \tag{1.6.12}$$

The functions $v_j(s, y, z)$ are assumed to be periodic in the "fast" variable z but not in the "slow" variables s and y. Inserting this expansion into (1.6.11), and collecting the terms with various powers of ε, we obtain in the leading order

$$\bar{v}_s(s, y) + H\Big(\frac{y}{\varepsilon}, \nabla_y \bar{v}(s, y) + \nabla_z v_1(s, y, \frac{y}{\varepsilon})\Big) = 0. \tag{1.6.13}$$

As is standard in such multiple scale expansions, we consider (1.6.13) as

$$\bar{v}_s(s, y) + H(z, \nabla_y \bar{v}(s, y) + \nabla_z v_1(s, y, z)) = 0, \quad z \in \mathbb{T}^n, \tag{1.6.14}$$

an equation for v_1 as a function of the fast variable $z \in \mathbb{T}^n$, for each $s > 0$ and $y \in \mathbb{R}^n$ fixed. In other words, for each pair of the "macroscopic" variables s and y we consider a microscopic problem in the z-variable. In the area of numerical analysis, one would call this "sub-grid modeling": the variables s and y live on the macroscopic grid, and the z-variable lives on the microscopic sub-grid.

The function $\bar{v}(s, y)$ will then be found from the solvability condition for (1.6.13). Indeed, the terms $\bar{v}_s(s, y)$ and $\nabla_y \bar{v}(s, y)$ in (1.6.14) do not depend on the fast variable z and should be treated as constants – we solve (1.6.14) independently for each s and y. Let us then, for each fixed $p \in \mathbb{R}^n$, consider the problem

$$H(z, p + \nabla_z w) = c, \quad z \in \mathbb{T}^n. \tag{1.6.15}$$

The case of interest is $p = \nabla_y \bar{v}(s, y)$ and $c = -\bar{v}_s(s, y)$ but one needs to momentarily look at (1.6.15) for an arbitrary choice of $p \in \mathbb{R}^n$ and $c \in \mathbb{R}$. The Lions-Papanicolaou-Varadhan theorem says that for each $p \in \mathbb{R}^n$ there is a unique c that we will denote by $\bar{H}(p)$ such that (1.6.15) has a solution. We then write (1.6.15) as

$$H(z, p + \nabla_z w) = \bar{H}(p), \quad z \in \mathbb{T}^n. \tag{1.6.16}$$

Hence, the solvability condition for (1.6.14) is that the function $\bar{v}(s, y)$ satisfies the homogenized (also known as "effective") equation

$$\bar{v}_s + \bar{H}(\nabla_y \bar{v}) = 0, \quad \bar{v}(0, y) = u_0(y), \quad s > 0, \ y \in \mathbb{R}^n, \tag{1.6.17}$$

and the function $\bar{H}(p)$ is called the effective, or homogenized Hamiltonian. Note that the effective Hamiltonian does not depend on the spatial variable – the "small scale" variations are averaged out via the above homogenization procedure. The point is that the solution $v^\varepsilon(s, y)$ of (1.6.11), an equation with highly oscillatory coefficients is well approximated by $\bar{v}(s, y)$, the solution of (1.6.17), an equation with spatially uniform coefficients, that is much simpler to study analytically or solve numerically.

Thus, the existence and uniqueness of the constant c for which solution of the steady equation (1.6.15) exists, is directly related to the homogenization (long time behavior) of the solutions to the Cauchy problem (1.6.8) with slowly varying initial conditions, as it provides the corresponding effective Hamiltonian. Unfortunately, there is a catch: not so much is known in general on how the effective Hamiltonian $\bar{H}(p)$ depends on the original Hamiltonian $H(x, p)$, except for some very generic properties. Estimating and computing numerically the effective Hamiltonian $\bar{H}(p)$ is a separate interesting line of research.

Exercise 1.6.3 *(The one-dimensional case)* Compute the effective Hamiltonian $\bar{H}(p)$ for

$$H(x,p) = R(x)\sqrt{1+p^2}, \quad x \in \mathbb{T}^1, p \in \mathbb{R},$$

where $R(x)$ is a smooth 1-periodic function.

Exercise 1.6.4 Show that for every $p \in \mathbb{R}^n$ one can find a periodic in x function $u(x;p)$, $x \in \mathbb{T}^n$, $p \in \mathbb{R}^n$ such that the function

$$v(t,x;p) = p \cdot x + u(x;p) - t\bar{H}(p)$$

is a solution to

$$v_t + H(x; \nabla v) = 0.$$

What is the function $u(x;p)$ in terms of the approximate expansion (1.6.12)? Explain why it is natural that the function $u(x;p)$ appears when we try to approximate the solution to

$$u_t^\varepsilon + H(x, \nabla u^\varepsilon) = 0,$$

with an initial condition of the form $u^\varepsilon(0,x) = \varepsilon^{-1}u_0(\varepsilon x)$.

The proof of the Lions-Papanicolaou-Varadhan theorem

Recall that our goal is to construct a solution to (1.6.7):

$$H(x, \nabla u) = c, \quad x \in \mathbb{T}^n. \tag{1.6.18}$$

As we have already proved uniqueness of the constant c, we only need to prove its existence, and, of course, construct the solution $u(x)$. We will make use of the viscosity solution to the auxiliary problem

$$H(x, \nabla u^\varepsilon) + \varepsilon u^\varepsilon = 0, \quad x \in \mathbb{T}^n, \tag{1.6.19}$$

with $\varepsilon > 0$. Note that the regularization parameter $\varepsilon > 0$ in (1.6.19) has nothing to do with the small parameter $\varepsilon > 0$ that we have used in the discussion of the periodic homogenization theory, where it referred to the separation of scales between the scale of variation of the initial condition and that of the periodic Hamiltonian. Unfortunately, it is common to use the notation ε in both of these settings. We hope that the reader will find it not too confusing.

We have already shown that (1.6.19) has at most one solution. Let us for the moment accept that the solution to the regularized problem (1.6.19) exists and show how one can finish the proof of Theorem 1.6.2 from here. Then, we will come back to the construction of a solution to (1.6.19). Our task is to pass to the limit $\varepsilon \downarrow 0$ in (1.6.19).

Exercise 1.6.5 Show that for all $\varepsilon > 0$, the solution $u^\varepsilon(x)$ of (1.6.19) satisfies

$$-\frac{\|H(\cdot,0)\|_{L^\infty}}{\varepsilon} \le u^\varepsilon(x) \le \frac{\|H(\cdot,0)\|_{L^\infty}}{\varepsilon}, \tag{1.6.20}$$

for all $x \in \mathbb{T}^n$. Hint: use the comparison principle.

Note that the fact that $u^\varepsilon(x)$ is of the size ε^{-1} is not a fluke of the estimate. For instance, if the function $H(x,p)$ is bounded from below by a positive constant c_0, then the solution to (1.6.19) will clearly satisfy $|u^\varepsilon(x)| \geq c_0/\varepsilon$ for all $x \in \mathbb{T}^n$. Therefore, one can not expect that the solution to (1.6.19) converges as $\varepsilon \to 0$ to a solution to (1.6.18). One can, however, hope that the solution becomes large but its gradient stays bounded, so if we subtract the large mean the difference will be bounded. Accordingly, we will decompose u^ε into its mean and oscillation:

$$u^\varepsilon(x) = \langle u^\varepsilon \rangle + v^\varepsilon(x), \tag{1.6.21}$$

where

$$\langle u^\varepsilon \rangle = \int_{\mathbb{T}^n} u^\varepsilon(y)dy. \tag{1.6.22}$$

Recall that the torus $\mathbb{T}^n$ is normalized so that $\text{Vol}(\mathbb{T}^n) = 1$. We will then show that there is a sequence $\varepsilon_k \to 0$ so that the limit

$$c = -\lim_{\varepsilon_k \to 0} \varepsilon_k \langle u^{\varepsilon_k} \rangle \tag{1.6.23}$$

exists, and $v^{\varepsilon_k}(x)$ also converge uniformly on $\mathbb{T}^n$ to a limit u that satisfies (1.6.18) with c given by (1.6.23).

In order to pass to the limit $\varepsilon \downarrow 0$ in (1.6.19), we need a modulus of continuity estimate on u^ε (and hence v^ε) that does not depend on $\varepsilon \in (0,1)$.

Lemma 1.6.6 *There is $C > 0$ independent of ε such that $|\text{Lip } u^\varepsilon| \leq C$.*

Proof. Again, we use the doubling of the independent variables. Fix $x \in \mathbb{T}^n$ and, for $K > 0$, consider the function

$$\zeta(y) = u^\varepsilon(y) - u^\varepsilon(x) - K|y - x|. \tag{1.6.24}$$

Let $\hat{x}$ be a maximum of $\zeta(y)$ (the point $\hat{x}$ depends on x). If $\hat{x} = x$ for all $x \in \mathbb{T}^n$, then, as $\zeta(x) = 0$, we obtain

$$u^\varepsilon(y) - u^\varepsilon(x) \leq K|x - y|, \tag{1.6.25}$$

for all $x, y \in \mathbb{T}^n$, which implies that u^ε is Lipschitz with the constant K. If there exists some x such that $\hat{x} \neq x$, then the function

$$\psi(y) = u^\varepsilon(x) + K|y - x|$$

is, in a vicinity of the point $y = \hat{x}$, an admissible test function, as a function of y. Moreover, the difference

$$\psi(y) - u^\varepsilon(y) = -\zeta(y)$$

attains its minimum at $y = \hat{x}$. As $u^\varepsilon(y)$ is a viscosity solution to (1.6.19), and

$$\nabla\psi(\hat{x}) = K\frac{\hat{x} - x}{|\hat{x} - x|},$$

it follows that

$$H\Big(\hat{x}, K\frac{\hat{x} - x}{|\hat{x} - x|}\Big) + \varepsilon u^\varepsilon(\hat{x}) \leq 0. \tag{1.6.26}$$

Since $\varepsilon u^\varepsilon(x)$ is bounded by $\|H(\cdot,0)\|_{L^\infty}$, as in (1.6.20), we deduce that

$$H\Big(\hat{x}, K\frac{\hat{x}-x}{|\hat{x}-x|}\Big) \le \|H(\cdot,0)\|_{L^\infty}. \tag{1.6.27}$$

On the other hand, the coercivity condition (1.6.5) implies that we can take K sufficiently large, so that

$$\|H(\cdot,0)\|_{L^\infty} < \inf_{x\in\mathbb{T}^n, |p|=K} H(x,p). \tag{1.6.28}$$

Hence, if we take K as in (1.6.28), then (1.6.27) can not hold. As a consequence, for such K we must have $\hat{x} = x$ for all $x \in \mathbb{T}^n$. It follows that for such K the inequality (1.6.25) holds for all $x, y \in \mathbb{T}^n$. This finishes the proof. □

To finish the proof of Theorem 1.6.2, we go back to the decomposition (1.6.21)-(1.6.22). The function

$$v^\varepsilon = u^\varepsilon - \langle u^\varepsilon\rangle$$

satisfies

$$H(x,\nabla v^\varepsilon) + \varepsilon\langle u^\varepsilon\rangle + \varepsilon v^\varepsilon = 0. \tag{1.6.29}$$

As

$$\int_{\mathbb{T}^n} v^\varepsilon(x)dx = 0,$$

and because of Lemma 1.6.6, the family v^ε is both uniformly bounded in L^∞ and is uniformly Lipschitz. As a consequence, it converges uniformly, up to extraction of a subsequence, to a function $v \in C(\mathbb{T}^n)$, and $\varepsilon v^\varepsilon \to 0$. The bound (1.6.20) implies that the family $\varepsilon\langle u^\varepsilon\rangle$ is bounded. We may, therefore, assume its convergence (along a subsequence) to a constant denoted by $-c$, as in (1.6.23). By the stability result in Exercise 1.5.8, we deduce that v is a viscosity solution of

$$H(x,\nabla v) = c. \tag{1.6.30}$$

This finishes the proof of Theorem 1.6.2 except for the construction of a solution to (1.6.19).

Existence of the solution to the auxiliary problem

Let us now construct a solution to (1.6.19).

Proposition 1.6.7 *If $H(x,p)$ satisfies the assumptions of Theorem 1.6.2, then for all $\varepsilon > 0$ the problem*

$$H(x,\nabla u) + \varepsilon u = 0, \quad x \in \mathbb{T}^n, \tag{1.6.31}$$

has a viscosity solution.

We will treat a solution to (1.6.31) as a fixed point of the map $\mathcal{S}[v] = u$ defined via

$$H(x,\nabla u) + Mu = (M-\varepsilon)v, \quad x \in \mathbb{T}^n, \tag{1.6.32}$$

with $M > 0$ to be chosen appropriately. The point is that if M is sufficiently large, we will be able to prove that this map is a contraction on $C(\mathbb{T}^n)$, hence has a fixed point. Any such fixed point is a solution to (1.6.31). Our first task is to prove the following lemma.

Lemma 1.6.8 *There exists $M_0 > 0$ so that for all $M > M_0$ and all $f \in C(\mathbb{T}^n)$ there exists a solution to*

$$H(x, \nabla u) + Mu = f, \quad x \in \mathbb{T}^n. \tag{1.6.33}$$

This lemma shows that the map $\mathcal{S}$ is well-defined for $M > M_0$. Its proof will use an explicit construction of the solutions via a limiting procedure that will give us sufficiently strong a priori bounds that will allow us to deduce that $\mathcal{S}$ is a contraction.

The proof of Lemma 1.6.8

We take a function $f \in C(\mathbb{T}^n)$, and consider a regularized problem

$$-\delta\Delta u^{\gamma,\delta} + H(x, \nabla u^{\gamma,\delta}) + Mu^{\gamma,\delta} = f_\gamma(x), \quad x \in \mathbb{T}^n, \tag{1.6.34}$$

with $\delta > 0$ and $\gamma > 0$, and

$$f_\gamma = G_\gamma \star f. \tag{1.6.35}$$

Here, G_γ is a compactly supported smooth approximation of identity:

$$G_\gamma(x) = \gamma^{-n} G\Big(\frac{x}{\gamma}\Big), \quad G(x) \geq 0, \quad \int_{\mathbb{R}^n} G(x)dx = 1,$$

so that $f_\gamma(x)$ is smooth, and $f_\gamma \to f$ in $C(\mathbb{T}^n)$. In particular, there exists K_γ that depends on $\gamma \in (0, 1)$ so that

$$\|f_\gamma\|_{L^\infty} \leq \|f\|_{L^\infty}, \quad \|f_\gamma\|_{C^1} \leq K_\gamma \|f\|_{L^\infty}. \tag{1.6.36}$$

It is relatively straightforward to show that (1.6.34) admits a smooth solution $u^{\gamma,\delta}$ for each $\gamma > 0$ and $\delta > 0$. The difficulty is to pass to the limit $\delta \downarrow 0$, followed by $\gamma \downarrow 0$ to construct in the limit a viscosity solution to (1.6.33). This will require a priori bounds on $u^{\gamma,\delta}$ summarized in the following lemma.

Lemma 1.6.9 *There exists $M_0 > 0$ so that if $M > M_0$ then the solution $u^{\gamma,\delta}$ to (1.6.34) obeys the following gradient bound, for all $\delta \in (0, 1)$:*

$$|\nabla u^{\gamma,\delta}(x)| \leq C_\gamma(1 + \|f\|_{L^\infty}) \text{ for all } x \in \mathbb{T}^n. \tag{1.6.37}$$

Here, the constant C_γ may depend on $\gamma \in (0, 1)$ but not on $\delta \in (0, 1)$. There also exists a constant $C > 0$ that does not depend on $\gamma \in (0, 1)$ or $\delta \in (0, 1)$ so that

$$|u^{\gamma,\delta}(x)| \leq \frac{C}{M}(1 + \|f\|_{L^\infty}) \text{ for all } x \in \mathbb{T}^n. \tag{1.6.38}$$

Proof. Let us look at the point x_0 where $|\nabla u^{\gamma,\delta}(x)|^2$ attains its maximum. Note that (we drop the super-scripts γ and δ for the moment)

$$\frac{\partial}{\partial x_i}(|\nabla u|^2) = 2\frac{\partial u}{\partial x_j}\frac{\partial^2 u}{\partial x_i \partial x_j},$$

so that, using (1.6.34), we compute

$$\Delta(|\nabla u|^2) = 2\sum_{i,j=1}^n \Big(\frac{\partial^2 u}{\partial x_i \partial x_j}\Big)^2 + 2\sum_{j=1}^n \frac{\partial u}{\partial x_j}\frac{\partial \Delta u}{\partial x_j} = 2\sum_{i,j=1}^n \Big(\frac{\partial^2 u}{\partial x_i \partial x_j}\Big)^2 + \frac{2M}{\delta}|\nabla u|^2$$
$$+\frac{2}{\delta}\sum_{j=1}^n \frac{\partial u}{\partial x_j}\frac{\partial H(x,\nabla u)}{\partial x_j} + \frac{2}{\delta}\sum_{k,j=1}^n \frac{\partial u}{\partial x_j}\frac{\partial H(x,\nabla u)}{\partial p_k}\frac{\partial^2 u}{\partial x_j \partial x_k} - \frac{2}{\delta}\sum_{j=1}^n \frac{\partial u}{\partial x_j}\frac{\partial f_\gamma}{\partial x_j}$$
$$= 2\sum_{i,j=1}^n \Big(\frac{\partial^2 u}{\partial x_i \partial x_j}\Big)^2 + \frac{2M}{\delta}|\nabla u|^2 + \frac{2}{\delta}\sum_{j=1}^n \frac{\partial u}{\partial x_j}\frac{\partial H(x,\nabla u)}{\partial x_j} + \frac{1}{\delta}\sum_{k=1}^n \frac{\partial H(x,\nabla u)}{\partial p_k}\frac{\partial |\nabla u|^2}{\partial x_k}$$
$$-\frac{2}{\delta}\sum_{j=1}^n \frac{\partial u}{\partial x_j}\frac{\partial f_\gamma}{\partial x_j}.$$

Thus, at the maximum x_0 of $|\nabla u|^2$ we have

$$0 \ge \Delta(|\nabla u|^2)(x_0) = 2\sum_{i,j=1}^n \Big(\frac{\partial^2 u}{\partial x_i \partial x_j}\Big)^2 + \frac{2M}{\delta}|\nabla u|^2 + \frac{2}{\delta}\sum_{j=1}^n \frac{\partial u}{\partial x_j}\frac{\partial H(x,\nabla u)}{\partial x_j} - \frac{2}{\delta}\sum_{j=1}^n \frac{\partial u}{\partial x_j}\frac{\partial f_\gamma}{\partial x_j}. \tag{1.6.39}$$

Let us recall the gradient bound (1.6.6) on $H(x,p)$:

$$|\nabla_x H(x,p)| \le K_0(1+|p|). \tag{1.6.40}$$

We see from (1.6.39) and (1.6.40) that

$$Q = |\nabla u(x_0)| = \sup_{x\in\mathbb{T}^n} |\nabla u(x)|$$

satisfies

$$MQ^2 \le K_0 Q(1+Q) + Q\|f_\gamma\|_{C^1} \le 5K_0(1+Q^2) + K_\gamma Q\|f\|_{L^\infty}. \tag{1.6.41}$$

We used (1.6.36) above. It follows from (1.6.41) that there exist $M_0 > 0$ and C_1 that depend on K_0 but not on $\gamma \in (0,1)$ and C_γ that depends on $\gamma \in (0,1)$ so that for all $M > M_0$ we have

$$Q \le C_1 + C_\gamma \|f\|_{L^\infty}. \tag{1.6.42}$$

This proves (1.6.37).

To prove (1.6.38) we look at the point x_M where u attains its maximum over $\mathbb{T}^n$. At this point we have

$$Mu(x_M) = f_\gamma(x_M) + \delta\Delta u(x_M) - H(x_M,0) \le \|f_\gamma\|_{L^\infty} + \|H(\cdot,0)\|_{L^\infty}, \tag{1.6.43}$$

hence

$$u(x_M) \le \frac{C}{M}(1+\|f\|_{L^\infty}).$$

A similar estimate holds at the minimum of u, proving (1.6.38). □

The Lipschitz bound (1.6.37) and (1.6.38) show that if $M > M_0$, after passing to a subsequence $\delta_k \downarrow 0$, the family $u^{\gamma,\delta_k}(x)$ converges uniformly in $x \in \mathbb{T}^n$, to a function $u^\gamma(x)$.

Exercise 1.6.10 Show that $u^\gamma(x)$ is the viscosity solution to

$$H(x, \nabla u^\gamma) + Mu^\gamma = f_\gamma(x), \quad x \in \mathbb{T}^n. \tag{1.6.44}$$

Hint: Exercise 1.5.8 and its solution should be helpful here.

The next step is to send $\gamma \to 0$.

Exercise 1.6.11 Mimic the proof of Lemma 1.6.6 to show that $u^\gamma(x)$ are uniformly Lipschitz: there exists a constant $C_f > 0$ that may depend on $\|f\|_{L^\infty}$ but is independent of $\gamma \in (0, 1)$ and of $M > M_0$ such that

$$|\text{Lip } u^\gamma| \le C_f. \tag{1.6.45}$$

Also show that

$$\|u^\gamma\|_{L^\infty} \le \frac{1}{M}(\|H(\cdot, 0)\|_{L^\infty} + \|f\|_{L^\infty}). \tag{1.6.46}$$

This exercise shows that as long as $M \ge M_0$, the family u^{γ_k} converges, along as subsequence $\gamma_k \downarrow 0$, uniformly in $x \in \mathbb{T}^n$, to a limit $u(x) \in C(\mathbb{T}^n)$ that obeys the same uniform Lipschitz and L^∞-bounds in Exercise 1.6.11. Invoking again the stability result of Exercise 1.5.8 shows that $u(x)$ is the unique viscosity solution to

$$H(x, \nabla u) + Mu = f(x), \quad x \in \mathbb{T}^n. \tag{1.6.47}$$

This finishes the proof of Lemma 1.6.8. □

The end of the proof of Proposition 1.6.7

We now explain how this construction implies the conclusion of Proposition 1.6.7. Let us take $\varepsilon < M$, and re-write equation (1.6.31)

$$H(x, \nabla u) + \varepsilon u = 0, \quad x \in \mathbb{T}^n. \tag{1.6.48}$$

for which we need to find a solution, as

$$H(x, \nabla u) + Mu = (M - \varepsilon)u, \quad x \in \mathbb{T}^n. \tag{1.6.49}$$

As we have mentioned, we define the map $\mathcal{S} : C(\mathbb{T}^n) \to C(\mathbb{T}^n)$ as follows: given $v \in C(\mathbb{T}^n)$, let $u = \mathcal{S}[v]$ be the unique viscosity solution to

$$H(x, \nabla u) + Mu = (M - \varepsilon)v, \quad x \in \mathbb{T}^n. \tag{1.6.50}$$

We claim that $\mathcal{S}$ is a contraction in $C(\mathbb{T}^n)$. We have shown that $u = \mathcal{S}[v]$ can be constructed via the above procedure of passing to the limit $\delta \to 0$, followed by $\gamma \to 0$ in the regularized problem

$$-\delta\Delta u^{\gamma,\delta} + H(x, \nabla u^{\gamma,\delta}) + Mu^{\gamma,\delta} = (M - \varepsilon)v_\gamma, \quad x \in \mathbb{T}^n. \tag{1.6.51}$$

Given $v_1, v_2 \in C(\mathbb{T}^n)$, consider the corresponding solutions to the regularized problems (1.6.51):

$$-\delta\Delta u_1^{\gamma,\delta} + H(x, \nabla u_1^{\gamma,\delta}) + Mu_1^{\gamma,\delta} = (M - \varepsilon)v_{1,\gamma}, \quad x \in \mathbb{T}^n, \tag{1.6.52}$$

and

$$-\delta\Delta u_2^{\gamma,\delta} + H(x, \nabla u_2^{\gamma,\delta}) + M u_2^{\gamma,\delta} = (M-\varepsilon)v_{2,\gamma}, \quad x \in \mathbb{T}^n. \tag{1.6.53}$$

Assume that the difference

$$w = u_1^{\gamma,\delta} - u_2^{\gamma,\delta}$$

attains its maximum at a point x_0. The function w satisfies

$$-\delta\Delta w + H(x, \nabla u_1^{\gamma,\delta}) - H(x, \nabla u_2^{\gamma,\delta}) + Mw = (M-\varepsilon)(v_{1,\gamma} - v_{2,\gamma}), \quad x \in \mathbb{T}^n. \tag{1.6.54}$$

Evaluating this at $x = x_0$, as $\nabla u_1^{\gamma,\delta}(x_0) = \nabla u_2^{\gamma,\delta}(x_0)$, we see that

$$-\delta\Delta w(x_0) + Mw(x_0) = (M-\varepsilon)(v_{1,\gamma}(x_0) - v_{2,\gamma}(x_0)), \quad x \in \mathbb{T}^n. \tag{1.6.55}$$

As x_0 is the maximum of w, we deduce that

$$w(x_0) \le \frac{M-\varepsilon}{M}\|v_{1,\gamma} - v_{2,\gamma}\|_{C(\mathbb{T}^n)}.$$

Using a nearly identical computation for the minimum, we conclude that

$$\|u_1^{\gamma,\delta} - u_2^{\gamma,\delta}\|_{C(\mathbb{T}^n)} \le \frac{M-\varepsilon}{M}\|v_{1,\gamma} - v_{2,\gamma}\|_{C(\mathbb{T}^n)}. \tag{1.6.56}$$

Passing to the limit $\delta \downarrow 0$ and $\gamma \downarrow 0$, we obtain

$$\|u_1 - u_2\|_{C(\mathbb{T}^n)} \le \frac{M-\varepsilon}{M}\|v_1 - v_2\|_{C(\mathbb{T}^n)}, \tag{1.6.57}$$

hence $\mathcal{S}$ is a contraction on $C(\mathbb{T}^n)$, as claimed. Thus, this map has a fixed point, which is the viscosity solution to

$$H(x, \nabla u) + \varepsilon u = 0, \quad x \in \mathbb{T}^n. \tag{1.6.58}$$

This completes the proof of Proposition 1.6.7. □

1.6.2 Existence of the solution to the Cauchy problem

We will now construct the viscosity solution to the Cauchy problem

$$\begin{aligned} &u_t + H(x, \nabla u) = 0, \quad t > 0,\ x \in \mathbb{T}^n, \\ &u(0,x) = u_0(x), \quad x \in \mathbb{T}^n, \end{aligned} \tag{1.6.59}$$

with a continuous initial condition $u_0(x)$. Recall that Exercise 1.5.13 implies the uniqueness of the solution with a given initial condition, so we do not need to address that issue. We make the same assumptions as in Theorem 1.6.2: there exists $C_L > 0$ so that

$$|H(x,p_1) - H(x,p_2)| \le C_L|p_1 - p_2|, \text{ for all } x, p_1, p_2 \in \mathbb{R}^n, \tag{1.6.60}$$

and

$$\lim_{|p|\to+\infty} H(x,p) = +\infty, \quad \text{uniformly in } x \in \mathbb{T}^n. \tag{1.6.61}$$

We will again assume the gradient bound (1.6.6):

$$|\nabla_x H(x,p)| \le K_0(1+|p|), \text{ for all } x \in \mathbb{T}^n, \text{ and } p \in \mathbb{R}^n. \tag{1.6.62}$$

Theorem 1.6.12 *The Cauchy problem (1.6.59) has a unique viscosity solution $u(t,x)$. Moreover, the weak contraction property holds: if $u(t,x)$ and $v(t,x)$ are two solutions to (1.6.59) with the corresponding initial conditions $u_0 \in C(\mathbb{T}^n)$ and $v_0 \in C(\mathbb{T}^n)$, then*

$$\|u(t,\cdot) - v(t,\cdot)\|_{L^\infty} \le \|u_0 - v_0\|_{L^\infty}. \tag{1.6.63}$$

The weak contraction property is recorded here simply for the sake of completeness: we have seen in Exercise 1.5.14 that it follows immediately from the comparison principle. Therefore, we will focus on the existence of the solutions.

An important consequence of the weak contraction principle is that we may restrict ourselves to initial conditions that are smooth. Indeed, suppose that we managed to prove the theorem for smooth initial conditions, and consider $u_0 \in C(\mathbb{T}^n)$. Let $u_0^{(k)}$ be a sequence of smooth functions converging to u_0 in $C(\mathbb{T}^n)$ as $k \to +\infty$, and $u^{(k)}(t,x)$ be the corresponding sequence of solutions to (1.6.59), with the initial conditions $u_0^{(k)}$. It follows from the weak contraction principle that

$$\|u^{(k)} - u^{(m)}\|_{L^\infty(\mathbb{R}_+\times\mathbb{T}^n)} \le \|u_0^{(k)} - u_0^{(m)}\|_{L^\infty},$$

ensuring that $u^{(k)}$ is a uniformly Cauchy sequence on $C([0,+\infty)\times\mathbb{T}^n)$. Hence, it converges uniformly to a continuous function $u \in C(\mathbb{R}_+\times\mathbb{T}^n)$. The stability result in Exercise 1.5.8 implies that u is a solution to the Cauchy problem (1.6.59) with the initial condition $u_0(x)$.

We are now left with the actual construction of a solution to (1.6.59), with the assumption that u_0 is smooth. We are going to use the most pedestrian way to do it: a time discretization. Take a family of time steps $\Delta t \to 0$. For a fixed $\Delta t > 0$, consider the sequence $u_{\Delta t}^n(x)$ defined by setting $u^0(x) := u_0(x)$ and the recursion relation:

$$\frac{u_{\Delta t}^{n+1} - u_{\Delta t}^n}{\Delta t} + H(x, \nabla u_{\Delta t}^{n+1}) = 0, \quad x \in \mathbb{T}^n, \tag{1.6.64}$$

that is an implicit time discretization of (1.6.59). Given $u_{\Delta t}^n(x)$, we look at (1.6.64) as a time-independent Hamilton-Jacobi equation

$$H(x, \nabla u_{\Delta t}^{n+1}) + \frac{1}{\Delta t} u_{\Delta t}^{n+1} = \frac{1}{\Delta t} u_{\Delta t}^n, \quad x \in \mathbb{T}^n. \tag{1.6.65}$$

It is of the type, for which Proposition 1.6.7 guarantees existence of a unique continuous solution $u_{\Delta t}^{n+1}$, as long as $u_{\Delta t}^n$ is continuous. This produces the sequence $u_{\Delta t}^n(x)$, for $n \ge 0$. An approximate solution $u_{\Delta t}$ to the Cauchy problem (1.6.59) is then constructed by interpolating linearly between the times $n\Delta t$ and $(n+1)\Delta t$:

$$u_{\Delta t}(t,x) = u_{\Delta t}^n(x) + \frac{t - n\Delta t}{\Delta t}\big(u_{\Delta t}^{n+1}(x) - u_{\Delta t}^n(x)\big), \quad t \in [n\Delta t, (n+1)\Delta t). \tag{1.6.66}$$

The help provided by the smoothness assumption on u_0 manifests itself in the next proposition.

Proposition 1.6.13 *There is $C > 0$, depending on $\|u_0\|_\infty$ and $Lip(u_0)$ but not on $\Delta t \in (0,1)$, such that the function $u_{\Delta t}(t,x)$ is uniformly Lipschitz continuous in t and x on $[0,+\infty)\times\mathbb{T}^n$, and the Lipschitz constant $Lip(u_{\Delta t})$ of $u_{\Delta t}$ both in t and x, over the set $[0,+\infty)\times\mathbb{T}^n$, satisfies*

$$Lip(u_{\Delta t}) \le C. \tag{1.6.67}$$

This ensures that there exists a sequence $\Delta t_n \to 0$, such that the corresponding sequence $u_{\Delta t_n}$ converges as $n \to \infty$ to a Lipschitz function $u(t,x)$ with the Lipschitz constant $\mathrm{Lip}(u) \le C$. The next step will be to prove

Proposition 1.6.14 *The function $u(t,x)$ is a viscosity solution to the Cauchy problem (1.6.59):*

$$
\begin{aligned}
&u_t + H(x, \nabla u) = 0, \quad t > 0,\ x \in \mathbb{T}^n, \\
&u(0,x) = u_0(x), \quad x \in \mathbb{T}^n.
\end{aligned}
\tag{1.6.68}
$$

Proof. Let us prove this claim first, assuming the conclusion of Proposition 1.6.13. Note that the initial condition $u(0,x) = u_0(x)$ is satisfied by construction, so we only need to check that u is a viscosity solution to (1.6.68). We will only prove that u is a super-solution, the sub-solution property of u can be proved identically. Let $\varphi(t,x)$ be a C^1-test function and (t_0,x_0) be a minimum point for the difference $u - \varphi$. As we have seen in the hint to Exercise 1.5.8, we may assume, possibly after subtracting a quadratic polynomial in t and x from the function φ, that the minimum is strict. Consider the linearly interpolated time discretization $\varphi_{\Delta t}$ of φ: set $\varphi^n(x) = \varphi(n\Delta t, x)$, for $n \ge 0$, and

$$
\varphi_{\Delta t}(t,x) = \varphi^n(x) + \frac{t - n\Delta t}{\Delta t}\big(\varphi^{n+1}(x) - \varphi^n(x)\big), \quad \text{for } t \in [n\Delta t, (n+1)\Delta t).
$$

Note a slight abuse of notation: the function $\varphi_{\Delta t}$ is a linear interpolation of the function φ, while $u_{\Delta t}$ is not the linear interpolation of the function u but rather the linear interpolation of the solution to the time-discretized problem (1.6.64),with the time step Δt. Nevertheless, as the minimum (t_0,x_0) of $u-\varphi$ is strict, and $u_{\Delta t}$ converges to u uniformly, for Δt sufficiently small, there exists a minimum point $(t_{\Delta t}, x_{\Delta t})$ for $u_{\Delta t} - \varphi_{\Delta t}$, such that

$$
\lim_{\Delta t \to 0} (t_{\Delta t}, x_{\Delta t}) = (t_0, x_0).
$$

In addition, because both $u_{\Delta t}$ and $\varphi_{\Delta t}$ are piecewise linear in t, we have $t_{\Delta t} = (n+1)\Delta t$ for some $n \ge 0$. Then we have, again, because $(t_{\Delta t}, x_{\Delta t})$ is a minimum for $u_{\Delta t} - \varphi_{\Delta t}$:

$$
\frac{u_{\Delta t}^{n+1}(x_{\Delta t}) - u_{\Delta t}^n(x_{\Delta t})}{\Delta t} = \partial_t^- u_{\Delta t}((n+1)\Delta t, x_{\Delta t}) \le \partial_t^- \varphi_{\Delta t}((n+1)\Delta t, x_{\Delta t}) = \partial_t \varphi(t_0,x_0) + o(1),
\tag{1.6.69}
$$

as $\Delta t \to 0$. We also have, in the vicinity of (t_0,x_0):

$$
\varphi(t,x) - \varphi_{\Delta t}(t,x) = O(\Delta t^2), \quad \partial_t\varphi(t,x) - \partial_t\varphi_{\Delta t}(t,x) = O(\Delta t), \text{ as } \Delta t \to 0,
\tag{1.6.70}
$$

with the slight catch here that we have to speak of the left and right derivatives of $\varphi_{\Delta t}$ at the discrete times $n\Delta t$. On the other hand, the point $x_{\Delta t}$ is a minimum of

$$
u_{\Delta t}^{n+1}(x) - \varphi_{\Delta t}((n+1)\Delta t, x)
$$

in the x-variable. Since $u_{\Delta t}^{n+1}$ is a viscosity solution to (1.6.64), we have

$$
\frac{u_{\Delta t}^{n+1}(x_{\Delta t}) - u_{\Delta t}^n(x_{\Delta t})}{\Delta t} \ge -H(x_{\Delta t}, \nabla\varphi_{\Delta t}((n+1)\Delta t, x_{\Delta t})) = -H(x_0, \nabla\phi(t_0,x_0)) + o(1),
\tag{1.6.71}
$$

as $\Delta t \to 0$. Putting together (1.6.69)-(1.6.71) and sending Δt to 0, we obtain

$$
\partial_t\varphi(t_0,x_0) + H(x_0, \nabla\varphi(t_0,x_0)) \ge 0,
$$

hence u is a super-solution to (1.6.68). This proves Proposition 1.6.14. $\Box$

Proof of Proposition 1.6.13

The reason behind this proposition is quite simple: if u is a smooth solution to

$$u_t + H(x, \nabla u) = 0, \tag{1.6.72}$$

then the function $v(t,x) = u_t(t,x)$ solves

$$v_t + \nabla_p H(x, \nabla u) \cdot \nabla v = 0, \tag{1.6.73}$$

with the initial condition $v(0,x) = -H(x, \nabla u_0(x))$. It follows from the maximum principle, or the method of characteristics for smooth solutions, that

$$\|v(t,\cdot)\|_{L^\infty} \le \|H(\cdot, \nabla u_0(\cdot))\|_{L^\infty}. \tag{1.6.74}$$

Moreover, (1.6.72) and (1.6.74) together with the coercivity of $H(x,p)$ yield the uniform boundedness of ∇u. The proof of the proposition consists in making this idea rigorous.

Let us recall that $u^n_{\Delta t}$ is the solution to the recursive equation (1.6.64)

$$\frac{u^{n+1}_{\Delta t} - u^n_{\Delta t}}{\Delta t} + H(x, \nabla u^{n+1}_{\Delta t}) = 0, \quad x \in \mathbb{T}^n, \tag{1.6.75}$$

interpolated between the times of the form $n\Delta t$ as in (1.6.66):

$$u_{\Delta t}(t,x) = u^n_{\Delta t}(x) + \frac{t - n\Delta t}{\Delta t}\big(u^{n+1}_{\Delta t}(x) - u^n_{\Delta t}(x)\big), \quad t \in [n\Delta t, (n+1)\Delta t). \tag{1.6.76}$$

The viscosity solution $u^{n+1}_{\Delta t}$ to (1.6.75) can be constructed using the by now familiar idea of a diffusive regularization:

$$-\delta\Delta u^{n+1,\delta}_{\Delta t} + H(x, \nabla u^{n+1,\delta}_{\Delta t}) + \frac{u^{n+1,\delta}_{\Delta t} - u^{n,\delta}_{\Delta t}}{\Delta t} = 0, \quad x \in \mathbb{T}^n, \tag{1.6.77}$$

with $\delta > 0$, and then sending $\delta \downarrow 0$. As we have assumed that $u_0(x)$ is smooth, all $u^{n,\delta}_{\Delta t}(x)$ are also smooth, for all $\delta > 0$.

Exercise 1.6.15 Show that

$$\|u^{n+1,\delta}_{\Delta t}\|_{L^\infty} \le \|u^{n,\delta}_{\Delta t}\|_{L^\infty} + (\Delta t)\|H(\cdot, 0)\|_{L^\infty}. \tag{1.6.78}$$

Hint: look at the maximum x_0 of the smooth function $u^{n+1,\delta}_{\Delta t}$ over $\mathbb{T}^n$.

Exercise 1.6.16 Use the argument in the proof of Lemma 1.6.9 and Exercise 1.6.15 to show that there exists a constant $C_{n,\Delta t}$ that may depend on n and Δt but not on $\delta > 0$, so that

$$\|\nabla u^{n,\delta}_{\Delta t}\|_{L^\infty} \le C_{n,\Delta t}. \tag{1.6.79}$$

The bound (1.6.79) is quite poor as we did not track the dependence of $C_{n,\Delta t}$ on n or Δt, but we have extra help. The differential quotient

$$v_{\Delta t}^{n,\delta} = \frac{u_{\Delta t}^{n+1,\delta} - u_{\Delta t}^{n,\delta}}{\Delta t}$$

satisfies

$$-\delta \Delta v_{\Delta t}^{n+1,\delta} + \frac{v_{\Delta t}^{n+1,\delta}}{\Delta t} + \frac{1}{\Delta t}\Big(H(x, \nabla u_{\Delta t}^{n+1,\delta}) - H(x, \nabla u_{\Delta t}^{n,\delta})\Big) = \frac{v_{\Delta t}^{n,\delta}}{\Delta t}, \tag{1.6.80}$$

for all $n \geq 0$. At the maximum x_M and minimum x_m of the smooth function $v_{\Delta t}^{n,\delta}$ we have

$$\nabla u_{\Delta t}^{n+1,\delta}(x_M) = \nabla u_{\Delta t}^{n,\delta}(x_M), \quad \nabla u_{\Delta t}^{n+1,\delta}(x_m) = \nabla u_{\Delta t}^{n,\delta}(x_m).$$

Using this in (1.6.80) we obtain

$$\|v_{\Delta t}^{n+1,\delta}\|_{L^\infty} \leq \|v_{\Delta t}^{n,\delta}\|_{L^\infty} \leq \cdots \leq \|v_{\Delta t}^{0,\delta}\|_{L^\infty}. \tag{1.6.81}$$

For the last term in the right side we observe that

$$v_{\Delta t}^{0,\delta} = \frac{u_{\Delta t}^{1,\delta} - u_0}{\Delta t}$$

satisfies, instead of (1.6.80), the equation

$$-\delta \Delta v_{\Delta t}^{0,\delta} + \frac{v_{\Delta t}^{0,\delta}}{\Delta t} + \frac{1}{\Delta t} H(x, \nabla u_{\Delta t}^{1,\delta}) = \frac{\delta}{\Delta t}\Delta u_0. \tag{1.6.82}$$

Again, the maximum principle implies

$$\|v_{\Delta t}^{0,\delta}\|_{L^\infty} \leq \|H(\cdot, \nabla u_0)\|_{L^\infty} + \delta \|\Delta u_0\|_{L^\infty}. \tag{1.6.83}$$

Using this in (1.6.81), we conclude that

$$\|v_{\Delta t}^{n,\delta}\|_{L^\infty} \leq \|H(\cdot, \nabla u_0)\|_{L^\infty} + \delta \|\Delta u_0\|_{L^\infty}, \tag{1.6.84}$$

for all $n \geq 0$. This bound is the reason why we have assumed that u_0 is smooth.

We may now pass to the limit $\delta \to 0$ in (1.6.84) and recall the convergence of $u_{\Delta t}^{n,\delta}$ to $u_{\Delta t}^n$, to conclude that

$$v_{\Delta t}^{n,\delta} = \frac{u_{\Delta t}^{n+1,\delta} - u_{\Delta t}^{n,\delta}}{\Delta t} \to v_{\Delta t}^n := \frac{u_{\Delta t}^{n+1} - u_{\Delta t}^n}{\Delta t} \quad \text{as } \delta \downarrow 0. \tag{1.6.85}$$

Combining this with the uniform bound (1.6.84) , we conclude that

$$\Big\|\frac{u_{\Delta t}^{n+1} - u_{\Delta t}^n}{\Delta t}\Big\|_{L^\infty} \leq \|H(\cdot, \nabla u_0)\|_{L^\infty}, \tag{1.6.86}$$

which is a uniform Lipschitz bound on $u_{\Delta t}$ in the t-variable that we need. The reader should compare it to the bound (1.6.74) that we have obtained easily for smooth solutions.

The Lipschitz bound for $u^n_{\Delta t}$ in the x-variable follows easily. Recall that the functions $u^n_{\Delta t}$ satisfy (1.6.64):

$$H(x, \nabla u^{n+1}_{\Delta t}) + \frac{1}{\Delta t} u^{n+1}_{\Delta t} = \frac{1}{\Delta t} u^n_{\Delta t}, \quad x \in \mathbb{T}^n. \tag{1.6.87}$$

We know from Exercise 1.6.16 that $u^n_{\Delta t}$ are Lipschitz – even though we do not know if they have a Lipschitz constant that does not depend on n or Δt. However, this already tells us that $u^n_{\Delta t}$ satisfy (1.6.87) almost everywhere. We write this equation in the form

$$H(x, \nabla u^{n+1}_{\Delta t}) = -v^n_{\Delta t}(x), \quad x \in \mathbb{T}^n. \tag{1.6.88}$$

The uniform bound on $v^n_{\Delta t}$ in (1.6.86) together with the coercivity of $H(x,p)$ imply that there exists a constant $K > 0$ that does not depend on n or Δt so that

$$\|\nabla u^{n+1}_{\Delta t}\|_{L^\infty} \leq K. \tag{1.6.89}$$

This finishes the proof of Proposition 1.6.13. □

Exercise 1.6.17 Prove the following elementary fact that we used in the very last step in the above proof: if $u(x)$ is a Lipschitz function then $\mathrm{Lip}(u) = \|\nabla u\|_{L^\infty}$.

Exercise 1.6.18 *(Hamiltonians that are coercive in u).* So far, we have been remarkably silent about Hamilton-Jacobi equations of the form

$$u_t + H(x, u, \nabla u) = 0, \quad t > 0, x \in \mathbb{T}^n, \tag{1.6.90}$$

with the Hamiltonian that depends also on the function u itself. There is one case when the above theory can be developed without any real input of new ideas: assume that $H(x,u,p)$ is non-decreasing in u, and that there exists $C_0 > 0$ so that for all $R > 0$, there exists $\delta_{1,2}(R)$ such that

$$0 < \delta_1(R) \leq \delta_2(R) < C_0,$$

and, for all $u \in [-R, R]$, we have

$$\delta_1(R)(|p| - 1) \leq H(x,u,p) \leq \delta_2(R)(|p| + 1) \text{ for all } |u| \leq R,\ x \in \mathbb{T}^n \text{ and } p \in \mathbb{R}^n.$$

Prove a well-posedness theorem analogous to Theorem 1.6.12. How far can one stretch the assumptions on $H(x,u,p)$? Hint: coercivity is really something one has to assume, one way or another.

1.7 When the Hamiltonian is strictly convex: the Lagrangian theory

Let us recall that in Section 1.4 we considered the Cauchy problem

$$u_t + \frac{1}{2}|\nabla u|^2 - R(x) = 0, \tag{1.7.1}$$

with an initial condition $u(0,x) = u_0(x)$. We have shown that when both $R(x)$ and $u_0(x)$ are convex, this problem has a smooth solution given by the (at first sight) strange looking expression (1.4.28)

$$u(t,x) = \inf_{\gamma(t)=x} \Big(u_0(\gamma(0)) + \int_0^t \big(\frac{|\gamma'(s)|^2}{2} + R(\gamma(s))\big)ds\Big). \tag{1.7.2}$$

Moreover, this expression is well-defined even if the boundary value problem for the characteristic curves may be not well-posed. Hence, a natural idea is to generalize this formula to other Hamiltonians and take this generalization as the definition of a solution. On the other hand, we already have the notion of a viscosity solution, so an issue is if these objects agree. In this section, we investigate when the variational approach is possible and whether the solution you construct in this way is, indeed, a viscosity solution. We also discuss how the strict convexity of the Hamiltonian gives an improved regularity of the solution.

1.7.1 The Lax-Oleinik formula and viscosity solutions

In the construction of the viscosity solutions, we assumed very little about the Hamiltonian H: all we really needed was coercivity and continuity. The other regularity assumptions we have made are mostly of the technical nature and can be avoided. From now on, we will adopt an even stronger technical assumption that $H(x,p)$ is $C^\infty(\mathbb{T}^n \times \mathbb{R}^n)$ smooth but more crucially we will assume that $H(x,p)$ is uniformly strictly convex in its second variable: there exists $\alpha > 0$ so that

$$D_p^2 H(x,p) \geq \alpha I, \qquad [D_p^2 H(x,p)]_{ij} = \frac{\partial^2 H(x,p)}{\partial p_i \partial p_j}, \tag{1.7.3}$$

in the sense of quadratic forms, for all $x \in \mathbb{T}^n$ and $p \in \mathbb{R}^n$. Unlike the regularity assumptions, the convexity of $H(x,p)$ in p is essential not only for this section, but also for many results on the Hamilton-Jacobi equations.

Exercise 1.7.1 The reader may be naturally concerned that in the construction of the viscosity solutions we have assumed that $H(x,p)$ is uniformly Lipschitz:

$$|H(x,p_1) - H(x,p_2)| \leq C_L|p_1 - p_2| \quad \text{for all } x \in \mathbb{T}^n \text{ and } p_1, p_2 \in \mathbb{R}^n, \tag{1.7.4}$$

and differentiable in x:

$$|\nabla_x H(x,p)| \leq C_0(1 + |p|) \quad \text{for all } x \in \mathbb{T}^n \text{ and } p \in \mathbb{R}^n, \tag{1.7.5}$$

These assumptions are, of course, incompatible with the strict convexity assumption on $H(x,p)$ in (1.7.3). Go through the proofs of existence and uniqueness of the viscosity solutions and show that the coercivity assumption

$$\lim_{|p|\to+\infty} H(x,p) = +\infty \tag{1.7.6}$$

together with the assumption that (1.7.4) and (1.7.5) hold locally in p, in the sense that for ever compact set $\mathcal{K} \subset \mathbb{R}^n$ there exist two constants $C_L(\mathcal{K})$ and $C_0(\mathcal{K})$ such that

$$\begin{aligned} &|H(x,p_1) - H(x,p_2)| \leq C_L|p_1 - p_2| \quad \text{for all } x \in \mathbb{T}^n \text{ and } p_1, p_2 \in \mathcal{K}, \\ &|\nabla_x H(x,p)| \leq C_0(1 + |p|) \quad \text{for all } x \in \mathbb{T}^n \text{ and } p \in \mathcal{K}, \end{aligned} \tag{1.7.7}$$

are sufficient to prove existence and uniqueness of the viscosity solutions both in the Lions-Papanicolaou-Varadhan Theorem 1.6.2 and in Theorem 1.6.12 for the solutions to the Cauchy problem.

The Legendre transform and extremal paths

Recall that in Section 1.2 we have informally argued as follows: given a path $\gamma(s)$, $t \le s \le T$, with the starting point $\gamma(t) = x$, we can define its cost as

$$\mathcal{C}(\gamma)(t) = \int_t^T \widetilde{L}(\dot{\gamma}(s))ds + f(x(T)). \tag{1.7.8}$$

Here, the function $\widetilde{L}(v)$ represents the running cost, and the function $f(x)$ is the terminal cost. The corresponding value function is

$$\widetilde{u}(t,x) = \inf_{\gamma:\ \gamma(t)=x} \mathcal{C}(\gamma)(t), \tag{1.7.9}$$

with the infimum taken over all curves $\gamma \in C^1$ such that $\gamma(t) = x$. We have shown, albeit very informally, that $\widetilde{u}(t,x)$ satisfies the Hamilton-Jacobi equation

$$\widetilde{u}_t + \widetilde{H}(\nabla \widetilde{u}) = 0, \tag{1.7.10}$$

with the terminal condition $u(T,x) = f(x)$. The Hamiltonian $\widetilde{H}(p)$ is given in terms of the running cost $\widetilde{L}(v)$ by (1.2.9):

$$\widetilde{H}(p) = \inf_{v\in\mathcal{A}} \big[\widetilde{L}(v) + v\cdot p\big]. \tag{1.7.11}$$

It is convenient to reverse the direction of time and set

$$u(t,x) = \widetilde{u}(T-t,x). \tag{1.7.12}$$

This function satisfies the forward Cauchy problem

$$u_t + H(\nabla u) = 0, \tag{1.7.13}$$

with the initial condition $u(0,x) = f(x)$ and the Hamiltonian given by

$$H(p) = -\widetilde{H}(p) = -\inf_{v\in\mathbb{R}^n} \big[\widetilde{L}(v) + v\cdot p\big] = \sup_{v\in\mathbb{R}^n}[-p\cdot v - \widetilde{L}(v)] = \sup_{v\in\mathbb{R}^n}[p\cdot v - L(v)], \tag{1.7.14}$$

with the time-reversed cost function

$$L(v) = \widetilde{L}(-v). \tag{1.7.15}$$

The natural questions are, first, if the above construction, using the minimizer in (1.7.9), indeed, produces a solution to the initial value problem for (1.7.13) – so far, our arguments were rather informal, and, second, how it is related to the notion of the viscosity solution.

This bring us to the terminology of the Legendre transforms. One of the standard references for the basic properties of the Legendre transform is [123], where an interested reader may find much more information on this beautiful subject. Given a function $L(v)$, known as the Lagrangian, we define its Legendre transform as in (1.7.14)

$$H(p) = \sup_{v\in\mathbb{R}^n}(p\cdot v - L(v)). \tag{1.7.16}$$

Exercise 1.7.2 Show the function $H(p)$ defined by (1.7.16) is convex. Hint: use the fact that $H(p)$ is the supremum of a family of linear functions in p.

This shows that if we hope to connect the Hamilton-Jacobi equations to the above optimal control problem, this can only be done for convex Hamiltonians. Hence, our assumption (1.7.3) that the Hamiltonian $H(x,p)$ is convex in p.

If the function $L(v)$ is smooth and strictly convex, then, for a given $p \in \mathbb{R}^n$, the maximizer $\bar{v}(p)$ in (1.7.16) is explicit: it is the unique solution to

$$p = \nabla L(\bar{v}). \tag{1.7.17}$$

Exercise 1.7.3 Show that if $L(v)$ is strictly convex, and $H(p)$ is its Legendre transform given by (1.7.16), then we have the duality

$$L(v) = \sup_{p \in \mathbb{R}^n} (p \cdot v - H(p)),$$

so that the Lagrangian L is the Legendre transform of the Hamiltonian H. Hint: this is easier to verify if $L(v)$ is smooth, in addition to being convex.

As a consequence, if a function $H(p)$ is strictly convex, then we can define the Lagrangian L as the Legendre transform of H. If the Hamiltonian $H(x,p)$ depends, in addition, on a variable $x \in \mathbb{T}^n$ as a parameter, then the Lagrangian $L(x,v)$ is defined as the Legendre transform of $H(x,p)$ in the variable p:

$$L(x,v) = \sup_{p \in \mathbb{R}^n} (p \cdot v - H(x,p)), \tag{1.7.18}$$

with the dual relation

$$H(x,p) = \sup_{v \in \mathbb{R}^n} (p \cdot v - L(x,v)). \tag{1.7.19}$$

We usually refer to x as the spatial variable, and to p as the momentum variable.

Exercise 1.7.4 Compute the Lagrangian $L(x,v)$ for the classical mechanics Hamiltonian

$$H(x,p) = \frac{|p|^2}{2m} + U(x),$$

with a given $m > 0$. Why is it called the classical mechanics Hamiltonian? What is the meaning of the two terms in its definition? Hint: consider the characteristic curves for this Hamiltonian.

Exercise 1.7.5 Consider a sequence of smooth strictly convex Hamiltonians $H_\varepsilon(p)$ that converges locally uniformly, as $\varepsilon \to 0$, to $H(p) = |p|$. What happens to the corresponding Lagrangians $L_\varepsilon(v)$ as $\varepsilon \to 0$?

In the context of the forward in time Hamilton-Jacobi equations, with the Hamiltonian that depends on the spatial variable as well, the variational problem (1.7.8)-(1.7.9) is defined as follows. For $t > 0$, and two points $x \in \mathbb{T}^n$ and $y \in \mathbb{T}^n$, we define the function

$$h_t(y, x) = \inf_{\gamma(0)=y, \gamma(t)=x} \int_0^t L(\gamma(s), \dot{\gamma}(s))\, ds. \tag{1.7.20}$$

Here, the infimum is taken over all paths γ on $\mathbb{T}^n$, that are piecewise $C^1[0, t]$, and $L(x, v)$ is the Lagrangian given by (1.7.18). The quantity

$$A(\gamma) = \int_0^t L(\gamma(s), \dot{\gamma}(s))\, ds$$

is usually referred to as the Lagrangian action, or simply the action. This is a classical minimization problem, which admits the following result (Tonelli's theorem).

Proposition 1.7.6 *Given any $(t, x, y) \in \mathbb{R}_+^* \times \mathbb{T}^n \times \mathbb{T}^n$, there exists at least one minimizing path $\gamma(s) \in C^2([0, t]; \mathbb{T}^n)$, such that*

$$h_t(y, x) = \int_0^t L(\gamma(s), \dot{\gamma}(s))\, ds.$$

Moreover there is $C(t, |x - y|) > 0$ such that

$$\|\dot{\gamma}\|_{L^\infty([0,t])} + \|\ddot{\gamma}\|_{L^\infty([0,t])} \le C(t, |x - y|). \tag{1.7.21}$$

The function C tends to $+\infty$ as $t \to 0$ – keeping $|x - y|$ fixed. The function $\gamma(s)$ solves the Euler-Lagrange equation

$$\frac{d}{ds} \nabla_v L(\gamma(s), \dot{\gamma}(s)) - \nabla_x L(\gamma(s), \dot{\gamma}(s)) = 0. \tag{1.7.22}$$

We leave the proof as an exercise but give a hint for the proof. Think of how we proceeded in Section 1.4.2 as blueprint. Consider a minimizing sequence γ_n. First, use the strict convexity of L to obtain the H^1-estimates for γ_n, thus ensuring compactness in the space of continuous paths and weak convergence to $\gamma \in H^1([0, t])$ with fixed ends. Next, show that the convexity of L implies that γ is, indeed, a minimizer. Finally, derive the Euler-Lagrange equation and show that γ is actually C^∞. Such a curve γ is called an *extremal*.

The Lax-Oleinik semigroup and viscosity solutions

We now relate the solutions to the Cauchy problem for the Hamilton-Jacobi equations

$$\begin{aligned} &u_t + H(x, \nabla u) = 0,\ t > 0,\ x \in \mathbb{T}^n, \\ &u(0, x) = u_0(x), \end{aligned} \tag{1.7.23}$$

with a strictly convex Hamiltonian $H(x, p)$, to the minimization problem. We let $L(x, v)$ be the Legendre transform of $H(x, p)$, and define the corresponding function $h_t(y, x)$. Given the initial condition $u_0 \in C(\mathbb{T}^n)$, we define the function

$$u(t, x) = \mathcal{T}(t)u_0(x) = \inf_{y \in \mathbb{T}^n} (u_0(y) + h_t(y, x)). \tag{1.7.24}$$

The following exercise gives the dynamic programming principle, the continuous in time analog of relation (1.2.5) in the time-discrete case we have considered in Section 1.2 .

Exercise 1.7.7 *Show that the infimum in (1.7.24) is attained. Also show that* $(\mathcal{T}(t))_{t>0}$ *is a semi-group: for all* $u_0 \in C(\mathbb{T}^n)$ *one has*

$$\mathcal{T}(t+s)u_0 = \mathcal{T}(t)\mathcal{T}(s)u_0, \quad \textit{for all } t \geq 0 \textit{ and } s \geq 0,$$

that is,

$$u(t,x) = \inf_{y \in \mathbb{T}^n} (u(s,y) + h_{t-s}(y,x)), \tag{1.7.25}$$

for all $0 \leq s \leq t$, *and* $\mathcal{T}(0) = I$.

This semigroup is sometimes referred to as the *Lax-Oleinik semigroup*. Here is its link to the Hamilton-Jacobi equations and the viscosity solutions.

Theorem 1.7.8 *Given* $u_0 \in C(\mathbb{T}^n)$, *the function* $u(t,x) := \mathcal{T}(t)u_0(x)$ *is the unique viscosity solution to the Cauchy problem*

$$\begin{aligned} &u_t + H(x, \nabla u) = 0, \\ &u(0,x) = u_0(x). \end{aligned} \tag{1.7.26}$$

Proof. The initial condition for $u(t,x)$ holds essentially automatically so we only need to check that u is the viscosity solution. We first show the super-solution property: take $t_0 > 0$ and $x_0 \in \mathbb{T}^n$ and let ϕ be a test function such that (t_0, x_0) is a minimum for $u - \phi$. As usual, without loss of generality, we may assume that $u(t_0,x_0) = \phi(t_0,x_0)$. Consider the minimizing point y_0 such that

$$u(t_0,x_0) = u_0(y_0) + h_{t_0}(y_0,x_0).$$

Let also γ be an extremal of the action between the times $t = 0$ and $t = t_0$, going from y_0 to x_0: $\gamma(0) = y_0$, $\gamma(t_0) = x_0$. We have, for all $0 \leq t \leq t_0$:

$$\phi(t,\gamma(t)) \leq u(t,\gamma(t)) \leq u_0(y_0) + \int_0^t L(\gamma(s),\dot{\gamma}(s))\, ds. \tag{1.7.27}$$

The first inequality above holds because (t_0,x_0) is a minimum of $u-\phi$ and $u(t_0,x_0) = \phi(t_0,x_0)$, and the second follows from the definition of $u(t,\gamma(t))$ in terms of the Lax-Oleinik semi-group. Note that at $t = t_0$ both inequalities in (1.7.27) become equalities: the first one because $u(t_0,x_0) = \phi(t_0,x_0)$, and the second because the curve γ is a minimizer for $u(t_0,x_0)$. This implies

$$\frac{d}{dt}\Big(u_0(y_0) + \int_0^t L(\gamma(s),\dot{\gamma})(s)\, ds - \phi(t,\gamma(t))\Big)\Big|_{t=t_0} \leq 0, \tag{1.7.28}$$

or, in other words

$$\phi_t(t_0,x_0) + \dot{\gamma}(t_0) \cdot \nabla\phi(t_0,x_0) - L(\gamma(t_0),\dot{\gamma}(t_0)) \geq 0. \tag{1.7.29}$$

Using the test point $v = \dot{\gamma}(t_0)$ in the definition (1.7.19) of $H(x,p)$, we then obtain

$$\phi_t(t_0,x_0) + H(x_0, \nabla\phi(t_0,x_0)) \geq 0. \tag{1.7.30}$$

Hence, $u(t,x)$ is a viscosity super-solution to (1.7.26).

To show the sub-solution property, consider a test function $\phi(t,x)$, as well as $t_0 > 0$ and $x_0 \in \mathbb{T}^n$, such that the difference $u - \phi$ attains its maximum at (t_0,x_0), and assume, once again, that $u(t_0,x_0)=\phi(t_0,x_0)$. Using the semigroup property (1.7.25), we obtain, for all $t\le t_0$ and any curve $\gamma(t)$ such that $\gamma(t_0)=x_0$:

$$u(t_0,x_0)\le u(t,\gamma(t))+h_{t_0-t}(\gamma(t),x_0)\le \phi(t,\gamma(t))+h_{t_0-t}(\gamma(t),x_0). \tag{1.7.31}$$

Given $v\in\mathbb{R}^n$, we take the test curve

$$\gamma(s)=x_0-(t_0-s)v$$

in (1.7.31), so that

$$\gamma(t)=x_0-(t_0-t)v.$$

Note that the curve

$$\gamma_1(s)=x_0-(t_0-t)v+sv,$$

can be used as a test curve in the definition of $h_{t_0-t}(\gamma(t),x_0)$ because we have $\gamma_1(0)=\gamma(t)$, and $\gamma_1(t_0-t)=x_0$. Using this in (1.7.31) gives

$$\begin{aligned} u(t_0,x_0)&\le \phi(t,x_0-(t_0-t)v)+\int_0^{t_0-t}L(x_0-(t_0-t)v+sv,v)ds\\ &=\phi(t,x_0-(t_0-t)v)+\int_0^{t_0-t}L(x_0-sv,v)ds,\end{aligned} \tag{1.7.32}$$

and, once again, this inequality becomes an equality at $t=t_0$, since $u(t_0,x_0)=\phi(t_0,x_0)$. Just as before, differentiating in t at $t=t_0$ gives

$$\phi_t(t_0,x_0)+v\cdot\nabla\phi(t_0,x_0)-L(x_0,v)\le 0. \tag{1.7.33}$$

As (1.7.33) holds for all $v\in\mathbb{R}^n$, it follows that

$$\phi_t(t_0,x_0)+H(x_0,\nabla\phi(t_0,x_0))\le 0. \tag{1.7.34}$$

Therefore, u is also a viscosity sub-solution to (1.7.26), and the proof is complete. $\square$

Exercise 1.7.9 Show the weak contraction and the finite speed of propagation properties, directly from the Lax-Oleinik formula.

Instant regularization to Lipschitz

We conclude this section with a remarkable result on instant smoothing. We will show that if the initial condition u_0 is continuous on $\mathbb{T}^n$, then the solution to the Cauchy problem

$$\begin{aligned} &u_t+H(x,\nabla u)=0,\quad t>0,\ x\in\mathbb{T}^n,\\ &u(0,x)=u_0(x)\end{aligned} \tag{1.7.35}$$

becomes instantaneously Lipschitz. The improved regularity comes from the strict convexity of the Hamiltonian: indeed, nothing of that sort is true without this assumption, as can be seen from the following exercise.

Exercise 1.7.10 Consider the initial value problem

$$\begin{aligned} &u_t + |u_x| = 0, \quad t > 0, \ x \in \mathbb{T}^1, \\ &u(0, x) = u_0(x). \end{aligned} \tag{1.7.36}$$

(i) Show that the solution to (1.7.36) is given by

$$u(t, x) = \inf_{|x-y| \le t} u_0(y). \tag{1.7.37}$$

Hint: one may do this directly but also by considering a family of strictly convex Hamiltonians $H_\varepsilon(p)$ that converges to $H(p) = |p|$ as $\varepsilon \to 0$, and using the Lax-Oleinik semi-group for

$$\begin{aligned} &u_t^\varepsilon + H_\varepsilon(u_x^\varepsilon) = 0, \quad t > 0, \ x \in \mathbb{T}^1, \\ &u^\varepsilon(0, x) = u_0(x). \end{aligned} \tag{1.7.38}$$

Exercise 1.7.5 may be useful here.
(ii) Given an example of a continuous initial condition $u_0(x)$ such that the viscosity solution to (1.7.36) is not Lipschitz.

On the other hand, if the Hamiltonian is strictly convex we have the following result.

Theorem 1.7.11 *Let $H(x, p)$ be strictly convex, and $u(t, x)$ be the unique solution to the Cauchy problem*

$$\begin{aligned} &u_t + H(x, \nabla u) = 0, \\ &u(0, x) = u_0(x), \end{aligned} \tag{1.7.39}$$

with $u_0 \in C(\mathbb{T}^n)$. Then, the function $u(t, x)$ is Lipschitz in t and x for all $t > 0$.

Let us point the key difference with Proposition 1.6.13: as can be seen from the proof of that proposition, we used the Lipschitz property of the initial condition u_0, and showed that the solution remains Lipschitz at $t > 0$. Here, the initial condition is not assumed to be Lipschitz but only continuous, and the improved regularity comes from the convexity of the Hamiltonian.

Proof. It is sufficient to consider time intervals of length one, and repeat the argument on the subsequent intervals. Given $0 < t \le 1$, and $x \in \mathbb{T}^n$, consider the extremal curve $\gamma(s)$ such that $\gamma(t) = x$, and

$$u(t, x) = u_0(\gamma(0)) + \int_0^t L(\gamma(s), \dot{\gamma}(s)) \, ds. \tag{1.7.40}$$

As $0 \le s \le 1$, both $\gamma(s)$ and $\dot{\gamma}(s)$ are uniformly bounded. Of course, on the torus $\gamma(s)$ is always bounded but it would also be bounded for $0 \le s \le 1$ if we were considering the problem on $\mathbb{R}^n$. Take $h \in \mathbb{R}^n$, and define the curve

$$\gamma_1(s) = \gamma(s) + \frac{s}{t} h, \quad 0 \le s \le t,$$

so that

$$\gamma_1(0) = \gamma(0), \quad \gamma_1(t) = x + h. \tag{1.7.41}$$

We may use the Lax-Oleinik formula for $u(t, x+h)$ and (1.7.40) for $u(t, x)$, as well as (1.7.41), to write

$$\begin{aligned} u(t, x+h) = u(t, \gamma_1(t)) &\le u(\gamma_1(0)) + \int_0^t L(\gamma_1(s), \dot\gamma_1(s))ds \\ &= u(t,x) + \int_0^t \left(L(\gamma_1(s), \dot\gamma_1(s)) - L(\gamma(s), \dot\gamma(s))\right) ds. \end{aligned} \tag{1.7.42}$$

The integral in the right side can be estimated as

$$\begin{aligned} &\int_0^t (L(\gamma_1(s), \dot\gamma_1(s)) - L(\gamma(s), \dot\gamma(s))\ ds = \int_0^t [L(\gamma(s) + \frac{s}{t}h, \dot\gamma(s) + \frac{1}{t}h) - L(\gamma(s), \dot\gamma(s))]\ ds \\ &\le \int_0^t \frac{1}{t}\Big(sh \cdot \nabla_x L(\gamma(s), \dot\gamma(s)) + h \cdot \nabla_v L(\gamma(s), \dot\gamma(s))\Big)\ ds + C_t|h|^2, \end{aligned} \tag{1.7.43}$$

with a constant $C_t > 0$ that may blow up as $t \downarrow 0$. We may now use the Euler-Lagrange equation

$$\frac{d}{ds}\nabla_v L(\gamma(s), \dot\gamma(s)) - \nabla_x L(\gamma(s), \dot\gamma(s)) = 0$$

to rewrite (1.7.43) as

$$\begin{aligned} &\int_0^t (L(\gamma_1(s), \dot\gamma_1(s)) - L(\gamma(s), \dot\gamma(s))\ ds \\ &\le \frac{1}{t}\int_0^t h \cdot \Big(s\frac{d}{ds}\nabla_v L(\gamma(s), \dot\gamma(s)) + \nabla_v L(\gamma(s), \dot\gamma(s))\Big)\ ds + C_t|h|^2 \\ &= h \cdot \nabla_v L(\gamma(t), \dot\gamma(t)) + C_t|h|^2. \end{aligned} \tag{1.7.44}$$

Using (1.7.44) in (1.7.42), we obtain

$$u(t, x+h) - u(t,x) \le h \cdot \nabla_v L(\gamma(t), \dot\gamma(t)) + C_t|h|^2, \tag{1.7.45}$$

which proves the Lipschitz regularity in the spatial variable for all $0 < t \le 1$, because both $\gamma(t)$ and $\dot\gamma(t)$ are bounded. Again, the boundedness of $\gamma(t)$ would only play a role if we considered the problem on $\mathbb{R}^n$, of course. Here, we use the fact that (1.7.45) holds for arbitrary x and $y = x + h$ so that the role of x and y can be switched.

In order to prove the Lipschitz regularity in time, let us examine a small variation of t, denoted by $t + \tau$ with $t + \tau > 0$. Perturbing the extremal curve γ into

$$\gamma_2(s) = \gamma\Big(\frac{t}{t+\tau}s\Big),$$

we still have

$$\gamma_2(0) = \gamma(0), \quad \gamma_2(t+\tau) = \gamma(t) = x.$$

The same computation as above gives

$$\begin{aligned} u(t+\tau, x) = u(t+\tau, \gamma_2(t+\tau)) &\le u(\gamma_2(0)) + \int_0^{t+\tau} L(\gamma_2(s), \dot\gamma_2(s))ds \\ &= u(t,x) + \int_0^t \left(L(\gamma_2(s), \dot\gamma_2(s)) - L(\gamma(s), \dot\gamma(s))\right) ds + \int_t^{t+\tau} L(\gamma_2(s), \dot\gamma_2(s))ds. \end{aligned} \tag{1.7.46}$$

It is now straightforward to see that there exists $C_t' > 0$ that depends on t so that

$$u(t+\tau, x) - u(t, x) \leq C_t'|\tau|.$$

Once again, the role of t and $t' = t + \tau$ can be switched, hence $u(t, x)$ is Lipschitz in t as well, for any $t > 0$, finishing the proof. □

Exercise 1.7.12 (i) Where did we use the strict convexity of the Hamiltonian in the above proof?
(ii) Consider again the initial value problem (1.7.36) with the convex but non strictly convex Hamiltonian $H(p) = |p|$ and a continuous initial condition $u_0(x)$ that is not Lipschitz continuous. Consider a sequence of smooth strictly convex Hamiltonians $H_\varepsilon(p)$ such that $H_\varepsilon(p) \to H(p)$ as $\varepsilon \to 0$, locally uniformly on $\mathbb{R}$. Review the above proof and see what will happen to the Lispchitz constant of the corresponding solution $u^\varepsilon(t, x)$ to the Cauchy problem

$$\begin{aligned} &u_t^\varepsilon + H(u_x^\varepsilon) = 0, \quad t > 0, \ x \in \mathbb{T}^1, \\ &u^\varepsilon(0, x) = u_0(x), \end{aligned} \tag{1.7.47}$$

constructed by the Lax-Oleinik formula. Hint: again, Exercise 1.7.5 may be useful here.

Exercise 1.7.13 Take $t > 0$ and $\gamma(s)$ an extremal such that u is differentiable at $x = \gamma(t)$. Show that

$$\nabla u(t, x) = \nabla_v L(x, \dot{\gamma}(t)). \tag{1.7.48}$$

and

$$u_t(t, x) = -H(x, \nabla u(t, x)). \tag{1.7.49}$$

1.7.2 Semi-concavity and $C^{1,1}$ regularity

As we have mentioned, the Cauchy problem for a Hamilton-Jacobi equation

$$u_t + H(x, \nabla u) = 0, \tag{1.7.50}$$

with a prescribed initial condition $u(0, x) = u_0(x)$, may have more than one Lipschitz solution, so it is worth asking whether the unique viscosity solution has some additional regularity when the Hamiltonian is strictly convex, so that the solution can be constructed by the Lax-Oleinik semigroup. A relevant notion is that of semi-concavity. Most of the material of this section comes from [62].

Semi-concavity

We begin with the following definition.

Definition 1.7.14 *If B is an open ball in $\mathbb{R}^n$, F a closed subset of B and K a positive constant, we say that $u \in C(B)$ is K-semi-concave on F if for all $x \in F$, there is $l_x \in \mathbb{R}^n$ such that for all $h \in \mathbb{R}^n$ satisfying $x + h \in B$, we have:*

$$u(x+h) \leq u(x) + l_x \cdot h + K|h|^2. \tag{1.7.51}$$

The function u is said to be K-semi convex on F if $-u$ is K-semi-concave on F.

Exercise 1.7.15 Examine the proof of Theorem 1.7.11 and check that it actually proves that for any $t > 0$ there exists $C_t > 0$ so that $u(t, x)$ is C_t-semi-concave in x.

The next theorem is crucial for the sequel. If u is continuous in an open ball B in $\mathbb{R}^n$, and F is a closed subset of B, we say that $u \in C^{1,1}(F)$ if u is differentiable in F and ∇u is Lipschitz over F.

Theorem 1.7.16 *Let B be an open ball of $\mathbb{R}^n$ and F closed in B. If $u \in C(B)$ is K-semi-concave and K-semi-convex in F, then $u \in C^{1,1}(F)$.*

Proof. As u is both K semi-concave and K-semi-convex, for all $x \in F$, there are two vectors l_x and m_x such that for all h such that $x + h \in B$ we have

$$\begin{aligned} u(x+h) &\leq u(x) + l_x \cdot h + K|h|^2, \\ u(x+h) &\geq u(x) + m_x \cdot h - K|h|^2 \end{aligned} \tag{1.7.52}$$

which yields

$$(m_x - l_x) \cdot h \leq 2K|h|^2.$$

As this is true for all h sufficiently small, we conclude that $l_x = m_x$ and, therefore, u is differentiable at x, and

$$l_x = m_x = \nabla u(x).$$

Next, we show that ∇u is Lipschitz over F. Given $(x, y, h) \in F \times F \times \mathbb{R}^n$, such that both $x + h \in B$ and $y + h \in B$, the semi-convexity and semi-concavity inequalities, written, respectively, between $x + h$ and x, x and y, and $x + h$ and y, give:

$$\begin{aligned} &|u(x+h) - u(x) - \nabla u(x) \cdot h| \leq K|h|^2 \\ &|u(x) - u(y) - \nabla u(y) \cdot (x-y)| \leq K|x-y|^2 \\ &|u(y) - u(x+h) + \nabla u(y) \cdot (x+h-y)| \leq K|x+h-y|^2. \end{aligned}$$

Adding the three inequalities above, we obtain:

$$|(\nabla u(x) - \nabla u(y)) \cdot h| \leq 3K(|h|^2 + |x-y|^2). \tag{1.7.53}$$

Taking

$$h = |x-y| \frac{\nabla u(x) - \nabla u(y)}{|\nabla u(x) - \nabla u(y)|},$$

in the inequality (1.7.53) gives

$$|\nabla u(x) - \nabla u(y)| \leq 6K|x-y|,$$

which is the Lipschitz property of ∇u that we sought. $\square$

Improved regularity of the viscosity solutions

Let us come back to the solution $u(t,x)$ to the Cauchy problem

$$\begin{aligned} &u_t + H(x,\nabla u) = 0, \\ &u(0,x) = u_0(x). \end{aligned} \tag{1.7.54}$$

We first prove that if γ is a minimizing curve for $u(t,x)$, with $\gamma(t) = x$, then it is also a minimizer for $u(s,\gamma(s))$ for all $0 \le s \le t$.

Proposition 1.7.17 *Fix $t > 0$ and $x \in \mathbb{T}^n$, and a minimizing curve γ such that $\gamma(t) = x$, and*

$$u(t,x) = u_0(\gamma(0)) + \int_0^t L(\gamma(s),\dot\gamma(s))\, ds. \tag{1.7.55}$$

Then for all $0 \le s \le s' \le t$ we have

$$u(s',\gamma(s')) = u_0(\gamma(0)) + \int_0^{s'} L(\gamma(\sigma),\dot\gamma(\sigma))\, d\sigma = u(s,\gamma(s)) + \int_s^{s'} L(\gamma(\sigma),\dot\gamma(\sigma))\, d\sigma. \tag{1.7.56}$$

Exercise 1.7.18 Relate the result of this proposition to the dynamic programming principle.

Proof. The Lax-Oleinik formula implies that for all $0 < s < t$ we have

$$u(s,\gamma(s)) \le u_0(\gamma(0)) + \int_0^s L(\gamma(\sigma),\dot\gamma(\sigma))\, d\sigma.$$

Assume that for some $0 < s < t$, we have a strict inequality

$$u(s,\gamma(s)) < u_0(\gamma(0)) + \int_0^s L(\gamma(\sigma),\dot\gamma(\sigma))\, d\sigma. \tag{1.7.57}$$

Then, there exists a curve $\gamma_1(s')$, $0 \le s' \le s$, such that $\gamma_1(s) = \gamma(s)$, and

$$u_0(\gamma_1(0)) + \int_0^s L(\gamma_1(\sigma),\dot\gamma_1(\sigma))\, d\sigma < u_0(\gamma(0)) + \int_0^s L(\gamma(\sigma),\dot\gamma(\sigma))\, d\sigma.$$

Then, we can consider the concatenated curve $\gamma_2(s)$ so that $\gamma_2(s') = \gamma_1(s')$ for $0 \le s' \le s$, and $\gamma_2(s') = \gamma(s')$ for $s \le s' \le t$. The resulting curve is piece-wise $C^1[0,t]$, hence is an allowed trajectory. This would give

$$\begin{aligned} u(t,\gamma(t)) &= u_0(\gamma(0)) + \int_0^s L(\gamma(\sigma),\dot\gamma(\sigma))\, d\sigma + \int_s^t L(\gamma(\sigma),\dot\gamma(\sigma))d\sigma \\ &> u_0(\gamma_2(0)) + \int_0^t L(\gamma_2(s),\dot\gamma_2(s))\, ds, \end{aligned} \tag{1.7.58}$$

which would contradict the extremal property of the curve γ between the times 0 and t. Therefore, (1.7.57) can not hold, and for all $0 \le s \le s' \le t$ we have:

$$u(s',\gamma(s')) = u_0(\gamma(0)) + \int_0^{s'} L(\gamma(\sigma),\dot\gamma(\sigma))\, d\sigma = u(s,\gamma(s)) + \int_s^{s'} L(\gamma(\sigma),\dot\gamma(\sigma))\, d\sigma. \tag{1.7.59}$$

This finishes the proof of Proposition 1.7.17. □

Definition 1.7.19 *We say that $\gamma : [0,t] \to \mathbb{T}^n$ is calibrated by u if (1.7.55) holds.*

Let us define the conjugate semigroup of the Lax-Oleinik semigroup by:

$$\widetilde{\mathcal{T}}(t)u_0(x) = \sup_{y\in\mathbb{T}^n}(u_0(y) - h_t(x,y)), \quad \forall u_0 \in C(\mathbb{T}^n),\ t > 0. \tag{1.7.60}$$

We will denote $\widetilde{u}(t,x) = \widetilde{\mathcal{T}}(t)u_0(x)$. The following lemma is proved exactly as Theorem 1.7.11.

Lemma 1.7.20 *Let $u_0 \in C(\mathbb{T}^n)$ and $\sigma > 0$. There is $K(\sigma) > 0$ such that $\widetilde{\mathcal{T}}(\sigma)u_0$ is $K(\sigma)$-semi-convex. The constant $K(\sigma)$ blows up as $\sigma \to 0$.*

Given $0 < s < s'$, we define the set $\Gamma_{s,s'}[u_0]$ as the union of all points $(s_1, x) \in [s, s'] \times \mathbb{T}^n$, so that the extremal calibrated by u, which passes through the point x at the time s_1 can be continued forward in time until the time s', and backward in time until the time s.

Corollary 1.7.21 *Let $u_0 \in C(\mathbb{T}^n)$ and $u(t,x) = \mathcal{T}(t)u_0(x)$, and $0 < s_1 < s_2$, then for any $\varepsilon > 0$, the function $u \in C^{1,1}(\Gamma_{s_1,s_2+\varepsilon} \cap ([s_1, s_2] \times \mathbb{T}^n))$.*

Proof. Let us take $(s, x_0) \in \Gamma_{s_1,s_2+\varepsilon}$, with $s_1 \le s \le s_2$, so that that the extremal γ such that $x_0 = \gamma(s)$ can be continued past the time s, until the time $s_2 + \varepsilon$.

Let us first deal with the spatial regularity. As we have mentioned in Exercise 1.7.15, there is $K > 0$ depending on s_1 such that the function $u(s,x)$ is K-semi-concave at all $x \in \mathbb{T}^n$ for all $s \ge s_1$, in particular, at x_0. Hence, we only need to argue that u is semi-convex at x_0, and here we are going to use the fact that $(s, x_0) \in \Gamma_{s_1,s_2+\varepsilon}$. Note that for all $y \in \mathbb{R}^n$ we have, by the Lax-Oleinik formula,

$$u(s_2 + \varepsilon, y) \le u(s, x_0) + h_{s_2+\varepsilon-s}(x_0, y). \tag{1.7.61}$$

In addition, the calibration relation (1.7.59) implies that if $x_0 = \gamma(s)$ and $(s, x_0) \in \Gamma_{s_1,s_2+\varepsilon}$, then equality is attained when $y = \gamma(s_2 + \varepsilon)$. We conclude that in this case we have

$$u(s, x_0) = \sup_{y\in\mathbb{T}^n}(u(s_2 + \varepsilon, y) - h_{s_2+\varepsilon-s}(x_0, y)) = \widetilde{\mathcal{T}}(s_2 + \varepsilon - s)[u(s_2 + \varepsilon, \cdot)](x_0).$$

It follows from Lemma 1.7.20 that there is a constant $\widetilde{K}$ depending on ε, such that $u(s,\cdot)$ is $\widetilde{K}$-semi-convex in x on $\Gamma_{s_1,s_2+\varepsilon} \cap ([s_1, s_2] \times \mathbb{T}^n)$.

Theorem 1.7.16 now implies that the function $u(s,\cdot)$ is $C^{1,1}$ in x on the set $\Gamma_{s_1,s_2+\varepsilon}$ for all $s_1 \le s \le s_2$. To end the proof, one just has to invoke relation (1.7.49) in Exercise 1.7.13 to obtain the corresponding regularity in the time variable. □

This corollary may not, at first sight, look so striking. To enjoy its scope, let us specialize it to the solutions to the stationary equation

$$H(x, \nabla u) = 0, \tag{1.7.62}$$

assuming that they exist. Corollary 1.7.21 allows us to discover the following

Corollary 1.7.22 *Consider a solution u of (1.7.62), and let F be the set of all points $x \in \mathbb{T}^n$ such that there exists $\varepsilon_x > 0$ and a C^1 curve $\gamma : (-\varepsilon_x, \varepsilon_x) \to \mathbb{T}^n$ such that $\gamma(0) = x$ and*

$$u(\gamma(\varepsilon_x)) - u(\gamma(-\varepsilon_x)) = \int_{-\varepsilon_x}^{\varepsilon_x} L(\gamma(s), \dot{\gamma}(s))ds. \tag{1.7.63}$$

Then $u \in C^{1,1}(F)$.

In other words, u is $C^{1,1}$ at every point through which an extremal of the Lagrangian passes, as opposed to ending at this point.

Let us examine some further consequences of this fact, in the form of a few exercises, just to give a glimpse of how far reaching these considerations can be. Their solution does not need more tools or ideas than the ones already presented, but they are fairly elaborate. We begin with an application of the finite speed of propagation property.

Exercise 1.7.23 Let $u(x)$ be a Lipschitz viscosity solution of

$$H(x, \nabla u) = 0$$

in a bounded open subset Ω of $\mathbb{R}^n$. Show that, for every open subset Ω_1 of Ω such that $\overline{\Omega}_1$ is compactly embedded in Ω, there is $\varepsilon > 0$ such that, for all $t \in [0, \varepsilon]$ and $x \in \Omega_1$ we have

$$u(x) = \mathcal{T}(t)u(x).$$

We continue with a statement that looks surprisingly elementary. However its solution is not.

Exercise 1.7.24 Let Ω be an open subset of $\mathbb{R}^n$ and u_p a sequence in $C^1(\Omega)$, such that

$$|\nabla u_p| = 1 \text{ for all } p.$$

Show that all uniform limits of u_p are C^1 functions. Hint: if $x_0 \in \Omega$, then, for small $\varepsilon > 0$, the function $u_p(x)$ coincides, in a small neighborhood of x, with both $\mathcal{T}(\varepsilon)u_p$ and $\widetilde{\mathcal{T}}(\varepsilon)u_p$. Note that the Hamiltonian is not strictly convex, so some care needs to be given to the definition of the Lax-Oleinik semigroup and its adjoint. If in doubt, look at (1.7.64) below.

We end the section with two regularity properties of the distance function. Recall that, if S is a subset of $\mathbb{R}^n$, the distance function to S is given by

$$d_S(x) = \inf_{v \in S} |x - v|.$$

It is, obviously, a Lipschitz function with Lipschitz constant 1. We can say much more, just recalling the age-old fact that the shortest path between two points is the line joining these two points: this makes d_S a viscosity solution of $|\nabla d| = 1$, or, even better:

$$|\nabla d|^2 = 1. \tag{1.7.64}$$

We may use the previous theory for the following results.

Exercise 1.7.25 If S is a compact set, $x_0 \notin S$ and v is such that

$$|x - v| = d_S(x),$$

then d_S is $C^{1,1}$ on the line segment $[v, x]$.

Exercise 1.7.26 If S is a convex set, then d_S is $C^{1,1}$ outside S.

If you are stuck with any of the above three exercises, see [62].

1.8 Large time behavior in particular cases

For the rest of this chapter, we go back to the long term behavior of the solutions to the Hamilton-Jacobi equations but unlike in Section 1.3, we now consider the inviscid case. In this initial section, we will focus on two examples. First, we will consider equations of the form

$$u_t + \frac{1}{2}|\nabla u|^2 = f(x), \tag{1.8.1}$$

with the classical Hamiltonian

$$H(x,p) = \frac{|p|^2}{2} - f(x). \tag{1.8.2}$$

This equation arises naturally in the context of classical mechanics. The strict convexity of the classical Hamiltonian (1.8.2) will allow us to use the Lax-Oleinik formula to understand the long time behavior for the solutions to (1.8.1), in a straightforward and elegant way.

Then, we will consider the Hamilton-Jacobi equation

$$u_t + R(x)\sqrt{1+|\nabla u|^2} = 0, \tag{1.8.3}$$

with the Hamiltonian

$$H(x,p) = R(x)\sqrt{1+|p|^2}. \tag{1.8.4}$$

The Hamiltonian in (1.8.4) is locally strictly convex in its second variable but not uniformly strictly convex. We could also attack the problem via the Lax-Oleinik formula, with a little extra technical argument due to the lack of the global strict convexity. We will not, however, rely on the strict convexity in any form in the analysis of the long time behavior for the solutions to (1.8.3). The separate arguments that we are going to display for this problem will work, at almost no additional cost, for the important class of Hamiltonians of the form

$$H(x,p) = |\nabla u| - f(x), \tag{1.8.5}$$

which are not strictly convex even locally. The proof is inspired by the arguments in [107].

Let us mention, looking ahead, that despite the difference in the approaches to the two cases, we will see some strong similarities in the underlying dynamics that will allow us to address the general case in the next section. We chose to start with these examples as the proofs here are much more concrete.

On the technical side, we will assume for (1.8.1) that the function $f(x)$ is smooth, and that the function $R(x)$ in (1.8.3) is smooth and positive: there exists $R_0 > 0$ so that

$$R(x) \geq R_0 > 0 \text{ for all } x \in \mathbb{T}^n, \tag{1.8.6}$$

and will use the notation

$$\bar{R} = \|R\|_{L^\infty}. \tag{1.8.7}$$

Note that the assumptions for the Hamiltonian $H(x,p) = R(x)\sqrt{1+|p|^2}$ fall in line with those made in Section 1.3 on the convergence to the viscous waves, and in Section 1.6 on the existence of the inviscid waves and of the solutions to the inviscid Cauchy problem. As usual, the smoothness assumptions on the function $f(x)$ and $R(x)$ can be greatly relaxed.

Let us first explain how equation (1.8.3) comes up from simple geometric considerations. Consider a family of hypersurfaces $\Sigma(t)$ of $\mathbb{R}^{n+1}$, moving according to an imposed normal velocity $R(y)$:

$$V_n = R(y), \quad y \in \mathbb{R}^{n+1}, \tag{1.8.8}$$

the function $R(y)$ being given and positive. Assume that, at each time $t \geq 0$, the surface $\Sigma(t)$ is the level set of a function $v(t, y)$:

$$\Sigma(t) = \{y \in \mathbb{R}^{n+1} : \ v(t, y) = 0\}.$$

It is easy to see that the normal velocity V_n at the point y, at time t, is given by

$$V_n(t, y) = \frac{v_t(t, y)}{|\nabla v(t, y)|},$$

so that the evolution equation for the function $v(t, y)$ is

$$v_t = R(y)|\nabla v| \quad \text{ on } \Sigma(t). \tag{1.8.9}$$

This evolution equation is interesting in itself, and is known in the literature on the mathematical theory of combustion as the G-equation. It also appears in many computational methods where it is often called the level sets equation. In particular, it allows to model coalescence of objects in digital animation.

We are going to consider a special situation when $\Sigma(t)$ is given in the form of a graph of a periodic function $u(t, x)$, $x \in \mathbb{R}^n$, that is, writing $y = (x, y_{n+1})$, with $x \in \mathbb{T}^n$ and $y_{n+1} \in \mathbb{R}$, we have

$$v(t, y) = y_{n+1} - u(t, x), \qquad x \in \mathbb{T}^n,$$

and also that $R(y)$ is actually a function of the form $R(x)$ – it depends only on the first n coordinates of y. Then we obtain from (1.8.9)

$$u_t + R(x)\sqrt{1 + |\nabla_x u|^2} = 0, \quad x \in \mathbb{T}^n, \tag{1.8.10}$$

which is (1.8.3).

We will begin with the analysis of the wave solutions to (1.8.1) and (1.8.3) – as we will soon see, this study is essentially identical for both problems. Then, we will consider the long time convergence to the wave solutions, and there the two analyses will diverge.

1.8.1 Counting the waves

The first step is to understand the wave solutions to (1.8.1) and (1.8.3). Note that a wave solution to (1.8.3) satisfies

$$R(x)\sqrt{1 + |\nabla u|^2} = c, \quad x \in \mathbb{T}^n, \tag{1.8.11}$$

an equation that can be alternatively stated as

$$|\nabla u(x)|^2 = g(x), \quad x \in \mathbb{T}^n, \tag{1.8.12}$$

with

$$g(x) = \frac{c^2}{R^2(x)} - 1. \tag{1.8.13}$$

On the other hand, a wave solution to (1.8.1) solves

$$\frac{1}{2}|\nabla u(x)|^2 = f(x) + c, \quad x \in \mathbb{T}^n, \tag{1.8.14}$$

that can also be re-stated as (1.8.13), but now with

$$g(x) = 2(f(x) + c). \tag{1.8.15}$$

Thus, in both cases, existence of the wave solutions is equivalent to the question of existence of steady solutions to (1.8.12).

Identification of the speed

We begin with the following.

Proposition 1.8.1 *A solution to an equation of the form*

$$|\nabla u(x)|^2 = f(x) + \gamma, \quad x \in \mathbb{T}^n, \tag{1.8.16}$$

with a smooth function f exists if and only if

$$\gamma = -\min_{x \in \mathbb{T}^n} f(x). \tag{1.8.17}$$

In other words, a solution to

$$|\nabla u(x)|^2 = f(x), \quad x \in \mathbb{T}^n, \tag{1.8.18}$$

exists if and only if

$$\min_{x \in \mathbb{T}^n} f(x) = 0. \tag{1.8.19}$$

A consequence of this proposition is that that the only c such that equation

$$R(x)\sqrt{1 + |\nabla u_\infty|^2} = c, \quad x \in \mathbb{T}^n, \tag{1.8.20}$$

has a solution $u_\infty(x)$ is $c = \bar{R}$, as seen from (1.8.12)-(1.8.13).

To understand the main idea of the proof, note that the unique γ for which (1.8.16) has a solution, can be alternatively defined as the only value of γ such that each solution to the Cauchy problem

$$\begin{aligned} &u_t + |\nabla u|^2 = f(x) + \gamma, \quad t > 0, \; x \in \mathbb{T}^n, \\ &u(0, x) = u_0(x), \end{aligned} \tag{1.8.21}$$

is uniformly bounded in time. This is an immediate consequence of the comparison principle for the viscosity solutions.

Exercise 1.8.2 (i) Explain this point: show that if $\gamma \neq c$ then the solution to the Cauchy problem (1.8.21) can not remain bounded as $t \to +\infty$, and, conversely, if $\gamma = c$ then it remains bounded as $t \to +\infty$.
(ii) Show also that c is the unique value γ such that there exists both a sub-solution $\underline{u}$ and a super-solution $\overline{u}$ to

$$|\nabla u|^2 = f(x) + \gamma, \quad x \in \mathbb{T}^n. \tag{1.8.22}$$

Hint: solutions to (1.8.21) may be helpful here.

Proof of Proposition 1.8.1. We know from the Lions-Papanicolaou-Varadhan theorem that for each $f \in C(\mathbb{T}^n)$ there exists some $\gamma \in \mathbb{R}$ such that a solution to

$$|\nabla u(x)|^2 = f(x) + \gamma, \quad x \in \mathbb{T}^n, \tag{1.8.23}$$

exists. We need to show that

$$\gamma = -\min_{x \in \mathbb{T}^n} f(x). \tag{1.8.24}$$

As in Exercise 1.8.2(ii), we only need to construct a sub-solution and a super-solution to (1.8.23) for γ as in (1.8.24). First, observe that if

$$\gamma + \min_{x \in \mathbb{T}^n} f(x) \geq 0, \tag{1.8.25}$$

then all constants are sub-solutions to (1.8.23).

On the other hand, a quadratic function of the form

$$\overline{u}(x) = \frac{\alpha}{2}|x - x_0|^2, \tag{1.8.26}$$

with some $x_0 \in \mathbb{T}^n$, is a super-solution to (1.8.23) if

$$\alpha^2 |x - x_0|^2 \geq f(x) + \gamma, \quad \text{for all } x \in \mathbb{T}^n. \tag{1.8.27}$$

It follows that, in particular,

$$f(x_0) + \gamma \leq 0,$$

hence such super-solution can exist only if

$$\gamma + \min_{x \in \mathbb{T}^n} f(x) \leq 0. \tag{1.8.28}$$

On the other hand, if (1.8.28) does hold, x_0 is a minimum of $f(x)$, and f is smooth, as we assume here, then (1.8.27) does hold if we choose $\alpha > 0$ to be sufficiently large.

Thus, if $\gamma = -\min_{x \in \mathbb{T}^n} f(x)$ then we can find both a sub-solution and a super-solution to (1.8.23), finishing the proof. □

Exercise 1.8.3 Note that the super-solution we have constructed in (1.8.26) is not periodic. Explain why this is not an issue.

Exercise 1.8.4 We did use the assumption that $f(x)$ is smooth in the construction of the super-solution in the above proof. Show that nevertheless the conclusion of Proposition 1.8.1 holds for $f \in C(\mathbb{T}^n)$. Hint: approximate $f \in C(\mathbb{T}^n)$ by a sequence of smooth functions f_k that converges uniformly to f and obtain a uniform Lipschitz bound for the solutions to

$$|\nabla u_k|^2 = f_k(x) + \gamma_k, \quad \gamma_k := -\min_{x\in\mathbb{T}^n} f_k(x),$$

such that $u_k(0) = 0$. Finally, use the stability property of the viscosity solutions to show that u_k converges, along a subsequence, to a viscosity solution to

$$|\nabla u|^2 = f(x) + \gamma, \quad \gamma := -\min_{x\in\mathbb{T}^n} f(x). \tag{1.8.29}$$

A simple example of the non-uniqueness of the waves

Before proceeding with the description of the set of the solutions to

$$|\nabla u(x)|^2 = f(x), \quad x \in \mathbb{T}^n, \tag{1.8.30}$$

under the assumption that

$$\min_{x\in\mathbb{T}^n} f(x) = 0, \tag{1.8.31}$$

let us explain why the solutions may be not unique. This is a big difference with the viscous case

$$-\Delta u + H(x, \nabla u) = c, \tag{1.8.32}$$

described in Theorem 1.3.1, where both the speed c and the solution u are unique.

We consider a very simple example in one dimension:.

$$|u'| = f(x), \quad x \in \mathbb{T}^1. \tag{1.8.33}$$

Assume that $f \in C^1(\mathbb{T}^1)$ is 1/2-periodic, satisfies

$$f(x) > 0 \text{ on } (0, 1/2) \cup (1/2, 1), \text{ and } f(0) = f(1/2) = f(1) = 0.$$

and is symmetric with respect to $x = 1/4$ (and thus $x = 3/4$). Let u_1 and u_2 be 1-periodic and be defined, over a period, as follows:

$$u_1(x) = \begin{cases} \displaystyle\int_0^x f(y)\,dy, & 0 \le x \le \dfrac{1}{2}, \\ \displaystyle\int_x^1 f(y)\,dy, & \dfrac{1}{2} \le x \le 1, \end{cases} \qquad u_2(x) = \begin{cases} \displaystyle\int_0^x f(y)\,dy, & 0 \le x \le \dfrac{1}{4}, \\ \displaystyle\int_x^{1/2} f(y)\,dy, & \dfrac{1}{4} \le x \le \dfrac{1}{2}, \\ u_2 \text{ is } \dfrac{1}{2}\text{-periodic.} \end{cases}$$

Note that $u_1(x)$ is continuously differentiable but $u_2(x)$ is only Lipschitz: its graph has corners at $x = 1/4$ and $x = 3/4$.

Exercise 1.8.5 Verify that both u_1 and u_2 are viscosity solutions of (1.8.33), and u_2 cannot be obtained from u_1 by the addition a constant. Pay attention to what happens at $x = 1/4$ and $x = 3/4$ with $u_2(x)$. Why can't you construct a solution that would have a corner at a minimum rather than the maximum?

Trajectories at very negative times

The above example of non-uniqueness inspires a more systematic study of the steady solutions to

$$|\nabla u|^2 = f(x), \quad x \in \mathbb{T}^n, \tag{1.8.34}$$

in order to understand how many steady solutions this problem may have. We assume that the function f is smooth and non-negative:

$$f(x) \geq 0 \ \text{ for all } x \in \mathbb{T}^n, \tag{1.8.35}$$

and

$$\min_{x\in\mathbb{T}^n} f(x) = 0. \tag{1.8.36}$$

This ensures existence of a solution to (1.8.34), via Proposition 1.8.1. The smoothness assumption on the function f is adopted merely for convenience, continuity of f would certainly suffice.

An important and non-technical assumption is that the function $f(x)$ has finitely many zeroes x_1,..., x_N. We will see that an absolutely crucial role in the analysis will be played by the set

$$\mathcal{Z} = \{x : \ f(x) = 0\} = \{x_1, \dots, x_N\}. \tag{1.8.37}$$

What follows is a (much simplified) adaptation of the last chapter of the book of Fathi [63].

As we have mentioned, the viscosity solutions to (1.8.34) exist by Proposition 1.8.1 and our assumptions on f. The Lagrangian associated to the Hamiltonian $H(x,p) = |p|^2 - f(x)$ is

$$L(x,v) = \frac{|v^2|}{4} + f(x). \tag{1.8.38}$$

Thus, the viscosity solutions to (1.8.34) satisfy the Lax-Oleinik formula: for any $t < 0$ we have

$$u(x) = \inf_{\gamma(0)=x} \Big(u(\gamma(t)) + \int_t^0 \Big(\frac{|\dot\gamma(s)|^2}{4} + f(\gamma(s)) \Big) ds \Big). \tag{1.8.39}$$

We know from the results of the preceding section that the infimum is, in fact, a minimum, attained at an extremal of the Lagrangian, that we denote $\gamma_t(s)$, $t \leq s \leq 0$. Note that $L(x,v)$ given by (1.8.38) is nonnegative and vanishes only at the points of the form $(x,v) = (x_i, 0)$, with $i \in \{1, ..., N\}$. Hence, we expect that the minimizers in (1.8.39) should prefer to stay near the points where f vanishes, and move very slowly around those points. To formalize this idea, we would like to send the starting time $t \to -\infty$ and say that each minimizing curve $\gamma_t(s)$ is near one of $x_i \in \mathcal{Z}$, for s sufficiently large and negative.

Proposition 1.8.6 *The function $u(x)$ can be written as*

$$u(x) = \inf_{x_i\in\mathcal{Z}} \inf_{\gamma(-\infty)=x_i, \gamma(0)=x} \Big(u(x_i) + \int_{-\infty}^0 \Big(\frac{|\dot\gamma(s)|^2}{4} + f(\gamma(s)) \Big) ds \Big), \tag{1.8.40}$$

with the infimum taken over all curves $\gamma(s)$ such that $\gamma(0) = x$ and $\gamma(s) \to x_i$ as $s \to -\infty$.

Proof. First, note that $u(x)$ is bounded from above by the right side of (1.8.40), as follows immediately from the Lax-Oleinik formula (1.8.39). We need to show that equality is actually attained. Let us fix $x \in \mathbb{T}^n$, take $t < 0$ large and negative, and consider the corresponding minimizer $\gamma_t(s)$, calibrated by u, so that

$$u(x) = u(\gamma_t(s)) + \int_s^0 \Big(\frac{|\dot{\gamma}_t(\sigma)|^2}{4} + f(\gamma_t(\sigma))\Big)d\sigma, \quad \text{for all } t \leq s \leq 0. \tag{1.8.41}$$

The uniform bounds on $\gamma_t(s)$ and $\dot{\gamma}_t(s)$ imply that there is a sequence $t_n \to -\infty$ such that $\gamma_{t_n}(s)$ converges, locally uniformly, to a limit $\gamma(s)$ that is defined for all $s < 0$. Passing to the limit $t_n \to -\infty$ in (1.8.41) we see that $\gamma(s)$ is also calibrated by u: for all $s < 0$ we have

$$u(x) = u(\gamma(s)) + \int_s^0 \Big(\frac{|\dot{\gamma}(\sigma)|^2}{4} + f(\gamma(\sigma))\Big)d\sigma. \tag{1.8.42}$$

We claim that there exists $x_k \in \mathcal{Z}$ so that

$$\lim_{s\to-\infty} \gamma(s) = x_k. \tag{1.8.43}$$

To see that (1.8.43) holds, take $\varepsilon > 0$ and consider the set

$$D_\varepsilon = \{y \in \mathbb{T}^n : \ |y - x_i| \leq \varepsilon \text{ for some } x_i \in \mathcal{Z}\}.$$

If $\varepsilon > 0$ is sufficiently small, then D_ε is a union of N pairwise disjoint balls

$$B_\varepsilon^{(k)} = \{y \in \mathbb{T}^n : \ |y - x_k| \leq \varepsilon\}.$$

The function $f(y)$ is strictly positive outside of D_ε: there exists $\lambda_\varepsilon > 0$ so that $f(y) > \lambda_\varepsilon$ for all $y \notin D_\varepsilon$. It follows from (1.8.42) that the total time that $\gamma(s)$ spends outside of D_ε is also bounded:

$$|\{s < 0 : \ \gamma(s) \notin D_\varepsilon\}| \leq \frac{2\|u\|_{L^\infty}}{\lambda_\varepsilon}. \tag{1.8.44}$$

Exercise 1.8.7 Show that there exists $\mu_\varepsilon > 0$ such that if $s_1 < s_2 < 0$, and $\gamma(s_1) \in B_\varepsilon^{(k)}$ while $\gamma(s_2) \in B_\varepsilon^{(k')}$ with $k \neq k'$, then

$$\int_{s_1}^{s_2} \Big(\frac{|\dot{\gamma}(s)|^2}{4} + f(\gamma(s))\Big)ds \geq \mu_\varepsilon. \tag{1.8.45}$$

Hint: show that if the switch from $B_\varepsilon^{(k)}$ to $B_\varepsilon^{(k')}$ happens "quickly" then the contribution of the first term inside the integral is bounded from below, and if this switch happens "slowly", then the contribution of the second term inside the integral is bounded from below.

A consequence of (1.8.44) and Exercise 1.8.7 is that there exists T_ε and $1 \leq k \leq N$ such that $\gamma(s) \in B_\varepsilon(x_k)$ for all $t < T_\varepsilon$. This implies (1.8.43).

Now, we may let $s \to -\infty$ in (1.8.42) to obtain

$$u(x) = u(x_k) + \int_{-\infty}^0 \Big(\frac{|\dot{\gamma}(\sigma)|^2}{4} + f(\gamma(\sigma))\Big)d\sigma. \tag{1.8.46}$$

It follows that $u(x)$ is bounded from below by the right side of (1.8.40), and the proof of Proposition 1.8.6 is complete. □

Classification of steady solutions

We can now classify all solutions to

$$|\nabla u|^2 = f(x), \quad x \in \mathbb{T}^n. \tag{1.8.47}$$

The reader may remember that the proof of uniqueness of the waves in Theorems 1.3.1 and the long time behavior in Theorem 1.3.4 in the viscous case relied crucially on the strong maximum principle and the Harnack inequality for parabolic equations. It is exactly the lack of these properties for the inviscid Hamilton-Jacobi equations that leads to the non-uniqueness of the solutions to (1.8.88), and to different possible long time behaviors of the solutions to the corresponding Cauchy problem.

Let us set

$$S(x_i, x) = \inf_{\gamma(-\infty)=x_i} \int_{-\infty}^{0} \Big(\frac{|\dot{\gamma}(s)|^2}{4} + f(\gamma(s)) \Big) ds. \tag{1.8.48}$$

It may be seen as the energy of a connection between x_i and x, or, in a more mathematically precise way, as a sort of distance between x_i and x. This fruitful point of view, developed in [63], will not be pushed further here. The next theorem classifies all solutions to (1.8.47).

Theorem 1.8.8 *Let $\{x_1, \ldots, x_N\}$ be the set of zeros of a smooth non-negative function $f(x)$. Given a collection of numbers $\{a_1, ..., a_N\}$ there is a unique solution $u(x)$ to (1.8.47), such that*

$$u(x_i) = a_i \text{ for all } 1 \leq i \leq N, \tag{1.8.49}$$

if and only if

$$a_j \leq a_i + S(x_i, x_j), \quad \text{for all } 1 \leq i, j \leq N. \tag{1.8.50}$$

Condition (1.8.50) has a simple interpretation: in order to be able to assign a value a_j at the zero x_j, the trajectory $\gamma(t) \equiv x_j$ for all $t < 0$, should be a minimizer.

Proof. Proposition 1.8.6 already shows that the values of $u(x_i)$ determine the value of $u(x)$ for all $x \in \mathbb{T}^n$, and that if a solution exists and (1.8.49) holds, then

$$a_j = \inf_{i \in \{1,...,N\}} \big(a_i + S(x_i, x_j) \big). \tag{1.8.51}$$

This implies (1.8.50).

To prove existence of a solution to (1.8.47) such that $u(x_i) = a_i$ for all $1 \leq i \leq N$, for given a_i, $i = 1, \ldots, N$, that satisfy (1.8.50), set

$$u(x) = \inf_{i \in \{1,...,N\}} \big(a_i + S(x_i, x) \big). \tag{1.8.52}$$

Using the by now familiar arguments, it is easy to see that u is a solution to (1.8.47). Moreover, we have

$$u(x_j) = \inf_{i \in \{1,...,N\}} \big(a_i + S(x_i, x_j) \big).$$

This, together with (1.8.50) implies $u(x_j) = a_j$. $\square$

Exercise 1.8.9 Apply the above theorem to the equation $|u'| = f(x)$ on $\mathbb{T}^1$, with a non-negative function $f(x)$ vanishing at 2 or 3 distinct points. Find out how many different solutions one may have.

Exercise 1.8.10 Let Ω be a smooth bounded subset of $\mathbb{R}^n$. Assume that f is nonnegative and vanishes only at a finite number of points and $u_0 \in C(\partial\Omega)$. Find a necessary and sufficient condition on the values of u_0 so that the boundary value problem

$$\begin{aligned} &|\nabla u|^2 = f(x), \quad x \in \Omega, \\ &u(x) = u_0(x), \quad x \in \partial\Omega, \end{aligned} \tag{1.8.53}$$

is well-posed. Count its solutions. If you have difficulty, we recommend that you read the very remarkable study of the non-uniqueness in Lions [93].

1.8.2 The large time behavior: a strictly convex example

The above analysis for the classification of the wave solutions can be adapted to understand the long time behavior of the solutions to the Cauchy problem

$$\begin{aligned} &u_t + |\nabla u|^2 = f(x), \quad t > 0, \; x \in \mathbb{T}^n, \\ &u(0, x) = u_0(x). \end{aligned} \tag{1.8.54}$$

The next theorem gives an (almost) explicit form of the asymptotic limit of the solution to (1.8.54), and exhibits again the role of the set $\mathcal{Z}$ in the dynamics.

Theorem 1.8.11 *Let $u(t, x)$ be the solution to (1.8.54) with a smooth non-negative function $f(x)$ that vanishes on a finite set $\mathcal{Z} = \{x_1, \dots, x_N\}$, and $u_0 \in C(\mathbb{T}^n)$. Then, the function $u(t, x)$ is non-increasing in t on the set $\mathcal{Z}$, so that for each $x_k \in \mathcal{Z}$ the limit*

$$a_k := \lim_{t \to +\infty} u(t, x_k) \tag{1.8.55}$$

exists. Moreover, for all $x \in \mathbb{T}^n$ we have

$$\lim_{t \to +\infty} u(t, x) = \inf_{x_k \in \mathcal{Z}} \big(a_k + S(x_k, x)\big), \tag{1.8.56}$$

with $S(x_i, x)$ as in (1.8.48):

$$S(x_i, x) = \inf_{\gamma(-\infty) = x_i} \int_{-\infty}^{0} \Big(\frac{|\dot\gamma(s)|^2}{4} + f(\gamma(s)) \Big) ds. \tag{1.8.57}$$

We will use throughout the proof the fact that the unique viscosity solution to (1.8.54) is uniformly bounded and is uniformly Lipschitz: there exists $C > 0$ so that for all $t \geq 1$ we have

$$\|u(t, \cdot)\|_{L^\infty} \leq C, \quad \|u_t(t, \cdot)\|_{L^\infty} + \|\nabla u(t, \cdot)\|_{L^\infty} \leq C. \tag{1.8.58}$$

The Lipschitz bound in (1.8.58) follows from Theorem 1.7.11, while the uniform bound on $u(t, x)$ is a simple consequence of the fact that steady solutions to (1.8.54) exist under our assumptions on $f(x)$. These estimates already tell us that there exists a sequence $t_n \to +\infty$ such that the sequence of functions $v_n(t, x) = u(t + t_n, x)$ converges in $L^\infty(\mathbb{T}^n)$ and locally uniformly in t, to a limit $\widetilde{u}(t, x)$. However, we do not know that the limit is unique, nor that it is time-independent, nor that it is a solution to (1.8.54).

Monotonicity on $\mathcal{Z}$

We first prove that $u(t,x)$ is non-increasing in t for $x \in \mathcal{Z}$. If $u(t,x)$ were actually smooth at $x_k \in \mathcal{Z}$, then, as $f(x_k) = 0$ for $x_k \in \mathcal{Z}$, we would have

$$u_t(t,x_k) = -|\nabla u(t,x_k)|^2 \le 0, \tag{1.8.59}$$

as desired. However, we only know that $u(t,x)$ is a viscosity solution, hence we can not use (1.8.59) directly. Instead, we fix $t_0 \ge 0$ and consider the function

$$\bar{u}(t,x) = u(t_0,x) + (t - t_0) f(x). \tag{1.8.60}$$

We claim that $\bar{u}(t,x)$ is a viscosity super-solution to (1.8.54). Consider a test function φ such that the difference $\bar{u} - \varphi$ attains its minimum at (t_1,x_1). As $\bar{u}(t,x)$ is smooth in t, we have, in particular, that

$$0 \le \varphi_t(t_1,x_1) - \bar{u}_t(t_1,x_1),$$

which implies

$$0 \le \varphi_t(t_1,x_1) - \bar{u}_t(t_1,x_1) = \varphi_t(t_1,x_1) - f(x_1) \le \varphi_t(t_1,x_1) + |\nabla\varphi(t_1,x_1)|^2 - f(x_1). \tag{1.8.61}$$

We deduce that $\bar{u}(t,x)$ is a super-solution to (1.8.54). Moreover, at $t = t_0$ we have

$$\bar{u}(t_0,x) = u(t_0,x) \quad \text{for all } x \in \mathbb{T}^n.$$

As a consequence, it follows that $u(t,x) \le \bar{u}(t,x)$ for all $t \ge t_0$ and $x \in \mathbb{T}^n$. Specifying this at $x_k \in \mathcal{Z}$ gives

$$u(t,x_k) \le u(t_0,x_k) \text{ for all } t \ge t_0,$$

thus $u(t,x)$ is non-increasing in t on $\mathcal{Z}$, proving the first claim of Theorem 1.8.11: the limit

$$a_k := \lim_{t\to+\infty} u(t,x_k) \tag{1.8.62}$$

exists for all $x_k \in \mathcal{Z}$, $1 \le k \le N$.

Convergence on the whole torus

The proof of the second part of Theorem 1.8.11 is similar to that of Proposition 1.8.6 but some technical points are different. For a fixed $t > 0$ and $x \in \mathbb{T}^n$, consider the Lax-Oleinik formula written as

$$u(t,x) = \inf_{\gamma(t)=x} \Big(u(s,\gamma(s)) + \int_s^t \Big(\frac{|\dot{\gamma}(\sigma)|^2}{4} + f(\gamma(\sigma)) \Big) d\sigma \Big), \tag{1.8.63}$$

with any $0 \le s \le t$. Taking a test curve $\gamma_{s,t}(\sigma)$, $s \le \sigma \le t$ such that $\gamma_{s,t}(s) = x_k \in \mathcal{Z}$, with both s and t large, and passing to the limit $t, s \to +\infty$ with $t - s \to +\infty$, we deduce that for all $x \in \mathbb{T}^n$ we have

$$\limsup_{t\to+\infty} u(t,x) \le \inf_{x_i\in\mathcal{Z}} \big(a_i + S(x_i,x) \big), \tag{1.8.64}$$

with $S(x_i,x)$ defined in (1.8.57). We used here the existence of the limit in (1.8.62).

The longer step is to show the reverse inequality to (1.8.64). Let $\gamma_t(\sigma)$, $0 \le \sigma \le t$, be a minimizer in the Lax-Oleinik formula

$$u(t,x) = \inf_{\gamma(t)=x} \Big(u_0(\gamma(0)) + \int_0^t \Big(\frac{|\dot\gamma(\sigma)|^2}{4} + f(\gamma(\sigma))\Big)d\sigma\Big). \tag{1.8.65}$$

As $\gamma_t(\sigma)$ is calibrated by u, we have

$$\begin{aligned} u(t,x) &= u(t+s,\gamma_t(t+s)) + \int_{t+s}^t \Big(\frac{|\dot\gamma_t(\sigma)|^2}{4} + f(\gamma_t(\sigma))\Big)d\sigma \\ &= u(t+s,\gamma_t(t+s)) + \int_s^0 \Big(\frac{|\dot\gamma_t(t+\sigma)|^2}{4} + f(\gamma_t(t+\sigma))\Big)d\sigma, \quad \text{for all } -t \le s \le 0. \end{aligned} \tag{1.8.66}$$

Let us introduce the path $\eta_t(\sigma) = \gamma_t(t+\sigma)$, $-t \le \sigma \le 0$, and write (1.8.66) as

$$u(t,x) = u(t+s,\eta_t(s)) + \int_s^0 \Big(\frac{|\dot\eta_t(\sigma)|^2}{4} + f(\eta_t(\sigma))\Big)d\sigma, \quad \text{for all } -t \le s \le 0. \tag{1.8.67}$$

We now pass to the limit $t \to +\infty$. The uniform a priori bounds on $\gamma_t(\sigma)$ and $\dot\gamma_t(\sigma)$ imply the corresponding bounds on $\eta_t(\sigma)$ and $\dot\eta_t(\sigma)$. Hence, there exists a sequence $t_n \to +\infty$ such that $\eta_{t_n}(\sigma)$ converges as $n \to +\infty$, locally uniformly in σ, to a limit $\eta(\sigma)$, $-\infty < \sigma \le 0$. In addition, $\eta(s)$ inherits the minimizing property of η_t: for any $s \le 0$, the curve $\eta(\sigma)$ is a minimizer of

$$\int_s^0 \Big(\frac{|\dot\gamma(\sigma)|^2}{4} + f(\gamma(\sigma))\Big)d\sigma,$$

over all curves $\gamma(\sigma)$, $s \le \sigma \le 0$, that connect the point $\gamma(s) = \eta(s)$ to $x = \gamma(0) = \eta(0)$.

By the same token, the bounds (1.8.58)

$$\|u(t,\cdot)\| \le C, \quad \|u_t(t,\cdot)\|_{L^\infty} + \|\nabla u(t,\cdot)\|_{L^\infty} \le C \tag{1.8.68}$$

on the function $u(t,x)$ imply that the sequence

$$v_n(s,x) = u(t_n+s,x),$$

possibly after extracting a subsequence, converges in $L^\infty(\mathbb{T}^n)$ and locally uniformly in s, to a limit $v(s,x)$ such that

$$v(s,x_k) = a_k, \quad \text{for all } x_k \in \mathcal{Z} \text{ and } s \in \mathbb{R}. \tag{1.8.69}$$

The uniformity of the limits of $\eta_t(\sigma)$ and $v_n(s,x)$ and the uniform in t Lipschitz bounds on $u(t,x)$ allow us to pass to the limit $t_n \to +\infty$ in (1.8.67), giving

$$\lim_{t_n\to+\infty} u(t_n,x) = v(0,x) = v(s,\eta(s)) + \int_s^0 \Big(\frac{|\dot\eta(\sigma)|^2}{4} + f(\eta(\sigma))\Big)d\sigma, \quad \text{for all } -\infty \le s \le 0. \tag{1.8.70}$$

As in the proof of Proposition 1.8.6, we deduce from (1.8.70) the boundedness of the integral

$$\int_{-\infty}^0 \Big(\frac{|\dot\eta(\sigma)|^2}{4} + f(\eta(\sigma))\Big)d\sigma < +\infty.$$

This, in turn, as in that proof, implies that there exists $x_j \in \mathcal{Z}$ such that

$$\lim_{s \to -\infty} \eta(s) = x_j.$$

Using (1.8.69) together with the uniform in s Lipschitz bound on $v(s, x)$ we may now pass to the limit $s \to -\infty$ in the right side of (1.8.70) to conclude that

$$\lim_{t_n \to +\infty} u(t_n, x) = a_j + \int_{-\infty}^{0} \Big(\frac{|\dot{\eta}(\sigma)|^2}{4} + f(\eta(\sigma)) \Big) d\sigma. \tag{1.8.71}$$

The minimizing property of $\eta(\sigma)$ implies that

$$\lim_{t_n \to +\infty} u(t_n, x) = a_j + S(x_j, x) \geq \inf_{x_k \in \mathcal{Z}} \big(a_k + S(x_k, x) \big). \tag{1.8.72}$$

Comparing to (1.8.64), we see that

$$\lim_{t_n \to +\infty} u(t_n, x) = \inf_{x_k \in \mathcal{Z}} \big(a_k + S(x_k, x) \big) := u_\infty(x). \tag{1.8.73}$$

On the other hand, as we have seen before, $u_\infty(x)$ is a solution to

$$|\nabla u_\infty|^2 = f(x).$$

The weak contraction property for the viscosity solutions implies that not only we have the limit along a sequence $t_n \to +\infty$ but actually

$$\lim_{t \to +\infty} u(t, x) = u_\infty(x). \tag{1.8.74}$$

This finishes the proof. $\Box$

Exercise 1.8.12 Explain how the weak contraction property is used in the very last step of the proof.

An equation with a drift

The minimizers for the problem

$$u_t + |\nabla u|^2 = f(x),$$

that we have just considered, spend most of their time near one of the finitely many points in the zero set $\mathcal{Z}$ of f. To illustrate a different possible behavior of the minimizers, consider the Cauchy problem

$$\begin{aligned} &u_t + c u_x + u_x^2 = 0, \quad t > 0, \; x \in \mathbb{T}^1, \\ &u(0, x) = u_0(x). \end{aligned} \tag{1.8.75}$$

The Lagrangian corresponding to the Hamiltonian

$$H(p) = |p|^2 + cp \tag{1.8.76}$$

is

$$L(v) = \sup_{p \in \mathbb{R}} \; [pv - cp - p^2] = \frac{(v - c)^2}{4}, \tag{1.8.77}$$

and the solution to (1.8.75) is given by the Lax-Oleinik formula:

$$u(t,x) = \inf_{\gamma(0)=y,\ \gamma(t)=x} \Big[u_0(y) + \frac{1}{4}\int_0^t (\dot\gamma(s) - c)^2 ds\Big]. \tag{1.8.78}$$

It is easy to see that the minimizer $\gamma_t(s;x)$ for (1.8.78) is a straight line $\gamma(s) = x + c_t(s-t)$. The optimal speed c_t is given by

$$c_t = \text{argmin}_{v\in\mathbb{R}}\Big[u_0(x - vt) + \frac{t}{4}(v-c)^2\Big]. \tag{1.8.79}$$

This is a very different behavior from that in Theorem 1.8.11: the minimizers visit every point on the torus infinitely many times. An immediate consequence of (1.8.78) is that

$$u(t,x) \ge \min_{y\in\mathbb{T}^n} u_0(y). \tag{1.8.80}$$

On the other hand, if x_0 is a minimum of $u_0(y)$, we can take

$$v = \frac{x - x_0 - [x - x_0 - ct]}{t} \tag{1.8.81}$$

in (1.8.78). Here, $[\xi]$ is the integer part of $\xi \in \mathbb{R}$. This gives

$$x - vt = x_0 + [x - x_0 - ct], \quad u_0(x - vt) = u_0(x_0), \tag{1.8.82}$$

leading to an upper bound

$$u(t,x) \le u_0(x - vt) + \frac{t(v-c)^2}{4} \le u_0(x_0) + \frac{1}{4t} = \min_{y\in\mathbb{T}^n} u_0(y) + \frac{1}{4t}. \tag{1.8.83}$$

We deduce from (1.8.80) and (1.8.83) that

$$\lim_{t\to+\infty} u(t,x) = \min_{y\in\mathbb{T}^n} u_0(y), \tag{1.8.84}$$

uniformly in $x \in \mathbb{T}^n$. Note that (1.8.84) holds even though the minimizers do not spend any more time near the minima of $u_0(y)$ than at any other points. Thus, the specific behavior of the minimizers we have seen in Theorem 1.8.11 is helpful but is not needed for the long time limit of the solution to exist. We will revisit this issue in a more general setting in Section 1.9.

1.8.3 The large time behavior: without the Lax-Oleinik formula

We now turn to the long time behavior of the solutions to the Cauchy problem (1.8.3):

$$\begin{aligned} &u_t + R(x)\sqrt{1 + |\nabla u|^2} = 0, \quad t > 0,\ x \in \mathbb{T}^n, \\ &u(0,x) = u_0(x). \end{aligned} \tag{1.8.85}$$

Let us recall that we assume that the function $R(x)$ is smooth and non-negative:

$$R(x) \ge R_0 > 0 \text{ for all } x \in \mathbb{T}^n, \tag{1.8.86}$$

and we use the notation

$$\bar{R} = \|R\|_{L^\infty}. \tag{1.8.87}$$

We will assume for simplicity that the set $\mathcal{Z}$ where $R(x)$ attains its maximum is finite, though this assumption may be very much relaxed. As we have seen in the discussion following Proposition 1.8.1, this problem admits wave solutions of the form $ct + u_\infty(x)$, moving with the speed $c = \bar{R}$. Our goal will be to prove the following long time behavior result.

Theorem 1.8.13 *Let $u(t,x)$ be the solution to (1.8.85) with $u_0 \in C(\mathbb{T}^n)$ and assume that $R(x)$ is smooth, satisfies (1.8.86), and attains its maximum on a finite set. There is a solution $u_\infty(x)$ to*

$$R(x)\sqrt{1 + |\nabla u_\infty|^2} = \bar{R}, \quad x \in \mathbb{T}^n, \tag{1.8.88}$$

such that we have, uniformly with respect to $x \in \mathbb{T}^n$:

$$\lim_{t\to+\infty} \big(u(t,x) + t\bar{R} - u_\infty(x)\big) = 0, \tag{1.8.89}$$

with $\bar{R}$ defined in (1.8.87).

Note that there is no claim of uniqueness of the solutions to (1.8.88) in Theorem 1.8.13, even up to addition of a constant. Indeed, as we have seen, uniqueness need not hold, as soon as the function $R(x)$ attains its maximum at more than one point. Unlike in the strictly convex case considered in the previous section, we will not use the Lax-Oleinik formula to understand the long time behavior, to illustrate the fact that the strict convexity of the Hamiltonian is also not needed for the solutions to have a long time limit. Nevertheless, the set

$$\mathcal{Z} = \{x \in \mathbb{T}^n : \ R(x) = \bar{R}\} \tag{1.8.90}$$

will play an important role in the proof, and in the dynamics, very similar to that of the minima of the function $f(x)$ in the proofs of Theorems 1.8.8 and 1.8.11.

We start the proof of Theorem 1.8.13 by writing

$$u(t,x) = v(t,x) + t\bar{R},$$

which transforms (1.8.85) into

$$\begin{aligned} &v_t + R(x)\sqrt{1 + |\nabla v|^2} - \bar{R} = 0, \quad x \in \mathbb{T}^n \\ &v(0,x) = u_0(x). \end{aligned} \tag{1.8.91}$$

Our goal is to show that there is a solution $u_\infty(x)$ to

$$R(x)\sqrt{1 + |\nabla u_\infty|^2} = \bar{R}, \quad x \in \mathbb{T}^n, \tag{1.8.92}$$

such that

$$\lim_{t\to+\infty} v(t,x) = u_\infty(x), \quad \text{uniformly in } x \in \mathbb{T}^n. \tag{1.8.93}$$

It is easy to see from the weak contraction principle that we may assume without loss of generality that the initial condition $u_0 \in C^1(\mathbb{T}^n)$. As a technical remark, we have seen that the unique viscosity solution to (1.8.91) is uniformly bounded and uniformly Lipschitz: there exists $C > 0$ so that for all $t \geq 1$ we have

$$\|v(t,\cdot)\| \leq C, \quad \|v_t(t,\cdot)\|_{L^\infty} + \|\nabla v(t,\cdot)\|_{L^\infty} \leq C. \tag{1.8.94}$$

The uniform bound on v in (1.8.94) follows from the existence of a steady solution to (1.8.92) and the comparison principle, and the Lipschitz bound is an implication of Theorem 1.7.11. These bounds will be useful again when we pass to the limit $t \to +\infty$.

Note that if we can show that $v(t,x)$ converges uniformly, as $t \to +\infty$, to a limit $u_\infty(x)$, as in (1.8.93), then the limit is a viscosity solution to (1.8.92). Indeed, in that case the functions $v_n(t,x) = v(t+n,x)$ are solutions to (1.8.91), and converge, as $n \to +\infty$, to $u_\infty(x)$, in $L^\infty(\mathbb{T}^n)$, and locally uniformly in t. The stability property of the viscosity solutions implies that $u_\infty(x)$ is a steady solution to (1.8.91), and thus solves (1.8.92). Thus, it suffices to prove that the limit in (1.8.93) exists. We will do this in two steps: first we will prove existence of the limit for $x \in \mathcal{Z}$, and then show that convergence on $\mathcal{Z}$ implies converges on $\mathbb{T}^n \setminus \mathcal{Z}$ as well. In other words, what happens on $\mathcal{Z}$ controls the behavior off $\mathcal{Z}$. This is very similar to the dynamics in Theorem 1.8.11 even though unlike in that case we will not use the Lax-Oleinik minimizers.

Convergence on $\mathcal{Z}$

To show convergence on $\mathcal{Z}$, we are going to prove that $v(t,x)$ is non-increasing in t on $\mathcal{Z}$. This is intuitively obvious: if $v(t,x)$ is continuously differentiable at $x \in \mathcal{Z}$ at some time $t > 0$, so that (1.8.91) holds in the classical sense, then, as $R(x) = \bar{R}$ for $x \in \mathcal{Z}$, we have

$$v_t(t,x) = \bar{R}\big(1 - \sqrt{1 + |\nabla v(t,x)|^2}\big) \leq 0,$$

so that $v(t,x)$ is non-increasing in t. The familiar problem is that $v(t,x)$ is merely Lipschitz, and not necessarily differentiable, hence (1.8.91) holds only almost everywhere, and we have no guarantee that it holds at any given (t,x).

To make the above simple reasoning rigorous, the argument is close to the corresponding step in the proof of Theorem 1.8.11: consider $t_0 > 0$ and $x_0 \in \mathcal{Z}$ and set

$$\overline{v}(t,x) = v(t_0,x) + (t - t_0)(\bar{R} - R(x)).$$

We claim that $\overline{v}$ is a super-solution to (1.8.91) on $[t_0, +\infty) \times \mathbb{T}^n$, such that

$$\overline{v}(t_0,x) = v(t_0,x) \text{ for all } x \in \mathbb{T}^n. \tag{1.8.95}$$

The latter follows immediately from the definition of $\overline{v}(t,x)$. To see the super-solution property, consider a test function $\varphi(t,x)$, and let $(t_1,x_1) \in [t_0,+\infty)$ be a minimum point for $\overline{v} - \varphi$. Since $\overline{v}(t,x)$ is smooth in t, we have

$$0 \leq \varphi_t(t_1,x_1) - \overline{v}_t(t_1,x_1) = \varphi_t(t_1,x_1) + R(x_1) - \bar{R}. \tag{1.8.96}$$

Hence, we have

$$\varphi_t(t_1,x_1) + R(x_1)\sqrt{1 + |\nabla\varphi(t_1,x_1)|^2} - \bar{R} \geq \varphi_t(t_1,x_1) + R(x_1) - \bar{R} \geq 0. \tag{1.8.97}$$

This proves the super-solution property of $\overline{v}(t,x)$. Together with (1.8.95), this implies

$$v(t,x) \le \overline{v}(t,x) \text{ for } (t,x) \in [t_0,+\infty) \times \mathbb{T}^n. \tag{1.8.98}$$

As $R(x_0) = \bar{R}$ for $x_0 \in \mathcal{Z}$, we obtain

$$v(t,x_0) \le v(t_0,x_0), \quad \text{for all } t \ge t_0 \text{ and } x_0 \in \mathcal{Z}. \tag{1.8.99}$$

Since t_0 is arbitrary, it follows that $v(t,x_0)$ is non-increasing in t. As a consequence, for each $x \in \mathcal{Z}$ the limit

$$u_\infty(x) = \lim_{t\to+\infty} v(t,x)$$

exists.

Exercise 1.8.14 Show that for any $\delta > 0$ we can find t_δ such that, for all $x \in \mathcal{Z}$, $h > 0$ and $t \ge t_\delta$ we have

$$0 \le v(t,x) - v(t+h,x) \le \delta. \tag{1.8.100}$$

Convergence outside of $\mathcal{Z}$

The heart of the proof is to show that convergence of $v(t,x)$ as $t \to +\infty$ on the set $\mathcal{Z}$ forces the convergence off $\mathcal{Z}$ as well, without the use of the Lax-Oleinik minimizers. Instead, the large time convergence of $v(t,x)$ outside of $\mathcal{Z}$ will follow from the fact that a transform of v solves a Hamilton-Jacobi equation that is more complex than (1.8.91), but that has the merit of carrying an absorption term. We will use the Kruzhkov transform:

$$w(t,x) = -e^{-v(t,x)}. \tag{1.8.101}$$

Because of the L^∞ and gradient bounds for the Lipschitz function v, the function w is also Lipschitz and satisfies L^∞ and gradient bounds of the same type, and, in particular, we have

$$w_t = |w|v_t = -wv_t, \quad \nabla w = |w|\nabla v = -w\nabla v.$$

Moreover, because the function $v \mapsto -e^{-v}$ is increasing in v, the function w is a viscosity solution to

$$w_t + R(x)\sqrt{w^2 + |\nabla w|^2} = -\bar{R}w, \tag{1.8.102}$$

which can be written as

$$w_t + R(x)\frac{|\nabla w|^2}{|w| + \sqrt{w^2 + |\nabla w|^2}} + (\bar{R} - R(x))w = 0, \quad t > 0,\ x \in \mathbb{T}^n. \tag{1.8.103}$$

The last term in the left side of (1.8.103) is the aforementioned absorption that will eventually save the day.

Exercise 1.8.15 Show that if $z(t,x)$ is a viscosity solution to

$$z_t + H(x, \nabla z) = 0,$$

and the function $G(z)$ is increasing, then $\zeta = G(z)$ is a viscosity solution to

$$\zeta_t + \frac{1}{Q'(\zeta)} H(x, Q'(\zeta)\nabla\zeta) = 0.$$

Here, $Q(\zeta)$ is the inverse function of $G(z)$. Is this necessarily true if the function G is not monotonic?

Let $\mathcal{Z}_\delta$ be the closed set of all points that are at distance at most $\delta > 0$ from $\mathcal{Z}$. Under our simplifying assumption that the set $\mathcal{Z}$ is finite, the set $\mathcal{Z}_\delta$ is a finite union of closed balls. The uniform bounds on ∇v, together with the result of Exercise 1.8.14 imply that there is $C > 0$ so that

$$|w(t,x) - w(t+h,x)| \leq C\delta \text{ for } t \geq t_\delta \text{ and } x \in \mathcal{Z}_\delta. \tag{1.8.104}$$

Our task is now to extend this inequality outside of $\mathcal{Z}_\delta$. Note that there is $\rho_\delta > 0$ such that

$$\bar{R} - R(x) \geq \rho_\delta \quad \text{for } x \text{ outside } \mathcal{Z}_\delta,$$

meaning that the pre-factor in the last term in the left side of (1.8.103) is uniformly positive off $\mathcal{Z}_\delta$. Intuitively this means that the dynamics for w outside of $\mathcal{Z}_\delta$ is "uniformly absorbing". Let us set

$$\underline{w}_\delta(t,x) = w(t+h,x) - C\delta - \|w(t_\delta,\cdot)\|_{L^\infty} e^{-\rho_\delta(t-t_\delta)}, \quad t \geq t_\delta,\ x \notin \mathcal{Z}_\delta. \tag{1.8.105}$$

To show that (1.8.104) holds outside of $\mathcal{Z}_\delta$, we are going to prove that $\underline{w}(t,x)$ is a sub-solution to (1.8.103) for $t \geq t_\delta$, and $x \notin \mathcal{Z}_\delta$, and, in addition,

$$\underline{w}_\delta(t,x) \leq w(t,x) \text{ for } (t,x) \in [t_\delta, +\infty) \times \mathcal{Z}_\delta, \text{ and } t = t_\delta,\ x \in \mathbb{T}^n. \tag{1.8.106}$$

This will imply $w(t,x) \geq \underline{w}_\delta(t,x)$ for $t \geq t_\delta$ and $x \notin \mathcal{Z}_\delta$, which, in turn, entails

$$w(t,x) \geq w(t+h,x) - C(\delta + e^{-\rho_\delta(t-t_\delta)}), \quad \text{for } t \geq t_\delta,\ x \in \mathbb{T}^n \text{ and } h > 0. \tag{1.8.107}$$

Since $\rho_\delta > 0$ is positive, and $\delta > 0$ is arbitrary, this implies the pointwise convergence of $w(t,x)$ to a limit $w_\infty(x)$ as $t \to +\infty$, and, consequently, its uniform convergence that follows from the Lipschitz bound on $w(t,x)$. Therefore, the function $v(t,x)$ also converges to a limit

$$v_\infty(x) = -\log(-w_\infty(x)),$$

as $t \to +\infty$. Note that the absorbing nature of the dynamics for w exhibits itself in the fact that $\rho_\delta > 0$ outside of $\mathcal{Z}_\delta$ – this is why the Kruzhkov transform is helpful here.

Thus, to finish the proof of Theorem 1.8.13, we only need to show that $\underline{w}_\delta$ is a sub-solution to (1.8.103) for $t \geq t_\delta$, and $x \notin \mathcal{Z}_\delta$, and check that (1.8.106) holds. We see from (1.8.104) that

$$w(t,x) \geq \underline{w}_\delta(t,x) \text{ for } t \geq t_\delta \text{ and } x \in \mathcal{Z}_\delta. \tag{1.8.108}$$

At the time $t = t_\delta$ we have

$$w(t_\delta,x) - \underline{w}_\delta(t_\delta,x) = w(t_\delta,x) + \|w(t_\delta,\cdot)\|_{L^\infty} + C\delta - w(t_\delta+h,x) \geq C\delta > 0, \quad \text{for all } x \in \mathbb{T}^n. \tag{1.8.109}$$

We used here the fact that $w(t,x) \leq 0$ for all $t > 0$ and $x \in \mathbb{T}^n$. Putting (1.8.108) and (1.8.109) together, we conclude that (1.8.106) does hold.

It remains to check the sub-solution property for $\underline{w}_\delta$, outside of $\mathcal{Z}_\delta$. Let φ be a test function and (t_1,x_1) a minimum point of $\varphi - \underline{w}_\delta$, with $x_1 \notin \mathcal{Z}_\delta$. In other words, (t_1,x_1) is a minimum point of the function

$$\big(\varphi(t,x) + C\delta + \|w(t_\delta,\cdot)\|_{L^\infty} e^{-\rho_\delta(t-t_\delta)}\big) - w(t,x+h).$$

As w is a viscosity solution to (1.8.103):

$$w_t + R(x)\frac{|\nabla w|^2}{|w| + \sqrt{w^2 + |\nabla w|^2}} + (\bar{R} - R(x))w = 0, \quad t > 0,\ x \in \mathbb{T}^n, \tag{1.8.110}$$

we deduce that the following inequality holds at (t_1, x_1):

$$\varphi_t - \|w(t_\delta, \cdot)\|_{L^\infty}\rho_\delta e^{-\rho_\delta(t-t_\delta)} + R(x_1)\frac{|\nabla\varphi|^2}{|\tau_h w| + \sqrt{(\tau_h w)^2 + |\nabla\varphi|^2}} + (\bar{R} - R(x_1))\tau_h w \le 0. \tag{1.8.111}$$

Here, we have set $\tau_h w(t,x) = w(t+h,x)$. The definition of $\underline{w}_\delta$ implies that

$$\tau_h w(t,x) \ge \underline{w}_\delta(t,x) \quad \text{ for } t \ge t_\delta \text{ and } x \notin \mathcal{Z}_\delta,$$

so that

$$|\tau_h w(t,x)| \le |\underline{w}_\delta(t,x)|. \tag{1.8.112}$$

Also, as

$$\bar{R} - R(x) \ge \rho_\delta \text{ for } x \notin \mathcal{Z}_\delta,$$

we have

$$\begin{aligned}
&(\bar{R} - R(x_1))\tau_h w(t,x_1) - \|w(t_\delta,\cdot)\|_{L^\infty}\rho_\delta e^{-\rho_\delta(t-t_\delta)} \\
&\ge (\bar{R} - R(x_1))w(\tau + h, x_1) - (\bar{R} - R(x_1))\|w(t_\delta,\cdot)\|_{L^\infty}e^{-\rho_\delta(t-t_\delta)} \\
&= (\bar{R} - R(x_1))[\underline{w}_\delta(t,x_1) + C\delta] \ge (\bar{R} - R(x_1))\underline{w}_\delta(t,x_1).
\end{aligned} \tag{1.8.113}$$

Using the inequalities (1.8.112) and (1.8.113) in (1.8.111) leads to

$$\varphi_t + R(x_1)\frac{|\nabla\varphi|^2}{|\underline{w}_\delta| + \sqrt{(\underline{w}_\delta)^2 + |\nabla\varphi|^2}} + (\bar{R} - R(x_1))\underline{w}_\delta \le 0, \tag{1.8.114}$$

at (t_1, x_1). This is the desired viscosity sub-solution inequality for $\underline{w}_\delta$. Thus, $\underline{w}_\delta(t,x)$ is, indeed, a sub-solution to (1.8.103) for $t \ge t_\delta$, and $x \notin \mathcal{Z}_\delta$, This finishes the proof of Theorem 1.8.13. □

Exercise 1.8.16 Carry out the same analysis for the equation

$$u_t + |\nabla u| = f(x), \quad t > 0,\ x \in \mathbb{T}^n,$$

where $f \in C(\mathbb{T}^n)$ satisfies the usual assumptions of this section: continuous, nonnegative, with a nontrivial zero set.

1.9 Convergence of the Lax-Oleinik semigroup

In this section, we prove that the solutions of

$$\begin{aligned}
&u_t + H(x, \nabla u) = 0, \quad x \in \mathbb{T}^n, \\
&u(0,x) = u_0(x)
\end{aligned} \tag{1.9.1}$$

converge to a wave solution as $t \to +\infty$, under the assumptions of uniform strict convexity of the Hamiltonian. So far, we have seen a very particular mechanism for convergence: the dynamics on a special set dictates in turn the convergence in the area where the equation is coercive. This was the zero set of the function $f(x)$ in Theorem 1.8.11, and the set $\mathcal{Z}$ where the function $R(x)$ attains its maximum in the proof of Theorem 1.8.13. In both cases, the minimizers spend most of their time near this "controlling' set. On the other hand, in the example following Theorem 1.8.11, we have seen a situation where the minima of the initial condition $u_0(y)$ dictate the long time behavior, even though the minimizing curves do not spend any extra time near these points.

It turns out that the existence of such "controlling" set is a general fact. For a general Hamilton-Jacobi equation of the type (1.9.1), there is a set where the extremals associated to the wave solutions accumulate, and which orchestrates the convergence to a steady solution. The reader who wishes to learn more may consult [61] or [62], where their general theory by Fathi is exposed. Our goal here is much more modest: we want to identify a set where, following the ideas of the preceding section, the dynamics of u will dictate the behavior on the whole torus. The following theorem is due to Fathi [61] but we present an alternative proof inspired by [125].

Theorem 1.9.1 *Let $H(x,p)$ be smooth and uniformly strictly convex in p:*

$$\alpha I \le D_p^2 H(x,p), \text{ in the sense of quadratic forms,}$$

and c be the corresponding wave speed: there exists a solution $u_\infty(x)$ to

$$H(x, \nabla u_\infty) = c, \; x \in \mathbb{T}^n. \tag{1.9.2}$$

Then, for any given $u_0 \in C(\mathbb{T}^n)$, there exists a solution $u_\infty(x)$ to (1.9.2) such that the solution to the Cauchy problem

$$\begin{aligned} &u_t + H(x, \nabla u) = 0, \quad x \in \mathbb{T}^n, \\ &u(0,x) = u_0(x) \end{aligned} \tag{1.9.3}$$

converges to $u_\infty(x)$ as $t \to +\infty$:

$$\lim_{t\to+\infty} \|u(t,\cdot) + ct - u_\infty(\cdot)\|_{L^\infty} = 0. \tag{1.9.4}$$

We will assume throughout the proof, without loss of generality, that $c = 0$. Otherwise, we would simply replace the Hamiltonian $H(x,p)$ by $H(x,p) - c$.

As usual, existence of the steady solutions implies that there exists $C_0 > 0$ such that

$$|u(t,x)| \le C_0, \quad \text{for all } t > 0 \text{ and } x \in \mathbb{T}^n. \tag{1.9.5}$$

In addition, we have the uniform Lipschitz bound:

$$\mathrm{Lip}_{t,x}[u] \le C, \quad \text{for all } t \ge 1 \text{ and } x \in \mathbb{T}^n. \tag{1.9.6}$$

Here, $\mathrm{Lip}_{t,x}[u]$ is the Lipschitz constant of u both in the t and x variables. These uniform bounds show that, at least along a subsequence, the uniform limits

$$u_\infty(t,x) = \lim_{t_n\to+\infty} u(t+t_n, x) \tag{1.9.7}$$

exist. Our goal is to show that there is actually a limit in (1.9.7) that does not depend on the sub-sequence, this limit is time-independent, and is a solution to the steady equation

$$H(x, \nabla u_\infty) = 0.$$

Before going directly into the proof of Theorem 1.9.1, we would like to explain the construction of the set $\mathcal{Z}$, and what sort of monotonicity we can use for the proof, as we did in the proofs of Theorem 1.8.11 and Theorem 1.8.13. This will require the notion of the ω-limit set of a solution, and that is where we will start the discussion. After introducing these objects and discussing their basic properties we will turn to the bona fide proof that will be quite short once we have obtained the desired properties of the set $\mathcal{Z}$.

The ω-limit set

The ω-limit set for a given initial condition $u_0 \in C(\mathbb{T}^n)$ with respect to the Lax-Oleinik semigroup is denoted by $\omega(u_0) \subset C(\mathbb{T}^n)$, and is constructed as follows. The uniform bounds (1.9.5) and (1.9.6) imply that there is a sequence $t_n \to +\infty$ such that the family $v_n(t,x) = u(t+t_n, x)$ converges :

$$v_n(t,x) = u(t+t_n, x) \to v(t,x), \tag{1.9.8}$$

in $L^\infty(\mathbb{T}^n)$ and uniformly on compact intervals of $t \in \mathbb{R}$. The function $v(t,x)$ is a solution to

$$v_t + H(x, \nabla v) = 0, \tag{1.9.9}$$

defined for all $t \in \mathbb{R}$, and not just for $t > 0$. Sometimes such solutions are called "entire in time", to indicate that they are also defined for negative times. The set $\omega(u_0)$ consists of all functions $\psi \in C(\mathbb{T}^n)$ such that there exists a sequence $t_n \to +\infty$ and the corresponding limit $v(t,x)$ so that (1.9.8) holds, and

$$\psi(x) = v(0,x) = \lim_{n\to+\infty} u(t_n, x). \tag{1.9.10}$$

An important observation is that if $\psi \in \omega_0(u_0)$ then the action of the Lax-Oleinik semigroup $\mathcal{T}(t)\psi$ is defined for all $t \in \mathbb{R}$, and not only for $t > 0$, via

$$\mathcal{T}(t)\psi(x) = v(t,x) = \lim_{n\to+\infty} \mathcal{T}(t_n + t)u_0(x), \quad \text{for } t \in \mathbb{R}. \tag{1.9.11}$$

Here, t_n is the sequence in (1.9.8). Note that $t_n + t > 0$ for n large enough, so that the action of $\mathcal{T}(t_n+t)$ in the right side above is well-defined even for $t < 0$. Taking the sequence $t_n' = t_n+s$ we see that if $\psi \in \omega(u_0)$ then $\mathcal{T}(s)\psi \in \omega(u_0)$ as well, for any $s \in \mathbb{R}$.

Exercise 1.9.2 (i) Assume that there are two sequences $t_n \to +\infty$ and $s_n \to +\infty$, and the corresponding limits

$$v(t,x) = \lim_{n\to+\infty} u(t_n + t, x), \quad w(t,x) = \lim_{t\to+\infty} u(s_n + t, x),$$

such that $v(0,x) = w(0,x) = \psi(x)$. Show that then $v(t,x) = w(t,x)$ for all $t \in \mathbb{R}$. This shows that the above definition of $\mathcal{T}(t)\psi$ does not depend on the choice of a sequence t_n such that (1.9.8) and (1.9.10) hold.
(ii) Show that for any $t, s \in \mathbb{R}$ and $\psi \in \omega(u_0)$ we have

$$\mathcal{T}(t)\mathcal{T}(s)\psi = \mathcal{T}(t+s)\psi. \tag{1.9.12}$$

The claim of Theorem 1.9.1 can be now reformulated as saying that for each $\psi \in \omega(u_0)$ the function $v(t,x) = \mathcal{T}(t)\psi(x)$ does not depend on t, and that $\omega(u_0)$ contains exactly one function ψ. The following exercise gives a sufficient condition for this to be true.

Exercise 1.9.3 (i) Assume that $\psi \in \omega(u_0)$ is such that $v(t,x) = \mathcal{T}(t)\psi(x)$ does not depend on t. Show that then $v(x)$ is a viscosity solution to

$$H(x, \nabla v) = 0. \tag{1.9.13}$$

(ii) Show that if there exists $\psi \in \omega(u_0)$ that satisfies the assumptions of part (i), then the limit

$$\lim_{t\to+\infty} \mathcal{T}(t)u_0,$$

exists, is unique, and is a viscosity solution to (1.9.13). Hint: use the contraction property for the solutions to (1.9.3) to show that if the ω-limit set of u_0 contains a time-independent solution $v(x)$ to (1.9.13), then v is the only element of $\omega(u_0)$.

Exercise 1.9.3 gives us a blueprint for the proof of Theorem 1.9.1: it suffices to show that for any $\psi \in \omega(u_0)$ the function $v(t,x) = \mathcal{T}(t)\psi(x)$ does not depend on t. As in the proof of Theorem 1.8.13, we will first identify a set $\mathcal{Z}$ where it is easier to show that $v(t,x)$ is time-independent, and then show this outside of $\mathcal{Z}$.

Monotonicity along the minimizers

Our first goal is to recycle the main idea of the proofs of Theorems 1.8.11 and 1.8.13, namely, to find a set where convergence will hold because of some monotonicity property. The following easy remark can be made: let $\phi(x)$ be a steady solution to

$$H(x, \nabla\phi) = 0, \tag{1.9.14}$$

and $\gamma : [0,t] \to \mathbb{T}^n$ be an extremal path calibrated by ϕ. For all $0 \le s \le s' \le t$ we have

$$\phi(\gamma(s')) = \phi(\gamma(s)) + \int_s^{s'} L(\gamma(\sigma), \dot\gamma(\sigma))d\sigma, \tag{1.9.15}$$

whereas, by the definition of the Lax-Oleinik semigroup we have

$$u(s', \gamma(s')) \le u(s, \gamma(s)) + \int_s^{s'} L(\gamma(\sigma), \dot\gamma(\sigma))d\sigma. \tag{1.9.16}$$

Subtracting, we obtain

$$u(s', \gamma(s')) - \phi(\gamma(s')) \le u(s, \gamma(s)) - \phi(\gamma(s)). \tag{1.9.17}$$

Thus, the difference $u(s,\gamma(s)) - \phi(\gamma(s)$ is non-increasing in s along the extremal path calibrated by ϕ. This simple observation will bear a lot of fruit.

Exercise 1.9.4 Interpret this observation for the problem

$$u_t + cu_x + u_x^2 = 0, \quad x \in \mathbb{T}^1,$$

that we have considered before.

The ω-limits of paths and the set $\mathcal{Z}$

We now use the monotonicity property (1.9.17) to construct a candidate for the set $\mathcal{Z}$. It will contain paths that calibrate all steady solutions but it will also do more. Let us fix a steady solution ϕ. We define $\mathcal{Z}_\phi$ as the collection of all "eternal" extremal paths calibrated by ϕ, the set of all trajectories $\gamma : \mathbb{R} \to \mathbb{T}^n$ such that

$$\phi(\gamma(s')) = \phi(\gamma(s)) + \int_s^{s'} L(\gamma(\sigma), \dot\gamma(\sigma))d\sigma, \quad \text{for all } -\infty < s \le s' < +\infty. \tag{1.9.18}$$

The next exercise shows that eternal extremal paths calibrated by ϕ exist, so the set $\mathcal{Z}_\phi$ is not empty.

Exercise 1.9.5 (i) Fix $x \in \mathbb{T}^n$ and consider a family of extremal paths $\gamma_t(s)$, $t \le s \le 0$, calibrated by ϕ. Show that there is a sequence $t_n \to -\infty$ such that $\gamma_{t_n}(s) \to \gamma(s)$, locally uniformly in $s \le 0$, and $\gamma(0) = x$.
(ii) Let $\gamma(s)$ be constructed as in part (i). Show that there exists a sequence $s_n \to -\infty$ such that the paths $\gamma_n(s) = \gamma(s_n + s)$, $-\infty \le s \le s_n$, converge, locally uniformly on $\mathbb{R}$, to a path $\bar\gamma(s)$, $s \in \mathbb{R}$.
(iii) Show that the path $\bar\gamma(s)$, $s \in \mathbb{R}$, is calibrated by ϕ.

The set $\mathcal{Z}$ is then defined as follows: a point $x \in \mathbb{T}^n$ is in $\mathcal{Z}$ if there is a path $\gamma_\infty : \mathbb{R} \to \mathbb{T}^n$ that passes through x, a path $\gamma \in \mathcal{Z}_\phi$, and a sequence $s_n \to +\infty$ such that

$$\gamma_\infty(s) = \lim_{n\to+\infty} \gamma(s + s_n), \tag{1.9.19}$$

with the limit uniform on every bounded interval of $\mathbb{R}$. In other words, $\mathcal{Z}$ is the union of ω-limits of the paths in $\mathcal{Z}_\phi$.

Exercise 1.9.6 (i) Find the set $\mathcal{Z}$ for the Hamiltonian $H(x,p) = |p|^2 - f(x)$ with a smooth non-negative function $f(x)$, $x \in \mathbb{T}^n$. Does it depend on the steady solution $\phi(x)$ with which you start?
(ii) Find the set $\mathcal{Z}$ for the Hamiltonian $H(p) = |p|^2 + cp$, $p \in \mathbb{T}^1$, with $c > 0$.

Exercise 1.9.7 Show that if $\gamma_\infty(\sigma)$, $\sigma \in \mathbb{R}$ is in $\mathcal{Z}$, then so is the time-shifted path

$$\gamma_\infty^{(s)}(\sigma) = \gamma_\infty(\sigma + s), \quad \sigma \in \mathbb{R},$$

for any $s \in \mathbb{R}$ fixed. Hint: this is because the original solution ϕ, that we used to construct $\mathcal{Z}_\phi$ and then $\mathcal{Z}$, is time-independent, so that a time-shift of a path $\gamma \in \mathcal{Z}_\phi$ that calibrates ϕ also calibrates ϕ, and is therefore in $\mathcal{Z}_\phi$.

Calibration by paths in $\mathcal{Z}$

Let us now take a path $\gamma_\infty(s)$, $s \in \mathbb{R}$ that lies in $\mathcal{Z}$, obtained as the limit in (1.9.19), with a given $\gamma \in \mathcal{Z}_\phi$. Writing

$$\begin{aligned}\phi(\gamma(s' + s_n)) &= \phi(\gamma(s + s_n)) + \int_{s+s_n}^{s'+s_n} L(\gamma(\sigma), \dot\gamma(\sigma))d\sigma \\ &= \phi(\gamma(s + s_n)) + \int_s^{s'} L(\gamma(\sigma + s_n), \dot\gamma(\sigma + s_n))d\sigma,\end{aligned} \tag{1.9.20}$$

and passing to the limit $s_n \to +\infty$, we see immediately that $\gamma_\infty(s)$ is also calibrated by ϕ. The miracle is that $\gamma_\infty(s)$ is also calibrated by every other steady solution $\psi(x)$ to (1.9.14). Indeed, it follows from the monotonicity property (1.9.17) used with $u(t,x) = \psi(x)$ that for any path $\gamma \in \mathcal{Z}_\phi$ the limit

$$\lim_{s\to+\infty}[\psi(\gamma(s)) - \phi(\gamma(s))] = K(\gamma),$$

exists. It follows that if $\gamma(s+s_n) \to \gamma_\infty(s)$ as $s_n \to +\infty$, then the two solutions differ by a constant on γ_∞:

$$\psi(\gamma_\infty(s)) = \phi(\gamma_\infty(s)) + K(\gamma), \quad \text{for all } -\infty < s < +\infty. \tag{1.9.21}$$

As $\gamma_\infty(s)$ is calibrated by ϕ, we conclude from (1.9.21) that it is calibrated by ψ as well. We stress that it is not true that every path in $\mathcal{Z}_\phi$ is calibrated by any other steady solution – this is only true for their ω-limits that form the set $\mathcal{Z}$. The following proposition shows that we can say even more.

Proposition 1.9.8 *Let $\gamma_\infty(s)$, $s \in \mathbb{R}$ be a path in $\mathcal{Z}$ and $\psi \in \omega(u_0)$. There exists $K(\gamma) \in \mathbb{R}$ such that $v(t,x) = \mathcal{T}(t)\psi(x)$ satisfies*

$$v(t, \gamma_\infty(t)) - \phi(\gamma_\infty(t)) = K(\gamma), \quad \textit{for all } t \in \mathbb{R}. \tag{1.9.22}$$

In particular, the path $\gamma_\infty(s)$, $s \in \mathbb{R}$, is calibrated by $v(t,x)$.

Proof. By definition of γ_∞, there is a global extremal path γ calibrated by ϕ, and a sequence $s_n \to +\infty$ such that

$$\gamma_\infty(\sigma) = \lim_{n\to+\infty} \gamma(s_n + \sigma),$$

uniformly in every compact in $\sigma \in \mathbb{R}$. Observe that to prove (1.9.22) it suffices to find a subsequence s_{n_k} such that

$$\lim_{k\to+\infty}[v(s, \gamma(s_{n_k}+s)) - \phi(\gamma(s_{n_k}+s))] = \text{const}, \quad \text{independent of } s \in \mathbb{R}. \tag{1.9.23}$$

Let $u(t,x) = \mathcal{T}(t)u_0(x)$ be the solution to

$$\begin{aligned} &u_t + H(x, \nabla u) = 0, \quad t > 0, \ x \in \mathbb{T}^n, \\ &u(0,x) = u_0(x). \end{aligned} \tag{1.9.24}$$

Since $\psi(x) = v(0,x)$ is in $\omega(u_0)$, we may also find both a sequence $t_k \to +\infty$ and a subsequence $s_{n_k} \to +\infty$ such that

$$v(0,x) = \lim_{k\to+\infty} u(t_k + s_{n_k}, x), \quad \text{uniformly in } x \in \mathbb{T}^n,$$

and for all $s \in \mathbb{R}$ we have

$$v(s,x) = \lim_{k\to+\infty} u(t_k + s_{n_k} + s, x), \ \text{ in } L^\infty(\mathbb{T}^n), \tag{1.9.25}$$

uniformly in every compact in $s \in \mathbb{R}$. Thus, for every compact set $K \subset \mathbb{R}$ and $\varepsilon > 0$ there exists $N_{\varepsilon,K}$ such that for all $k > N_{\varepsilon,K}$ we have

$$|v(s, \gamma(s_{n_k} + s)) - u(t_k + s_{n_k} + s, \gamma(s_{n_k} + s))| < \varepsilon \text{ for all } s \in K. \qquad (1.9.26)$$

Hence, (1.9.23) would follow if we can show that, possibly after a further extraction of a sub-sequence k_j (to avoid too cumbersome notation, we still denote the corresponding subsequences by t_k and s_{n_k}), we have

$$\lim_{k \to +\infty} [u(t_k + s_{n_k} + s, \gamma(s_{n_k} + s)) - \phi(\gamma(s_n + s))] = \text{const}, \quad \text{independent of } s \in \mathbb{R}. \quad (1.9.27)$$

Note that, after passing to a yet another sub-sequence, we may assume that there is $\psi_1 \in \omega(u_0)$ such that

$$\psi_1(\cdot) = \lim_{k \to +\infty} u(t_k, \cdot), \quad \text{in } L^\infty(\mathbb{T}^n). \qquad (1.9.28)$$

Let us denote $\widetilde{v}(t, x) = \mathcal{T}\psi_1(x)$. By the weak contraction property, we have

$$\|u(t_k + t, \cdot) - \widetilde{v}(t, \cdot)\|_{L^\infty} \le \|u(t_k, \cdot) - \psi_1\|_{L^\infty} \to 0, \quad \text{as } k \to +\infty, \qquad (1.9.29)$$

uniformly in $t \ge 0$. Thus, for any $\varepsilon > 0$ we can find N_ε so that for all $k > N_\varepsilon$ we have

$$|u(t_k + s_{n_k} + s, \gamma(s_{n_k} + s)) - \widetilde{v}(s_{n_k} + s, \gamma(s_{n_k} + s))| \le \varepsilon, \qquad (1.9.30)$$

locally uniformly in $s \in \mathbb{R}$. Hence, (1.9.27) would follow if we show that

$$\lim_{k \to +\infty} [\widetilde{v}(s_{n_k} + s, \gamma(s_{n_k} + s)) - \phi(\gamma(s_{n_k} + s))] = \text{const}, \quad \text{independent of } s \in \mathbb{R}. \qquad (1.9.31)$$

However, the monotonicity property (1.9.17) along the extremals implies that the limit

$$\ell = \lim_{\xi \to +\infty} \big(\widetilde{v}(\xi + s, \gamma(\xi + s)) - \phi(\gamma(\xi + s))\big).$$

exists and is independent of s. It follows that

$$\lim_{k \to +\infty} [\widetilde{v}(s_{n_k} + s, \gamma(s_{n_k} + s)) - \phi(\gamma(s_{n_k} + s))] = \ell \quad \text{for all } s \in \mathbb{R}, \qquad (1.9.32)$$

finishing the proof. □

We will also need the following proposition which says that paths calibrated by solutions in $\omega(u_0)$ come arbitrarily close to the set $\mathcal{Z}$ – this is what eventually leads to the fact that the behavior of the solutions on $\mathcal{Z}$ controls the behavior outside of $\mathcal{Z}$ as well.

Proposition 1.9.9 *Let $\psi \in \omega(u_0)$ and $v(t, x) = \mathcal{T}(t)\psi(x)$. Given any $t \in \mathbb{R}$, and a path $\gamma(s)$, defined for $s \le t$, and calibrated by v, there exists a sequence $s_n \to -\infty$ such that dist$(\gamma(s_n), \mathcal{Z}) \to 0$ as $n \to +\infty$.*

Proof. Let $\gamma(s)$, $s \in \mathbb{R}$ be a path calibrated by $v(t, x)$, and $\phi(x)$ be the steady solution used to generate $\mathcal{Z}$. The Lax-Oleinik formula tells us that for any $s < s'$ we have

$$v(s', \gamma(s')) = v(s, \gamma(s)) + \int_s^{s'} L(\gamma(\sigma), \dot{\gamma}(\sigma)) d\sigma,$$

and

$$\phi(\gamma(s')) \leq \phi(\gamma(s)) + \int_s^{s'} L(\gamma(\sigma), \dot{\gamma}(\sigma)) d\sigma.$$

Subtracting we get the monotonicity relation

$$v(s', \gamma(s')) - \phi(\gamma(s')) \leq v(s, \gamma(s)) - \phi(\gamma(s)), \text{ for all } s < s'. \tag{1.9.33}$$

Hence, the limit

$$\lim_{s \to -\infty} [v(s, \gamma(s)) - \phi(\gamma(s)] \tag{1.9.34}$$

exists. The uniform bounds on $\gamma(s)$ and $\dot{\gamma}(s)$ imply that there exists a sequence $s_n \to -\infty$ so that the paths $\gamma_n(t) = \gamma(s + s_n)$ converge, as $n \to +\infty$, to a limit $\gamma_\infty(s)$, $s \in \mathbb{R}$, locally uniformly in s.

Exercise 1.9.10 Show that the path $\gamma_\infty(s)$ is calibrated by any steady solution, in particular, by ϕ.

Exercise 1.9.10 shows that $\gamma_\infty(s)$, $s \in \mathbb{R}$, lies in $\mathcal{Z}_\phi$ but we do not yet know that the path $\gamma_\infty(s)$, $s \in \mathbb{R}$, is in $\mathcal{Z}$. Since $\gamma_\infty(s)$ inherits the uniform bounds obeyed by $\gamma(s)$ and $\dot{\gamma}(s)$, we can find a sequence $s'_n \to +\infty$ such that

$$\gamma_\infty^{(n)}(s) := \gamma_\infty(s + s'_n) \to \bar{\gamma}_\infty(s),$$

locally uniformly in $s \in \mathbb{R}$. As all $\gamma_\infty^{(n_k)}(s)$ are calibrated by ϕ, we know that the limiting path $\bar{\gamma}_\infty(s)$, $s \in \mathbb{R}$, lies in $\mathcal{Z}$, by the definition of the set $\mathcal{Z}$.

To finish the proof of the proposition, consider the points $\gamma(s_n + s'_m)$. First, we fix m and choose $n = N_m$ sufficiently large, so that both

$$|\gamma(s_{N_m} + s'_m) - \gamma_\infty(s'_m)| < \frac{\varepsilon}{2},$$

and $s_{N_m} + s'_m < -m$. Next, we choose m sufficiently large, so that $|\gamma_\infty(s'_m) - \bar{\gamma}_\infty(0)| < \varepsilon/2$. This gives

$$|\gamma(s_{N_m} + s'_m) - \bar{\gamma}_\infty(0)| < \varepsilon.$$

Since $\bar{\gamma}_\infty(0)$ is in $\mathcal{Z}$ and $\tau_m = s_{N_m} + s'_m \to -\infty$, the proof is complete. □

Convergence on $\overline{\mathcal{Z}}$

After setting up the required objects, we turn to the proof of of Theorem 1.9.1. The strategy comes from Exercise 1.9.3: we need to show that any solution to

$$v_t + H(x, \nabla v) = 0, \quad t \in \mathbb{R}, \ x \in \mathbb{T}^n, \tag{1.9.35}$$

with $v(0, x) = \psi(x) \in \omega(u_0)$, is time-independent. The reader has certainly guessed what will happen: the set $\mathcal{Z}$ will play the same role as the zero set of the function $f(x)$ in Theorem 1.8.11, and the set where the function $R(x)$ attains its maximum in the proof of Theorem 1.8.13. This is confirmed by the following proposition, showing that such $v(t, x)$ is independent of $t \in \mathbb{R}$ on the closure $\overline{\mathcal{Z}}$ of the set $\mathcal{Z}$, though we do not know yet that this happens everywhere.

Proposition 1.9.11 *If $\psi \in \omega(u_0)$, then $v(t,x) = \mathcal{T}(t)\psi(x)$ does not depend on $t \in \mathbb{R}$ for all $x \in \overline{\mathcal{Z}}$.*

Proof. Consider an eternal extremal path $\gamma_\infty(s)$, $s \in \mathbb{R}$, in $\mathcal{Z}$. We are going to show that

$$\partial_t v(t, \gamma_\infty(t)) = 0, \quad \text{for all } t \in \mathbb{R}. \tag{1.9.36}$$

Let ϕ be a steady solution to (1.9.35). We have shown in Proposition 1.9.8 that $\gamma_\infty(s)$, $s \in \mathbb{R}$, is calibrated both by ϕ and by v. It follows then from Corollary 1.7.21 that both ϕ and v are $C^{1,1}$ on γ_∞, and we have

$$\nabla v(t, \gamma_\infty(t)) = \nabla_v L(\gamma_\infty(t), \dot\gamma_\infty(t)), \quad \nabla\phi(\gamma_\infty(t)) = \nabla_v L(\gamma_\infty(t), \dot\gamma_\infty(t)),$$

for all $t \in \mathbb{R}$, as in (1.7.48) in Exercise 1.7.13. This gives

$$\nabla v(t, \gamma_\infty(t)) = \nabla\phi(\gamma_\infty(t)), \quad \text{for all } t \in \mathbb{R}. \tag{1.9.37}$$

This relation holds in the classical sense, as both v and ϕ are $C^{1,1}$ on $\gamma_\infty(t)$. Since ϕ is a solution to the steady equation (1.9.14):

$$H(x, \nabla\phi) = 0, \tag{1.9.38}$$

we deduce that

$$H(\gamma_\infty(t), \nabla v(t, \gamma_\infty(t))) = 0.$$

As v is $C^{1,1}$ at $(t, \gamma_\infty(t))$, this entails (1.9.36):

$$\partial_t v(t, \gamma_\infty(t)) = 0, \quad \text{for all } t \in \mathbb{R}. \tag{1.9.39}$$

Consider $x \in \mathcal{Z}$ and an eternal extremal path $\gamma_\infty(\sigma)$, $\sigma \in \mathbb{R}$, in $\mathcal{Z}$ that passes through x, so that $\gamma_\infty(t) = x$, with some $t \in \mathbb{R}$. Given any $s \in \mathbb{R}$, Exercise 1.9.7 allows us to use (1.9.39) with the shifted path

$$\gamma_\infty^{(s)}(\sigma) = \gamma_\infty(\sigma + t - s).$$

Note that

$$x = \gamma_\infty(t) = \gamma_\infty^{(s)}(s),$$

and (1.9.39) applied to $\gamma_\infty^{(s)}(\sigma)$ at $\sigma = s$ gives

$$0 = \partial_t v(s, \gamma_\infty^{(s)}(s)) = \partial_t v(s, \gamma_\infty(t)) = \partial_t v(s, x). \tag{1.9.40}$$

Since s is arbitrary, we conclude that $v(t,x)$ does not depend on t for all $x \in \mathcal{Z}$. The continuity of $v(t,x)$ implies that the same is true for $\overline{\mathcal{Z}}$ as well. □

Convergence away from $\mathcal{Z}$

To finish the proof of Theorem 1.9.1 we now show that the claim of Proposition 1.9.11 holds also outside of $\overline{\mathcal{Z}}$.

Proposition 1.9.12 *If $\psi \in \omega(u_0)$, then $v(t,x) = \mathcal{T}(t)\psi(x)$ does not depend on $t \in \mathbb{R}$ for all $x \in \mathbb{T}^n$.*

Proof. Proposition 1.9.11 shows that we only need to consider $x \notin \overline{\mathcal{Z}}$. Using the by now familiar arguments based on the uniform bounds on minimizers, for any $x \in \mathbb{T}^n$ and $t > 0$ fixed, we can find a path $\gamma_t(\sigma)$, $\sigma \le t$, calibrated by v, so that for any $s < t$ we have

$$v(t,x) = v(s,\gamma_t(s)) + \int_s^t L(\gamma_t(\sigma),\dot\gamma_t(\sigma))d\sigma.$$

Proposition 1.9.9 shows that there exists a sequence $\tau_n \to +\infty$ such that

$$\mathrm{dist}(\gamma_t(-\tau_n),\mathcal{Z}) \to 0, \quad \text{as } n \to +\infty.$$

Hence, there exists $z \in \overline{\mathcal{Z}}$ and a subsequence $\tau_{n_k} \to +\infty$ such that

$$\gamma_t(-\tau_{n_k}) \to z \quad \text{as } k \to +\infty.$$

Since the function $v(t,x)$ is Lipschitz in x, uniformly in $t \in \mathbb{R}$, we know that

$$\Delta_k(t,x) = v(-\tau_{n_k},\gamma_t(-\tau_{n_k})) - v(-\tau_{n_k},z) \to 0 \quad \text{as } k \to +\infty. \tag{1.9.41}$$

Let us then write

$$v(t,x) = v(-\tau_{n_k},z) + \int_{-\tau_{n_k}}^t L(\gamma_t(\sigma),\dot\gamma_t(\sigma))d\sigma + \Delta_k(t,x). \tag{1.9.42}$$

As $z \in \overline{\mathcal{Z}}$, by Proposition 1.9.11 we know that $v(s,z)$ does not depend on s, so that for any $s \in \mathbb{R}$ we have

$$v(-\tau_{n_k},z) = v(-\tau_{n_k} - t + s, z) = \psi(z),$$

and (1.9.42) can be written as

$$\begin{aligned}
v(t,x) =& v(-\tau_{n_k} - t + s, z) + \int_{-\tau_{n_k}}^t L(\gamma_t(\sigma),\dot\gamma_t(\sigma))d\sigma + \Delta_k(t,x)\\
&= v(-\tau_{n_k} - t + s, z) + \int_{-\tau_{n_k}-t+s}^s L(\gamma_t(\sigma+t-s),\dot\gamma_t(\sigma+t-s))d\sigma + \Delta_k(t,x)\\
&= v(-\tau_{n_k} - t + s, z) + \int_{-\tau_{n_k}-t+s}^s L(\widetilde\gamma_t(\sigma),\dot{\widetilde\gamma}_t(\sigma))d\sigma + \Delta_k(t,x),
\end{aligned} \tag{1.9.43}$$

with the path

$$\widetilde\gamma_t(\sigma) = \gamma_t(\sigma + t - s), \quad \sigma \le s.$$

Note that $\widetilde\gamma_t(s) = \gamma_t(t) = x$, and

$$\widetilde\gamma_t(-\tau_{n_k} - t + s) = \gamma_t(-\tau_{n_k}) = z + \delta_k, \quad \delta_k \to 0 \ \text{ as } k \to +\infty. \tag{1.9.44}$$

Therefore, the Lax-Oleinik formula, together with the uniform Lipschitz property of the function $v(t,x)$, tells us that

$$\begin{aligned}
v(s,x) &\le v(-\tau_{n_k} - t + s, \widetilde\gamma_t(-\tau_{n_k} - t + s)) + \int_{-\tau_{n_k}-t+s}^s L(\widetilde\gamma_t(\sigma),\dot{\widetilde\gamma}_t(\sigma))d\sigma\\
&\le v(-\tau_{n_k} - t + s, z) + \int_{-\tau_{n_k}-t+s}^s L(\widetilde\gamma_t(\sigma),\dot{\widetilde\gamma}_t(\sigma))d\sigma + \Delta_k(t,x) + C\delta_k.
\end{aligned} \tag{1.9.45}$$

Comparing to (1.9.43) and passing to the limit $k \to +\infty$ we conclude that

$$v(s,x) \leq v(t,x) \quad \text{for all } t,s \in \mathbb{R}. \tag{1.9.46}$$

As t and s are arbitrary, it follows that $v(t,x)$ does not depend on t, finishing the proof of Proposition 1.9.12, and thus that of Theorem 1.9.1 as well. □

Our tour of the Hamilton-Jacobi equations ends here. One could say much more on the organization of the steady solutions, and the reader should consult [62]. They would be, however, outside the scope of this book. Let us just notice that the results of the present section provide a complete parallel with the large time behavior of the solutions to viscous Hamilton-Jacobi equations, which was the goal we wanted to achieve: the viscosity solutions of the inviscid problem still converge to waves, although their organization, that we have largely uncovered, is much more complicated.

Chapter 2

Introduction to mean field games

This chapter is based on the wonderful notes by Pierre Cardaliaguet on the mean-field games, often following them verbatim. I am very thankful to Pierre for allowing me to use them. All errors are mine.

2.1 What is a mean-field game?

The mean-field game system consists of a backward-in-time Hamilton-Jacobi equation for a value function $u(t, x)$ and a forward-in-time equation for a mean-field density $m(t, x)$. A typical example would look like the following system

$$
\begin{aligned}
&- \partial_t u - \nu\Delta u + H(x, Du) = f(x, m(x, t)),\\
&\partial_t m - \nu\Delta m - \operatorname{div}(D_p H(x, Du)m) = 0,\\
&m(0, x) = m_0(x)\ ,\ u(x, T) = G(x).
\end{aligned}
\tag{2.1.1}
$$

As we will see, such systems come up in an optimization problem approximating a large number of agents (players), where the behavior of each player is governed by the rest of the players via the mean-field. Accordingly, the evolution of the value function $u(t, x)$ for an individual player, is coupled to that of the density $m(t, x)$ of all players. The term mean-field refers to the fact that the strategy of each player is affected only by the average density (mean-field) of the other players, and not by a particular stochastic configuration of the system. The function $H(x, p)$ is the Hamiltonian, and the function $f(x, m)$ is a coupling between the value function of the optimal control problem and the density of the players. Of course, the coupling need not be local, and we will consider non-local couplings as well.

The most unusual feature of (2.4.1) is that it couples the forward Fokker-Planck equation that has an initial condition for $m(0, x)$ at the initial time $t = 0$ to the backward in time Hamilton-Jacobi equation for $u(t, x)$ that has a prescribed terminal value at $t = T$. Thus, this is not a Cauchy problem that normally arises in PDE problems, and has novel features compared to what we are used to see.

Mean field game theory is devoted to the analysis of differential games with infinitely many players. For such large population dynamic games, it is unrealistic for a player to collect detailed state information about all other players. Fortunately, this impossible task is useless: mean field game theory explains that one just needs to implement strategies based on the distribution of the other players. Such a strong simplification is well documented in the (static) game community since the seminal works of Aumann [154]. However, for differential games, this idea has been considered only very recently: the starting point is a series of papers by Lasry and Lions [242, 243, 244, 245], who introduced the terminology in around 2005. The term mean field comes for an analogy with the mean field models in mathematical physics, which analyse the behavior of many identical particles (see, for instance, Sznitman's notes [262]). Here, the particles are replaced by agents or players, whence the name of mean field games. Related ideas have been developed independently, and at about the same time, by Caines, Huang and Malhamé [228, 229, 230, 231], under the name of Nash certainty equivalence principle.

The Cardaliaguet notes we are following (copy-pasting almost 100% of the time) aim to give a basic presentation of the topic. They are largely based on Lions' series of lectures at the College de France [248] and on Lasry and Lions seminal papers on the subject [242, 243, 244, 245], but also on other notes taken from Lions lectures: Yves Achdou's survey for a CIME course [144] and Guéant's notes [222] (see also the survey by Gomes and Saude [212]).

There are several approaches to the analysis of differential games with an infinite number of agents. A first one is to look at the limit of Nash equilibria in differential games with a large number of players and try to pass to the limit as this number tends to infinity. A second approach consists in guessing the equations that Nash equilibria of differential games with infinitely many players should satisfy and to show that the resulting solutions of these equations allow to solve differential games with finitely many players.

Concerning the first approach, little was completely understood until very recently. Lions explains in [248] how to derive formally an equation for the limit to Nash equilibria: it is a nonlinear transport equation in the space of measures (the "master equation"). Existence, uniqueness of solution for this equation is an open problem in general, and, beside the linear-quadratic case, one did not know how to pass to the limit is the Nash system. Progress has been made very recently on both questions [169, 191, 202] and we explain some of the ideas in the second part of these notes. The starting point is that, as observed by Lions [248], the "characteristics" of the infinite dimensional transport equations solve—at least formally—a system coupling of a Hamilton-Jacobi equation with a Kolmogorov-Fokker-Planck equation: this is the MFG system, which is the main object of the first part of these notes.

A very nice point is that this system also provides a solution to the second approach: indeed, the feedback control, given by the solution of the mean field game system, provides $\varepsilon-$Nash equilibria in differential games with a large (but finite) number of players. This point was first noticed by Huang, Caines and Malham [229] and further developed in several papers (Carmona, Delarue [184], Kolokoltsov, Li, Yang [235], etc...).

To complete the discussion on the master equation, let us finally underline that, beside the MFG system, another possible and natural simplification of this equation is a space discretization, which yields to a more standard transport equation in finite space dimension: see the discussion by Lions in [248], by Gomes, Mohr, Souza [203, 204, 205] and Guant [220].

We now describe the mean field game system in a more precise way. The system has two unknowns u and m, which solve the equations

$$\begin{aligned} &-\partial_t u - \nu\Delta u + H(x,m,Du) = 0, \\ &\partial_t m - \nu\Delta m - \operatorname{div}\left(D_p H(x,m,Du)m\right) = 0, \\ &m(0,x) = m_0(x),\ u(x,T) = G(x,m(T)). \end{aligned} \tag{2.1.2}$$

In the above system, the diffusivity $\nu \geq 0$ is nonnegative so that it includes both the viscous and inviscid cases. The first equation in (2.1.2) is backward in time: a terminal condition is prescribed for $u(t,x)$, and the second one is forward in time: an initial condition is prescribed for the function $m(t,x)$.

There are two crucial structure conditions for this system: the first one is the convexity of the Hamiltonian $H = H(x,m,p)$ with respect to the last variable. This condition means that the Hamilton-Jacobi equation (the first equation in (2.1.2)) is associated to an optimal control problem. The solution to the first equation is interpreted as the value function associated with a typical small player. The second structure condition is that the initial condition m_0 is density of a probability measure. It follows from the structure of the Fokker-Planck equation (the second equation in (2.1.2)) that this property is preserved for all times:

$$\int_{\mathbb{R}^n} m(t,x)dx = \int_{\mathbb{R}^n} m_0(x)dx = 1. \tag{2.1.3}$$

The heuristic interpretation of this system is the following. An average agent controls the stochastic differential equation

$$dX_t = \alpha_t dt + \sqrt{2\nu}dB_t.$$

Here, B_t is a standard Brownian motion. He aims at minimizing the quantity

$$\mathbb{E}\left[\int_0^T \frac{1}{2}L(X_s, m(s), \alpha_s)ds + G(X_T, m(T))\right],$$

where L is the Legendre transform of H with respect to the p variable. Note that the evolution of the measure $m(s,\cdot)$ enters as a parameter in this cost function.

The value function of our average player is then given by the solution to (2.1.2-(i)). The corresponding optimal control is, at least heuristically, given in feedback form by

$$\alpha^*(x,t) = -D_p H(x,m,Du). \tag{2.1.4}$$

If all agents behave in this way and if their associated noises are independent, then by the law of large numbers their density moves with a velocity which is due, on the one hand, to the diffusion, and, one the other hand, to the drift term $-D_p H(x,m,Du)$. This leads to the Fokker-Planck equation (2.1.2-(ii)).

The aim of these notes is to collect, with detailed proofs, various existence and uniqueness results obtained by Lasry and Lions for the above system when the Hamiltonian H is "separated": $H(x,m,p) = H(x,p) - F(x,m)$. The coupling between the two equations is then via the function $F(x,m)$. There are two types of coupling which are appear in the mean field game literature. First, we may take F as nonlocal and regularizing. That is, we view $F = F(x, m(t))$ as a map on the space of probability measures. This is typically the case when two players who are not too close to each other can still influence themselves. Second, F may be of a local nature. That is, $F = F(x, m(x, t))$ depends on the value of the density at the point (t, x), meaning that the players only take into account their very nearest neighbors. Although the second coupling can be seen as a limit case of the first one, in practice, the proofs are more demanding in the local case. In particular, while we provide existence and uniqueness results for nonlocal couplings both when $\nu = 1$ (the viscous case) and $\nu = 0$ (the inviscid case), we will consider local couplings only for viscous equations. We willavoid the inviscid case with the local coupling. This case, described in [248], is only understood under specific structure conditions and requires several a priori estimates which, unfortunately, exceed the modest framework of these notes.

Warning: the literature comments below come from the original notes by Pierre Cardaliaguet, things have evolved since then but I am not in a position to review them. His comments are still very much relevant. Some comments on the literature are now in order. Since the pioneering works by Lasry and Lions and by Huang, Caines and Malham, the literature on the MFG has grown very fast: it is by now almost impossible to give a reasonable account of the activity on the topic. Many references on the subject can be found, for instance, in the survey by Gomes and Saud [212] and in the monograph by Bensoussan, Frehse and Yam [165]. We only provide here a few references, without the smallest pretense of completeness.

Let us start with the probabilistic aspects. As the value function of an optimal control problem is naturally described in terms of backward stochastic differential equations (BSDEs), it is very natural to understand the MFG system as a BSDE with a mean field term of McKean-Vlasov type: this is the approach generally followed the probabilistic part of the literature on mean field games: beside the papers by Huang, Caines and Malham already quoted, see also Buckdahn, Li, Peng [168], Buckdahn, Djehiche, Li, Peng [168], Andersson, Djehiche [153] (where a linear MFG system appears as optimality condition of a control of mean field type). Forward-backward stochastic differential equation (FBSDE) of the McKean-Vlasov type, are analyzed by Carmona, Delarue [184], Kolokoltsov, Li, Yang [235] (with nonlinear diffusions). MFG models with a major player are discussed by Huang [225], while Nourian, Caines, Malhame, Huang [255] deal with mean field LQG control in leader-follower stochastic multi-agent systems. Differential games in which several major players influence an overall population but compete with each others lead to differential games of mean field type, as considered by Bensoussan, Frehse [164]. Linear quadratic MFG system have also been very much investigated: beside Huang, Caines and Malham work, see Bensoussan, Sung, Yam, Yung [163], Carmona, Delarue [183] for probabilistic arguments, and Bardi [155] from a PDE view point.

In terms of PDE, the analysis of mean field games boils down—more or less—to solve the coupled system (2.1.2) with various assumptions on the coefficients. Beside Lasry and Lions' papers, other existence results and estimates for classical MFG system can be found in Guant [217, 221] (by use of Hopf-Cole transform for 2nd order of MFG systems with local coupling), Cardaliaguet, Lasry, Lions, Porretta [180] (2nd order MFG systems with local unbounded coupling), Bardi, Feleqi [156] (stationary MFG systems with general diffusions and boundary conditions), Gomes, Pirez, Sanchez-Morgado [206] (estimates for stationary MFG systems), Cardaliaguet [178] (1rst order MFG system, local coupling by methods of calculus of variation). Models with several populations are discussed by Feleqi [196], Bardi, Feleqi [156], Cirant [187]. Other models are considered in the literature: the so-called extended mean field games, i.e., MFG systems in which the HJ equation also depends on the velocity field of the players have been studied by Gomes, Patrizi, Voskanyan [207], Gomes, Voskanyan [208]; Santambrogio [259] discusses MFG models with density constraints; mean field kinetic model for systems of rational agents interacting in a game theoretical framework is discussed in [189] and [190].

Numerical aspects of the theory have been developed in particular by Achdou, Capuzzo Dolcetta [141], Achdou, Camilli, Capuzzo Dolcetta [142], [143], Achdou, Perez [146] Camilli, Silva [172], Lachapelle, Salomon, Turinici [237].

As shown by numerical studies, solutions of time dependent MFG systems, such as (2.1.2) quickly stabilize to stationary MFG systems: the analysis of the phenomenon (i.e., the long time behavior of solutions of the mean field game system) has been considered for discrete systems by Gomes, Mohr, Souza [203] and for continuous ones in Lions' lectures, and subsequently developed by Cardaliaguet, Lasry, Lions, Porretta [180, 181] for second order MFG game system with local and nonlocal couplings, in Cardaliaguet [177], from 1rst order MFG systems with nonlocal coupling.

It is impossible to cover all the applications of MFG to economics, social science, biological science, and engineering—and this part is even less complete than the previous ones. Let us just mention that the early work on large population differential games was motivated by wireless power control problems: see Huang, Caines, Malham [226, 227]. Application to economic models can be found in Guant [216], Guant, Lions, Lasry, [218, 246], Lachapelle [236], Lachapelle, Wolfram [238], Lucas, Moll [249]. A price formation model, inspired by the MFG, has been introduced in Lasry, Lions [242] and analyzed by Markowich, Matevosyan, Pietschmann, Wolfram [250], Caffarelli, Markowich, Wolfram [171].

2.2 Nonatomic games

Before starting the analysis of differential games with a large number of players, it is helpful to look at this question for classical games. The general framework is as follows: let N be a (large) number of players. We will usually assume that the players are identical, in the sense that the set Q of available strategies is the same for all players. We denote by $F_i^N = F_i^N(x_1, \ldots, x_N)$ the payoff (or the cost) of player $i \in \{1, \ldots, N\}$ given the "all-players" state $(x_1, \ldots, x_N)$. The symmetry assumption means that

$$F^N_{\sigma(i)}(x_{\sigma(1)}, \ldots, x_{\sigma(N)}) = F_i(x_1, \ldots, x_N)$$

for all permutations σ on $\{1, \ldots, N\}$. Our goal is to analyze the behavior of the Nash equilibria for this game as $N \to +\infty$.

For this, we first recall the notion of Nash equilibria. In order to proceed with the analysis of large population games, we describe next the limit of maps of many variables. Then we explain the limit, as the number of players tends to infinity, of Nash equilibria in pure, and then in mixed, strategies. This is how the mean-field game equation comes about. We finally discuss the uniqueness of the solution of the limit equation and present some examples.

2.2.1 Nash equilibria in classical differential games

Here, we introduce the notion of Nash equilibria in one-shot games. We do not assume that the players are identical. Let $Q_1, \dots, Q_N$ be compact metric spaces – the elements of Q_i are the possible strategies of player i, and $J_1, \dots, J_N$ be continuous real valued functions on $\prod_{i=1}^N Q_i$.

Definition 2.2.1 *A Nash equilibrium in pure strategies is a $N-$tuple $(\bar{s}_1, \dots, \bar{s}_N) \in \prod_{i=1}^N Q_i$ such that, for any $i = 1, \dots, N$,*

$$J_i(\bar{s}_1, \dots, \bar{s}_N) \le J_i\left(s_i, (\bar{s}_j)_{j\neq i}\right) \qquad \forall s_i \in Q_i\ .$$

In other words, a Nash equilibrium is a set of strategies $\bar{s}_1, \dots, \bar{s}_N$ such that it is "expensive" for a player i to deviate from $\bar{s}_i$ provided that all other players uses strategies $\bar{s}_k$, $k \neq i$. Let us consider a couple of examples.

Example 2.2.2 Consider two players who can set prices p_1 and p_2, with $0 \le p_1, p_2 \le 1$, and sell fractions $x_1(p_1, p_2)$ and $x_2(p_1, p_2)$ of units respectively, with

$$x_2(p_1, p_2) = \frac{2}{3}(p_1 - p_2), \text{ if } p_1 \ge p_2, \text{ and } x_2(p_1, p_2) = 0 \text{ if } p_1 < p_2,$$

and $x_1(p_1, p_1) = 1 - x_2(p_1, p_2)$. The profit of the two players is

$$u_1(p_1, p_2) = p_1 x_1(p_1, p_2), \quad u_2(p_1, p_2) = p_2 x_2(p_1, p_2).$$

Then, given the strategy p_2, the optimization problem for the first player is to maximize the function

$$u_1(p_1) = p_1 x_1(p_1, p_2),$$

with

$$x_1 = 1 - \frac{2}{3}(p_1 - p_2).$$

A simple computation shows that the optimal value of p_1 (again, given p_2) is

$$\widetilde{p}_1(p_2) = \min(1, \frac{3}{4} + \frac{p_2}{2}).$$

The second player optimizes $u_2 = p_2 x_2$, subject to the constraint $x_2 = (2/3)(p_1 - p_2)$, so the optimal price for him (given p_1) is $\widetilde{p}_2(p_1) = p_1/2$. Then the unique Nash equilibrium is $\bar{p}_1 = 1$ and $\bar{p}_2 = 1/2$.

Example 2.2.3 Let us look at a similar example but with slightly different constraints. Again, the profits of the two players are $u_1(p_1, p_2) = p_1 x_1(p_1, p_2)$ and $u_2(p_1, p_2) = p_2 x_2(p_1, p_2)$, with $x_1(p_1, p_1) = 1 - x_2(p_1, p_2)$. However, the fraction $x_2(p_1, p_2)$ is now determined by

$$p_1 = p_2 + l(x_2), \quad l(x) = \frac{x - 1/2}{\varepsilon}, \text{ if } x \ge 1/2, \text{ and } l(x) = 0 \text{ if } 0 \le x \le 1/2.$$

In other words,

$$x_2(p_1, p_2) = \frac{1}{2} + \varepsilon(p_1 - p_2), \quad \text{if } p_1 > p_2, \tag{2.2.1}$$

and $x_2 = 0$ if $p_1 \le p_2$, be while

$$x_1(p_1,p_2) = \frac{1}{2} - \varepsilon(p_1 - p_2), \quad \text{if } p_1 > p_2, \tag{2.2.2}$$

and $x_1 = 1$ if $p_1 \le p_2$. Then, given p_2, the first player optimizes the function

$$u_1(p_1,p_2) = p_1 x_1(p_1,p_2) = \begin{cases} p_1, & \text{if } p_1 \le p_2, \\ p_1\big(\frac{1}{2} - \varepsilon(p_1 - p_2)\big), & \text{if } p_1 > p_2. \end{cases} \tag{2.2.3}$$

Note that

$$\max_{p_1} u_1(p_1,p_2) = \max\Big(p_2, \frac{1}{2} - \varepsilon + \varepsilon p_2\Big). \tag{2.2.4}$$

Thus, the optimal strategy of the first player is

$$\widetilde{p}_1(p_2) = p_2, \quad \text{if } p_2 \ge \frac{1/2 - \varepsilon}{1 - \varepsilon}, \tag{2.2.5}$$

and

$$\widetilde{p}_1(p_2) = 1, \quad \text{if } p_2 < \frac{1/2 - \varepsilon}{1 - \varepsilon}. \tag{2.2.6}$$

On the other hand, for the second player we have

$$u_2(p_1,p_2) = p_2 x_2(p_1,p_2) = \begin{cases} 0, & \text{if } p_1 < p_2, \\ p_2\big(\frac{1}{2} + \varepsilon(p_1 - p_2)\big), & \text{if } p_1 > p_2. \end{cases} \tag{2.2.7}$$

We see that

$$\max_{p_2} u_2(p_1,p_2) = \frac{p_1}{2}, \tag{2.2.8}$$

and the optimal strategy of the second player is

$$\widetilde{p}_2(p_1) = p_1. \tag{2.2.9}$$

Then one can directly check that a pure Nash equilibrium does not exist when $\varepsilon > 0$ is sufficiently small, according to some MIT slides.

Example 2.2.4 Consider the symmetric setting where $Q_1 = Q_2 = \mathbb{T}^1$, and there is a function $F(x_1,x_2)$ so that $J_1(x_1,x_2) = F(x_1,x_2)$, $J_2(x_1,x_2) = F(x_2,x_1)$. Then a point (y_1,y_2) is a pure Nash equilibrium if

$$\frac{\partial J_1(y_1,y_2)}{\partial y_1} = \frac{\partial J_2(y_1,y_2)}{\partial y_2} = 0. \tag{2.2.10}$$

This translates into

$$\frac{\partial F}{\partial y_1}(y_1,y_2) = 0 \text{ and } \frac{\partial F}{\partial y_1}(y_2,y_1) = 0. \tag{2.2.11}$$

It is easy to construct a function F such that (2.2.11) has no solutions – the only requirement on the function $\partial F/\partial x_1$ is that F is periodic and

$$\int_0^1 \frac{\partial F(x_1,x_2)}{\partial x_1} dx_1 = 0 \text{ for all } 0 \le x_2 \le 1,$$

so the requirement that its zero set contains no two points symmetric with respect to the line $x_1 = x_2$ can be satisfied, and then (2.2.11) has no solutions.

Thus, Nash equilibria in pure strategies do not necessarily exist and, to ensure their existence, we have to introduce the notion of mixed strategies. This means that each player uses a family of strategies with a certain probability distribution. Let us denote by $\mathcal{P}(Q_i)$ the space of all Borel probability measures on Q_i. A mixed strategy of player i will be an element of $\mathcal{P}(Q_i)$. The set $\mathcal{P}(Q)$ is endowed with the weak-* topology: a sequence m_N in $\mathcal{P}(Q)$ converges to $m \in \mathcal{P}(Q)$ if

$$\lim_{N\to\infty} \int_Q \varphi(x) dm_N(x) = \int_Q \varphi(x) dm(x) \qquad \forall \varphi \in \mathcal{C}(Q) .$$

Recall that $\mathcal{P}(Q)$ is a compact metric space for this topology, which can be metrized by the Kantorowich-Rubinstein distance:

$$d_1(\mu,\nu) = \sup\{ \int_Q f d(\mu-\nu) : \ \|f\|_{Lip(Q)} \leq 1 \text{ and } \sup_{x\in Q} |f(x)| \leq 1.\} .$$

Alternatively, this distance can be stated in terms of optimal transportation:

$$d_1(\mu,\nu) = \inf_M \int_{Q\times Q} d(x,y) dM(x,y),$$

with the infimum taken over all probability measures $dM(x,y)$ on $Q \times Q$ such that the marginals of $M(x,y)$ in x and y are μ and ν, respectively.

Definition 2.2.5 *A Nash equilibrium in mixed strategies is an $N-$tuple $(\bar\pi_1,\dots,\bar\pi_N) \in \prod_{i=1}^N \mathcal{P}(Q_i)$ such that, for any $i = 1,\dots,N$,*

$$J_i(\bar\pi_1,\dots,\bar\pi_N) \leq J_i\left((\bar\pi_j)_{j\neq i}, \pi_i\right) \qquad \forall \pi_i \in \mathcal{P}(Q_i) . \tag{2.2.12}$$

where, with some abuse of notation, we set

$$J_i(\pi_1,\dots,\pi_N) = \int_{Q_1\times\cdots\times Q_N} J_i(s_1,\dots,s_N) d\pi_1(s_1)\dots d\pi_N(s_N) .$$

Theorem 2.2.6 (Nash (1950), Glicksberg (1952)) *Under the above assumptions, there exists at least one equilibrium point in mixed strategies.*

Proof. Consider the best response map $\mathcal{R}_i : X := \prod_{j=1}^N \mathcal{P}(Q_i) \to 2^{\mathcal{P}(S_i)}$ of player i:

$$\mathcal{R}_i(\pi_1,\dots,\pi_N) = \left\{ \pi \in \mathcal{P}(Q_i) \ , \ J_i((\pi_j)_{j\neq i}, \pi) = \min_{\pi' \in \mathcal{P}(S_i)} J_i((\pi_j)_{j\neq i}, \pi') \right\}, \tag{2.2.13}$$

and define $\phi(\pi_1,\dots,\pi_N) = \prod_{i=1}^N \mathcal{R}_i(\pi_1,\dots,\pi_N) : X \to 2^X$. Then, any fixed point x of ϕ such that $x \in \phi(x)$ is a Nash equilibrium of mixed strategies.

Existence of such fixed point is established using Fan's fixed point Theorem [195]. It says the following. Let X be a non-empty, compact and convex subset of a locally convex topological vector space. We say that a set-valued function $\phi : X \to 2^X$ is upper-semicontinuous if for every open set $W \subset X$, the set $\{x \in X : \ \phi(x) \subseteq W\}$ is open in X. Equivalently, for every closed set $H \subset X$, the set $\{x \in X : \ \phi(x) \cap H \neq \emptyset\}$ is closed in X. Assume also that $\phi(x)$ is non-empty, compact and convex for all $x \in X$. Then ϕ has a fixed point: $\exists \bar{x} \in X$ with $\bar{x} \in \phi(\bar{x})$.

Note that in our setting ϕ is upper semicontinuous. Indeed, let $W \subset X$ be an open set and take $x = (\pi_1, \ldots, \pi_n) \in X$ such that $\phi(x) \in W$. Then for $x' = (\pi_1', \ldots, \pi_n')$ sufficiently close to x, the minimizers in (2.2.13) for π_j', $j \neq i$ fixed, will be close to the minimizers corresponding to π_j, $j \neq i$ fixed, so that $\phi(x') \in W$. It is also easy to see that the values $\phi(x)$ are compact, convex and non-empty. Therefore, ϕ has a fixed point, which is a Nash equilibrium in mixed strategies by the definition of ϕ. □

Let us now consider the special case where the game is symmetric. Namely, we assume that, for all $i \in \{1, \ldots, N\}$, $Q_i = Q$ and $J_i(s_1, \ldots, s_N) = J_{\theta(s_i)}(s_{\theta(1)}, \ldots, s_{\theta(N)})$ for all i and all permutations θ on $\{1, \ldots, N\}$.

Theorem 2.2.7 (Symmetric games) *If the game is symmetric, then there is an equilibrium of the form $(\bar{\pi}, \ldots, \bar{\pi})$, where $\bar{\pi} \in \mathcal{P}(Q)$ is a mixed strategy.*

Proof. Let $X = \mathcal{P}(Q)$ and $\mathcal{R} : X \to 2^X$ be the set-valued map defined by

$$\mathcal{R}(\pi) = \left\{ \sigma \in X \ , \ J_1(\sigma, \pi, \ldots, \pi) = \min_{\sigma' \in X} J_1(\sigma', \pi, \ldots, \pi) \right\} .$$

Then $\mathcal{R}$ is upper semicontinuous with nonempty convex compact values. By Fan's fixed point Theorem, it has a fixed point $\bar{\pi}$ and, from the symmetry of the game, the $N-$tuple $(\bar{\pi}, \ldots, \bar{\pi})$ is a Nash equilibrium. □

2.2.2 Symmetric functions of many variables

Let Q be a compact metric space and $u_N : Q^N \to \mathbb{R}$ be a symmetric function:

$$u_N(x_1, \ldots, x_N) = u_N(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \quad \text{for any permutation } \sigma \text{ on } \{1, \ldots, n\}.$$

Our aim is to define a limit for u_N – note that the number of unknowns depends on N also, so something slightly non-standard needs to be done. The idea is to associate to the points $x_1, \ldots, x_N$ the measure

$$m_X^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}.$$

Next, we interpret $u_N(X)$ as the value of a certain functional on m_X^N. To this end, we make the following two assumptions on u_N. First, a uniform bound: there exists $C > 0$ so that

$$\|u_N\|_{L^\infty(Q)} \leq C \tag{2.2.14}$$

Second, uniform continuity: there is a modulus of continuity ω independent of n such that

$$|u_N(X) - u_N(Y)| \leq \omega(d_1(m_X^N, m_Y^N)) \qquad \forall X, Y \in Q^N, \ \forall N \in \mathbb{N}. \tag{2.2.15}$$

Under these assumptions, define the maps $U^N : \mathcal{P}(Q) \to \mathbb{R}$ by

$$U^N(m) = \inf_{X \in Q^N} \left\{ u_N(X) + \omega(d_1(m_X^N, m)) \right\} \qquad \forall m \in \mathcal{P}(Q) .$$

Then, by assumption (2.2.15), we have

$$U^N(m_X^N) = \inf_{Y \in Q^N} \left\{ u_N(Y) + \omega(d_1(m_Y^N, m_X^N)) \right\} = u_N(X), \text{ for any } X \in Q^N.$$

With this interpretation, instead of talking about the convergence of the functions u_N that are defined on different spaces Q_N that depend on N, we can talk about convergence of the functionals U^N that are all defined on $\mathcal{P}(Q)$.

Theorem 2.2.8 *If u_N are symmetric and satisfy (2.2.14) and (2.2.15), then there is a subsequence u_{N_k} of u_N and a continuous map $U : \mathcal{P}(Q) \to \mathbb{R}$ such that*

$$\lim_{k\to+\infty} \sup_{X \in Q^{N_k}} |u_{N_k}(X) - U(m_X^{N_k})| = 0 .$$

Proof. [Proof of Theorem 2.2.8.] Without loss of generality we can assume that the modulus ω is concave. Let us show that the U^N have ω for modulus of continuity on $\mathcal{P}(Q)$: if $m_1, m_2 \in \mathcal{P}(Q)$ and if $X \in Q^N$ is $\varepsilon-$optimal in the definition of $U^N(m_2)$:

$$u_N(x) + \omega(d_1(m_X^N, m_2)) \le U^N(m_2) + \varepsilon,$$

then we have

$$\begin{array}{rcl} U^N(m_1) & \le & u_N(X) + \omega(d_1(m_X^N, m_1)) \le u_N(X) + \omega(d_1(m_X^N, m_2) + d_1(m_1, m_2)) \\ & \le & U^N(m_2) + \varepsilon + \omega(d_1(m_X^N, m_2) + d_1(m_1, m_2)) - \omega(d_1(m_X^N, m_2)) \\ & \le & U^N(m_2) + \omega(d_1(m_1, m_2)) + \varepsilon, \end{array}$$

because ω is concave. Hence the family U^N are equicontinuous on the compact set $\mathcal{P}(Q)$ and uniformly bounded. We complete the proof thanks to the Ascoli Theorem. ◻

Remark 2.2.9 Some uniform continuity condition is needed: for instance if Q is a compact subset of $\mathbb{R}^d$ and $u_N(X) = \max_i |x_i|$, then u_N "converges" to $U(m) = \sup_{x \in spt(m)} |x|$ which is not continuous. Of course the convergence is not uniform.

Remark 2.2.10 If Q is a compact subset of some finite dimensional space $\mathbb{R}^d$, a typical condition which ensures (2.2.15) is the existence of a constant $C > 0$, independent of N, such that

$$\sup_{i=1,\dots,N} \|D_{x_i} u_N\|_\infty \le \frac{C}{N} \qquad \forall N.$$

2.2.3 Limits of Nash equilibria in pure strategies

Let us assume is that the payoffs $F_1^N, \dots, F_N^N$ of the players are symmetric. In particular, under suitable bounds and uniform continuity, we know from Theorem 2.2.8 that F_i^N have a limit, which has the form $F(x, m)$. Here, the dependence on x is to keep track of the dependence on i of the function F_i^N. So the payoffs of the players are very close to the form

$$F(x_1, \frac{1}{N-1} \sum_{j\ge2} \delta_{x_j}), \dots, F(x_N, \frac{1}{N-1} \sum_{j\le N-1} \delta_{x_j}).$$

In order to keep the presentation as simple as possible, we suppose that the payoffs already have this form. That is, we suppose that there is a continuous map $F : Q \times \mathcal{P}(Q) \to \mathbb{R}$ such that, for any $i \in \{1, \dots, N\}$

$$F_i^N(x_1, \dots, x_N) = F\Big(x_i, \frac{1}{N-1}\sum_{j\neq i} \delta_{x_j}\Big) \qquad \forall (x_1, \dots, x_N) \in Q^N \ .$$

Let us recall that a pure Nash equilibrium for the game $(F_1^N, \dots, F_N^N)$ is $(\bar{x}_1^N, \dots, \bar{x}_N^N) \in Q^N$ such that

$$F_i^N(\bar{x}_1^N, \dots, \bar{x}_{i-1}^N, y_i, \bar{x}_{i+1}^N, \dots, \bar{x}_N^N) \geq F_i^N(\bar{x}_1^N, \dots, \bar{x}_N^N) \qquad \forall y_i \in Q \ .$$

We set

$$\bar{X}^N = (\bar{x}_1^N, \dots, \bar{x}_N^N) \qquad \text{and} \qquad m_{\bar{X}^N}^N = \frac{1}{N}\sum_{i=1}^N \delta_{\bar{x}_i^N} \ .$$

Theorem 2.2.11 *Assume that $\bar{X}^N = (\bar{x}_1^N, \dots, \bar{x}_N^N)$ is a Nash equilibrium in pure strategies for the game $F_1^N, \dots, F_N^N$. Then up to extraction of a subsequence, the sequence of measures $m_{\bar{X}^N}^N$ converges to a measure $\bar{m} \in \mathcal{P}(Q)$ such that*

$$\int_Q F(y, \bar{m}) d\bar{m}(y) = \inf_{m\in\mathcal{P}(Q)} \int_Q F(y, \bar{m}) dm(y) \ . \tag{2.2.16}$$

Remark 2.2.12 The "mean field equation" (2.2.16) is equivalent to saying that the support of $\bar{m}$ is contained in the set of minima of $F(y, \bar{m})$. Indeed, if $\text{Spt}(\bar{m}) \subset \text{argmin}_{y\in Q} F(y, \bar{m})$, then clearly $\bar{m}$ satisfies (2.2.16). Conversely, if (2.2.16) holds, then choosing $m = \delta_x$ shows that

$$\int_Q F(y, \bar{m}) d\bar{m}(y) \leq F(x, \bar{m}) \text{ for any } x \in Q.$$

Therefore, we have

$$\int_Q F(y, \bar{m}) d\bar{m}(y) \leq \min_{x\in Q} F(x, \bar{m}),$$

which implies that $\bar{m}$ is supported in $\text{argmin}_{y\in Q} F(y, \bar{m})$.

Proof. Without loss of generality we can assume that the sequence $m_{\bar{X}^N}^N$ converges to some $\bar{m}$. Let us check that $\bar{m}$ satisfies (2.2.16). Note that, by the definition of a pure Nash equilibrium, the measure $\delta_{\bar{x}_i^N}$ is a minimizer of the problem

$$\inf_{m\in\mathcal{P}(Q)} \int_Q F(y, \frac{1}{N-1}\sum_{j\neq i} \delta_{\bar{x}_j^N}) dm(y).$$

Since

$$d_1\Big(\frac{1}{N-1}\sum_{j\neq i} \delta_{\bar{x}_j^N}, m_{\bar{X}^N}^N\Big) \leq \frac{2}{N} \ ,$$

and since F is uniformly continuous, the measure $\delta_{\bar{x}_i^N}$ is also $\varepsilon-$optimal for the problem

$$\inf_{m\in\mathcal{P}(Q)}\int_Q F(y, m^N_{\bar{X}^N})dm(y),$$

as soon as N is sufficiently large, and this is true for all $i = 1, \ldots, N$. By linearity, so is $m^N_{\bar{X}^N}$:

$$\int_Q F(y, m^N_{\bar{X}^N})dm^N_{\bar{X}^N}(y) \leq \inf_{m\in\mathcal{P}(Q)}\int_Q F(y, m^N_{\bar{X}^N})dm(y) + \varepsilon.$$

Letting $N \to +\infty$ gives the result. ☐

2.2.4 Limit of the Nash equilibria in mixed strategies

Theorem 2.2.11 is not completely satisfying because it requires the existence of a pure Nash equilibrium in the $N-$player game, which does not always hold. However a Nash equilibrium in mixed strategies always exists, and we now discuss the corresponding result.

We now assume that the players play the same game $F_1^N, \ldots, F_N^N$ as before, but they are allowed to play in mixed strategies – they minimize over elements of $\mathcal{P}(Q)$ instead of minimizing over elements of Q. If the players play the mixed strategies $\pi_1, \ldots, \pi_N \in \mathcal{P}(Q)$, then the outcome of player i (still denoted, by abuse of notation, F_N^i) is

$$F_i^N(\pi_1, \ldots, \pi_N) = \int_{Q^N} F\Big(x_i, \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j}\Big)d\pi_1(x_1)\ldots d\pi_N(x_N)\ . \tag{2.2.17}$$

Recall that that symmetric Nash equilibria do exist for mixed strategies, unlike for pure strategies.

Theorem 2.2.13 *Assume that F is Lipschitz continuous. Let $(\bar{\pi}^N, \ldots, \bar{\pi}^N)$ be a symmetric Nash equilibrium in mixed strategies for the game $F_1^N, \ldots, F_N^N$. Then, up to a subsequence, $\bar{\pi}^N$ converges to a measure $\bar{m}$ satisfying (2.2.16).*

Remark 2.2.14 In particular the above Theorem proves the existence of a solution to the "mean field equation" (2.2.16).

Proof. Let $\bar{m}$ be a limit, up to extracting a subsequence, of $\bar{\pi}^N$. Fix $y \in Q$ and consider the map

$$\widetilde{F}(y, x) = F(y, \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j}) :\ Q^{N-1} \to \mathbb{R}.$$

Note that $\widetilde{F}$ is $\mathrm{Lip}(F)/(N-1)-$Lipschitz continuous in each coordinate $x_j \in Q$, hence twe have, by the definition of the distance d_1:

$$\Big|\int_{Q^{N-1}}\widetilde{F}(y, x)\prod_{j\neq i}d\bar{\pi}^N(x_j) - \int_{Q^{N-1}}\widetilde{F}(y, x)\prod_{j\neq i}d\bar{m}(x_j)\Big| \leq \mathrm{Lip}(F)d_1(\bar{\pi}^N, \bar{m}) \qquad \forall y \in Q\ . \tag{2.2.18}$$

Since $(\bar{\pi}_1, \ldots, \bar{\pi}_N)$ is a Nash equilibrium, inequality (2.2.18) implies that, for any $\varepsilon > 0$ and if we choose N large enough, we have

$$\int_{Q^N} F(y, \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j})\prod_{j\neq i} d\bar{m}(x_j)d\bar{m}(y) \leq \int_{Q^N} F(y, \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j})\prod_{j\neq i} d\bar{m}(x_j)dm(y) + \varepsilon \ , \tag{2.2.19}$$

for any $m \in \mathcal{P}(Q)$. Note also that we have

$$\lim_{N\to+\infty}\int_{Q^{N-1}} F(y, \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j})\prod_{j\neq i} d\bar{m}(x_j) = F(y, \bar{m}) \ , \tag{2.2.20}$$

where the convergence is uniform with respect to $y \in Q$ thanks to the (Lipschitz) continuity of F. Letting $N \to +\infty$ in both sides of (2.2.19) gives, in view of (2.2.20),

$$\int_Q F(y, \bar{m})d\bar{m}(y) \leq \int_Q F(y, \bar{m})dm(y) + \varepsilon \qquad \forall m \in \mathcal{P}(Q) \ ,$$

which finishes the proof, since ε is arbitrary. □

We can also investigate the converse statement: suppose that a measure $\bar{m}$ satisfying the equilibrium condition (2.2.16) is given. To what extent can it be used in an $N-$player game?

Theorem 2.2.15 *Let F be as in Theorem 2.2.13. For any $\varepsilon > 0$, there exists $N_0 \in \mathbb{N}$ such that, if $N \geq N_0$, the symmetric mixed strategy $(\bar{m}, \cdot, \bar{m})$ is $\varepsilon-$optimal in the $N-$player game with costs (F_i^N) defined by (2.2.17). Namely, we have*

$$F_i^N(\bar{m}, \ldots, \bar{m}) \leq F_i^N(x_i, (\bar{m})_{j\neq i}) + \varepsilon \qquad \forall x_i \in Q.$$

Proof. Indeed, as explained in the proof of Theorem 2.2.13, see (2.2.20). we have

$$\lim_{N\to+\infty} F_i^N(x_i, (\bar{m})_{j\neq i}) = F(x_i, \bar{m})$$

and this limit holds uniformly with respect to $x_i \in Q$. So we can find N_0 such that

$$\sup_{x_i\in Q}\left|F_i^N(x_i, (\bar{m})_{j\neq i}) - F(x_i, \bar{m})\right| \leq \varepsilon/2 \qquad \forall N \geq N_0. \tag{2.2.21}$$

Then, for any $x_i \in Q$, we have

$$F_i^N(x_i, (\bar{m})_{j\neq i}) \geq F(x_i, \bar{m}) - \varepsilon/2 \geq \int_Q F(y_i, \bar{m})d\bar{m}(y_i) - \varepsilon/2 \tag{2.2.22}$$

where the last inequality comes from the mean-field equaiton (2.2.16) for $\bar{m}$ by using $m = \delta_{x_i}$. Using again (2.2.21) and (2.2.22), we finally get

$$F_i^N(x_i, (\bar{m})_{j\neq i}) \geq \int_Q F(y_i, \bar{m})d\bar{m}(y_i) - \varepsilon/2 \geq F_i^N(\bar{m}, \ldots, \bar{m}) - \varepsilon.$$

□

2.2.5 A uniqueness result

One obtains the full convergence of the measure $m^N_{\hat{X}^N}$ (or $\bar{\pi}^N$), rather than along a subsequence, if there is a unique measure $\bar{m}$ satisfying the mean-field equation (2.2.16). This is the case under the following (very strong) assumption:

Proposition 2.2.16 *Assume that F satisfies*

$$\int_Q (F(y,m_1) - F(y,m_2))d(m_1 - m_2)(y) > 0 \qquad \forall m_1 \neq m_2 \ . \tag{2.2.23}$$

Then there is at most one measure satisfying (2.2.16).

Remark 2.2.17 Requiring at the same time the continuity of F and the above monotonicity condition seems rather restrictive for applications.

Condition (2.2.23) is more easily fulfilled for mappings defined on strict subsets of $\mathcal{P}(Q)$. For instance, if Q is a compact subset of $\mathbb{R}^d$ of positive measure and $\mathcal{P}_{ac}(Q)$ is the set of absolutely continuous measures on Q, with respect to the Lebesgue measure, then

$$F(y,m) = \begin{cases} G(m(y)) & \text{if } m \in \mathcal{P}_{ac}(Q) \\ +\infty & \text{otherwise} \end{cases}$$

satisfies (2.2.23) as soon as $G : \mathbb{R} \to \mathbb{R}$ is continuous and increasing. Here, we denote by $m(y)$ the density of m at y.

If we assume that Q is the closure of a smooth open bounded subset Ω of $\mathbb{R}^d$, another example is given by

$$F(y,m) = \begin{cases} u_m(y) & \text{if } m \in \mathcal{P}_{ac}(Q) \cap L^2(Q) \\ +\infty & \text{otherwise} \end{cases}$$

where u_m is the solution in $H^1(Q)$ of

$$\begin{cases} -\Delta u_m = m & \text{in } \Omega \\ u_m = 0 & \text{on } \partial\Omega \end{cases}$$

Note that in this case the map $y \to F(y,m)$ is continuous.

Proof. [Proof of Proposition 2.2.16] Let $\bar{m}_1, \bar{m}_2$ satisfying (2.2.16). Then

$$\int_Q F(y,\bar{m}_1)d\bar{m}_1(y) \leq \int_Q F(y,\bar{m}_1)d\bar{m}_2(y)$$

and

$$\int_Q F(y,\bar{m}_2)d\bar{m}_2(y) \leq \int_Q F(y,\bar{m}_2)d\bar{m}_1(y) \ .$$

Therefore

$$\int_Q (F(y,\bar{m}_1) - F(y,\bar{m}_2))d(\bar{m}_1 - \bar{m}_2)(y) \leq 0 \ ,$$

which implies that $\bar{m}_1 = \bar{m}_2$ thanks to assumption (2.2.23). □

2.2.6 An example: potential games

We now consider a class of nonatomic games for which the mean-field game equilibria can be found by minimizing a functional. To fix the idea, we assume that $Q \subset \mathbb{R}^d$, and that $F(x, m)$ has the form

$$F(y, m) = \begin{cases} F(m(y)) & \text{if } m \in \mathcal{P}_{ac}(Q) \\ +\infty, & \text{otherwise} \end{cases}$$

where $\mathcal{P}_{ac}(Q)$ is the set of absolutely continuous measures on Q, with respect to the Lebesgue measure, and $m(y)$ is the density of m at $y \in Q$. If $F(x, m)$ can be represented as the derivative of some mapping $\Phi(x, m)$ with respect to the $m-$variable, and if the problem

$$\inf_{m \in \mathcal{P}(Q)} \int_Q \Phi(x, m) dx$$

has a minimum $\bar{m}$, then the first variation tells us that

$$\int_Q \Phi'(x, \bar{m})(dm - d\bar{m}) \geq 0 \qquad \forall m \in \mathcal{P}(Q),$$

so

$$\int_Q F(x, \bar{m}) dm \geq \int_Q F(x, \bar{m}) d\bar{m} \qquad \forall m \in \mathcal{P}(Q) ,$$

which shows that $\bar{m}$ is a solution of the mean-field game equation.

For instance let us assume that

$$F(x, m) = \begin{cases} V(x) + G(m(x)) & \text{if } m \in \mathcal{P}_{ac}(Q) \\ +\infty & \text{otherwise} \end{cases}$$

where $V : Q \to \mathbb{R}$ is continuous and $G : (0, +\infty) \to \mathbb{R}$ is continuous, strictly increasing, with $G(0) = 0$ and $G(s) \geq cs$ for some $c > 0$. Then let

$$\Phi(x, m) = V(x)m(x) + H(m(x)) \qquad \text{if } m \text{ is a.c.}$$

where H is a primitive of G with $H(0) = 0$. Note that H is strictly convex with

$$H(s) \geq \frac{c}{2} s^2.$$

Hence the problem

$$\inf_{m \in \mathcal{P}_{ac}(Q)} \int_Q V(x)m(x) + H(m(x)) dx$$

has a unique solution $\bar{m} \in L^2(Q)$. Then we have, for any $m \in \mathcal{P}_{ac}(Q)$,

$$\int_Q (V(x) + G(\bar{m}(x)))m(x) dx \geq \int_Q (V(x) + G(\bar{m}(x)))\bar{m}(x) dx ,$$

so that $\bar{m}$ satisfies (a slightly modified version of) the mean field equation (2.2.16). In particular, we have

$$V(x) + G(m(x)) = \min_y V(y) + G(\bar{m}(y)) \text{ for any } x \in Spt(\bar{m}).$$

Let us set $\lambda = \min_y V(y) + G(\bar{m}(y))$. Then

$$\bar{m}(x) = G^{-1}((\lambda - V(x))_+)$$

For instance, if we plug formally $Q = \mathbb{R}^d$, $V(x) = |x|^2/2$ and $G(s) = \log(s)$ into the above equality, we get $m(x) = e^{-|x|^2/2}/(2\pi)^{d/2}$.

2.2.7 Comments

There is a huge literature on games with a continuum of players, starting from the seminal work by Aumann [154]. Schmeidler [260], and then Mas-Colell [251], introduced a notion of non-cooperative equilibrium in games with a continuum of agents and established several existence results in a much more general framework where the agents have *types*, i.e., personal characteristics; in that set-up, the equilibria are known under the name of Cournot-Nash equilibria. Blanchet and Carlier [159] investigated classes of problems in which such equilibrium is unique and can be fully characterized.

The variational approach described in Section 2.2.6 presents strong similarities with the potential games of Monderer and Shapley [253].

2.3 The mean field game system with a non-local coupling

This part is devoted to the mean field game (MFG) system

$$\begin{array}{rl} (i) & -\partial_t u - \Delta u + H(x, Du) = F(x, m) \\ (ii) & \partial_t m - \Delta m - \operatorname{div}(m D_p H(x, Du(t,x))) = 0 \\ (iii) & m(0) = m_0\ ,\ u(T,x) = G(x, m(T)). \end{array} \tag{2.3.1}$$

The Hamiltonian $H : \mathbb{T}^d \times \mathbb{R}^d \to \mathbb{R}$ is assumed to be convex with respect to the second variable. The two equations in (2.3.1) are coupled via the functions F and G. For simplicity, we work with the data which are periodic in space: although this situation is completely unrealistic in terms of applications, this assumption simplifies the proofs and avoids the technical discussion on the boundary conditions. Note that we have set the diffusivity to be equal to one, to simplify the notation. We will also consider the case when the diffusivity vanishes, so the system is of the first order.

The MFG system can be interpreted as a Nash equilibrium for a system for nonatomic agents with a cost (or pay-off) depending of the density of the other agents. More precisely, at the initial time $t = 0$ the agents are distributed according to the probability density m_0. We make the strong assumption that the agents also share a common belief on the future behavior of the density of agents $m(t)$, with, of course, $m(0) = m_0$. Each player starting from a position x at time $t = 0$, has to solve an optimization problem of the form

$$\inf_{\alpha} \mathbb{E}\left[\int_0^T (L(X_s, \alpha_s) + F(X_s, m(s)))ds + G(X_T, m(T))\right]$$

where L is the Legendre transform of H with respect to the last variable:

$$H(p, x) = \inf_{v \in A} \left[L(x, v) + v \cdot p\right], \quad p \in \mathbb{R}^n, \tag{2.3.2}$$

and X_s is the solution to the SDE

$$dX_s = \alpha_s ds + \sqrt{2} dB_s, \qquad X_0 = x.$$

Here, B_s is a standard $d-$dimensional Brownian motion and the infimum is taken over controls $\alpha : [0,T] \to \mathbb{R}^d$ adapted to the filtration generated by B_s. Note that the final cost $G(X_T, m(T))$ depends not only on the final position but also on the distribution of the other players at the final time T, and that the component $F(X_s, m(s))$ of the running cost depends on the position X_s and $m(s)$ but not directly on the control α_s. The cost associated with the control comes only into the Lagrangian $L(X_s, \alpha_s)$.

As it is standard in the control theory, it is convenient to introduce the value function $u(t,x)$ for this problem:

$$u(t,x) := \inf_{\alpha} \mathbb{E}\left[\int_t^T (L(X_s,\alpha_s) + F(X_s, m_s))ds + G(X_T, m_T))\right]$$

where

$$dX_s = \alpha_s ds + \sqrt{2}dB_s,\ X_t = x.$$

As we have discussed, if m_t is known, then u is a classical solution to the Hamilton-Jacobi equation (2.3.1)-(i) with the terminal condition $u(T,x) = G(x, m_T)$. Moreover, the optimal control parameter of each agent is given by

$$\alpha^*(t,x) := -D_pH(x, Du(t,x)).$$

Hence, the best policy for each individual agent at position x at time t, is to "play" (use the control) $\alpha^*(t,x)$. Then, the actual density $\widetilde{m}(t)$ of agents would evolve according to the Fokker-Planck equation (2.3.1)-(ii), with the initial condition $\widetilde{m}(0) = m_0$. We say that the pair (u,m) is a Nash equilibrium of the game if the pair (u,m) satisfies the MFG system (2.3.1). This agrees with our discussion in the previous section.

We discuss here several regimes for the MFG system: first, the uniformly parabolic case, for which existence of a classical solution for the system is expected to hold. When there is no diffusion, one has to introduce a suitable notion of a weak solution. We will also have to consider various smoothing properties of the couplings F and G, depending on whether the couplings are regularizing or not. This is what leads us to separate the "more regularizing" non-local couplings from "not so much regularizing" local couplings.

2.3.1 The existence theorem

Let us start with the second order mean field games with a nonlocal coupling:

$$\begin{array}{rl} (i) & -\partial_t u - \Delta u + H(x, Du) = F(x,m) \qquad \text{in } (0,T)\times\mathbb{T}^d, \\ (ii) & \partial_t m - \Delta m - \operatorname{div}(m\ D_pH(x, Du(t,x))) = 0 \qquad \text{in } (0,T)\times\mathbb{T}^d, \\ (iii) & m(0) = m_0\ ,\ u(T,x) = G(x, m(T)) \qquad \text{in } \mathbb{T}^d. \end{array} \tag{2.3.3}$$

Our aim is to prove the existence of classical solutions for this system and give some interpretation in terms of a game with finitely many players.

Let us describe various assumptions used throughout the section. Our main hypothesis is that F and G are regularizing on the set of probability measures on $\mathbb{T}^d$ in the following sense. Let $\mathcal{P}(\mathbb{T}^d)$ be the set of such Borel probability measures on $\mathbb{T}^d$ endowed with the Kantorovitch-Rubinstein distance:

$$d_1(\mu,\nu) = \sup\left\{\int_{\mathbb{T}^d} \phi(x)(\mu-\nu)(dx) \text{ s.t } \phi : \mathbb{T}^d \to \mathbb{R} \text{ is 1-Lipschitz continuous}\right\}. \tag{2.3.4}$$

Recall that the distance metricizes the weak-* topology on $\mathcal{P}(\mathbb{T}^d)$ and that $\mathcal{P}(\mathbb{T}^d)$ is a compact space.

Here are our main assumptions on F, G and m_0:
(i) The functions $F(x,m)$ and $G(x,m)$ are Lipschitz continuous in $\mathbb{T}^d \times \mathcal{P}(\mathbb{T}^d)$,
(ii) Uniform regularity of F and G in space: $F(\cdot,m)$ and $G(\cdot,m)$ are bounded in $C^{1+\beta}(\mathbb{T}^d)$ and $C^{2+\beta}(\mathbb{T}^d)$ (for some $\beta \in (0,1)$) uniformly with respect to $m \in \mathcal{P}(\mathbb{T}^d)$.
(iii) The Hamiltonian $H : \mathbb{T}^d \times \mathbb{R}^d \to \mathbb{R}$ is locally Lipschitz continuous, D_pH exists and is continuous on $\mathbb{T}^d \times \mathbb{R}^d$, and H satisfies the condition

$$\langle D_x H(x,p), p\rangle \geq -C_0(1+|p|^2) \tag{2.3.5}$$

for some constant $C_0 > 0$.
(iv) The probability measure m_0 is absolutely continuous with respect to the Lebesgue measure, and has a $C^{2+\beta}$ continuous density, still denoted m_0.

Let us comment on the Lipschitz continuity in m assumption. For example, if we fix a Lipschitz function f and take

$$F(m) = \int_{\mathbb{T}^d} f(x)dm,$$

then

$$|F(m_1) - F(m_2)| \leq \|f\|_{Lip} d_1(m_1, m_2),$$

thus $F(m)$ is Lipschitz continuous. On the other hand, if we take a function $g : \mathbb{R} \to \mathbb{R}$ and define $F(x,m) = g(m(x))$ for measures m that are absolutely continuous with respect to the Lebesgue measure, then $F(x,m)$ is not Lipschitz continuous in the d_1-metric on $\mathcal{P}(\mathbb{T}^d)$, no matter how nice g is. This is why this assumption implies that coupling is non-local.

A pair (u,m) is a classical solution to (2.3.3) if $u, m : \mathbb{R}^d \times [0,T] \to \mathbb{R}$ are continuous, of class C^2 in space and C^1 in time and (u,m) satisfies (2.3.3) in the classical sense. The main result of this section is the following:

Theorem 2.3.1 *Under the above assumptions, there is at least one classical solution to (2.3.3).*

The proof is relatively easy and relies on the basic estimates for Hamilton-Jacobi equations and on some remarks on the Fokker-Planck equation (2.3.3-(ii)). We give the details below.

2.3.2 On the Fokker-Planck equation

Let $b : \mathbb{T}^n \times [0,T] \to \mathbb{R}^n$ be a given vector field. Our aim is to analyze the Fokker-Planck equation

$$\begin{aligned} &\partial_t m - \Delta m - \operatorname{div}(mb) = 0, \\ &m(0,x) = m_0(x), \end{aligned} \tag{2.3.6}$$

as an evolution equation in the space of probability measures. We assume here that the vector field $b(t,x)$ is continuous in time and Lipschitz continuous in space.

Definition 2.3.2 *We say that* $m \in L^1(\mathbb{T}^d \times [0,T])$ *is a weak solution to (2.3.6) if for any test function* $\varphi \in C_c^\infty(\mathbb{R}^d \times [0,T))$*, we have*

$$\int_{\mathbb{T}^d} \phi(x,0)dm_0(x) - \int_{\mathbb{T}^d} \phi(x,t)dm(t)(x)$$
$$+ \int_0^T \int_{\mathbb{T}^d} (\partial_t \varphi(t,x) + \Delta\varphi(t,x) - \langle D\varphi(t,x), b(t,x)\rangle)\, dm(t)(x) = 0.$$

In order to analyze some particular solutions of (2.3.6), it is convenient to introduce the following stochastic differential equation (SDE)

$$\begin{aligned} &dX_t = -b(t,X_t)dt + \sqrt{2}dB_t, \qquad t \in [0,T] \\ &X_0 = Z_0. \end{aligned} \tag{2.3.7}$$

Here, the initial condition $Z_0 \in L^1(\Omega)$ is possibly random and independent of the Brownian motion B_t. Under the above assumptions on b, there is a unique solution to (2.3.7). This solution is closely related to the Fokker-Planck equation (2.3.6).

Lemma 2.3.3 *If* $\mathcal{L}(Z_0) = m_0$*, then* $m(t) := \mathcal{L}(X_t)$ *a weak solution of (2.3.6).*

Proof. This is a straightforward consequence of the Itô formula: if $\varphi(t,x)$ is smooth with compact support, then

$$\varphi(t,X_t) = \varphi(0,Z_0) + \int_0^t \left[\partial_s \varphi(s,X_s) - \langle D\varphi(s,X_s), b(X_s,s)\rangle + \Delta\varphi(s,X_s)\right] ds$$
$$+ \int_0^t \langle D\varphi(s,X_s), dB_s\rangle\ .$$

Taking the expectation on both sides (with respect to the Brownian motion and the randomness in the initial condition) gives

$$\mathbb{E}\left[\varphi(t,X_t)\right] = \mathbb{E}\Big[\varphi(0,Z_0) + \int_0^t \left[\varphi_t(s,X_s) - \langle D\varphi(s,X_s), b(s,X_s)\rangle + \Delta\varphi(s,X_s)\right] ds\Big]\ .$$

So by definition of $m(t)$, we get

$$\int_{\mathbb{R}^d} \varphi(t,x)dm(t)(x) = \int_{\mathbb{R}^d} \varphi(0,x)dm_0(x)$$
$$+ \int_0^t \int_{\mathbb{R}^d} \left[\varphi_t(s,x) - \langle D\varphi(s,x), b(s,x)\rangle + \Delta\varphi(s,x)\right] dm(s)(x)ds,$$

thus m is a weak solution to (2.3.6). □

The interpretation of the solution of the continuity equation as the law of the corresponding solution of the SDE allows us to get a Hölder regularity estimate on $m(t)$ in $\mathcal{P}(\mathbb{T}^d)$.

Lemma 2.3.4 *There is a constant* $c_0 = c_0(T)$*, independent of* $\nu \in (0,1]$*, such that*

$$d_1(m(t),m(s)) \le c_0(1 + \|b\|_\infty)|t-s|^{1/2} \qquad \forall s,t \in [0,T]\ .$$

Proof. We write

$$d_1(m(t), m(s)) = \sup\left\{ \int_{\mathbb{T}^d} \phi(x)(m(t) - m(s))(dx) \text{ s.t } \phi \text{ is 1-Lipschitz continuous}\right\}$$
$$\leq \sup\left\{\mathbb{E}\left[\phi(X_t) - \phi(X_s)\right] \text{ s.t } \phi \text{ is 1-Lipschitz continuous}\right\} \leq \mathbb{E}\left[|X_t - X_s|\right].$$

Moreover, if, for instance, $s < t$ we have

$$\mathbb{E}\left[|X_t - X_s|\right] \leq \mathbb{E}\left[\int_s^t |b(X_\tau, \tau)|\, d\tau + \sqrt{2}\,|B_t - B_s|\right] \leq \|b\|_\infty(t-s) + \sqrt{2\nu(t-s)}.$$

This finishes the proof. □

2.3.3 Proof of the existence Theorem

We are now ready to prove Theorem 2.3.1. We fix the initial condition $m_0(x)$ and the terminal condition $G(x, m)$ as in the assumptions of this theorem.

For a large constant C_1 to be chosen below, let $\mathcal{C}$ be the set of maps $\mu \in C^0([0, T], \mathcal{P}(\mathbb{T}^d))$ such that

$$\sup_{s \neq t} \frac{d_1(\mu(s), \mu(t))}{|t - s|^{1/2}} \leq C_1. \tag{2.3.8}$$

Then $\mathcal{C}$ is a convex closed subset of $C^0([0, T], \mathcal{P}(\mathbb{T}^d))$. It is actually compact thanks to Ascoli's Theorem and the compactness of the set $\mathcal{P}(\mathbb{T}^d)$.

The proof is based on a fixed point theorem. To any $\mu \in \mathcal{C}$, we associate $m = \Psi(\mu) \in \mathcal{C}$ as follows. Let $u(t, x)$ be the unique solution to the terminal problem

$$\begin{aligned} &-\partial_t u - \Delta u + H(x, Du) = F(x, \mu(t)) \quad \text{in } (0, T) \times \mathbb{T}^d, \\ &u(x, T) = G(x, \mu(T)) \qquad \text{in } \mathbb{T}^d. \end{aligned} \tag{2.3.9}$$

Then we define $m = \Psi(\mu)$ as the solution of the initial value problem for the Fokker-Planck equation

$$\begin{aligned} &\partial_t m - \Delta m - \operatorname{div}(m D_p H(x, Du)) = 0 \quad \text{in } (0, T) \times \mathbb{T}^d, \\ &m(0, x) = m_0(x) \qquad \text{in } \mathbb{T}^d. \end{aligned} \tag{2.3.10}$$

Let us check that Ψ is a well-defined and continuous map $\mathcal{C} \to \mathcal{C}$. It is convenient to set

$$\widetilde{H}(t, x, p) = H(x, p) - F(x, \mu(t)).$$

The theory of the viscous Hamilton-Jacobi equations shows that under our present assumptions, equation (2.3.9) has a unique classical solution $u(t, x)$. Moreover, because the diffusivity in the viscous Hamilton-Jacobi equation (2.3.9) is strictly positive, we have an estimate

$$\|u\|_{C^{1+\alpha/2, 2+\alpha}([0,T] \times \mathbb{T}^n)} \leq C, \tag{2.3.11}$$

where $\alpha > 0$ and $C > 0$ do not depend on μ, because of the a priori bounds on F we have assumed. Recall that the bounds on $F(x, m)$ are uniform in the probability measure m. The constant C in (2.3.11) may depend on T though.

Next we turn to the Fokker-Planck equation (2.3.10), that we write in the form

$$\partial_t m - \Delta m - \langle Dm, D_pH(x, Du)\rangle - m \operatorname{div}[D_pH(x, Du)] = 0 \ .$$

Since $u \in C^{1+\alpha/2,2+\alpha}([0,T]\times\mathbb{T}^n)$, the maps $(t,x) \to D_pH(x,Du)$ and $(t,x) \to \operatorname{div}D_pH(x,Du)$ belong to $C^\alpha([0,T] \times \mathbb{T}^n)$, so that this advection-diffusion equation is uniquely solvable and the solution m belongs to $C^{2+\alpha,1+\alpha/2}([0,T] \times \mathbb{T}^n)$. Moreover, from Lemma 2.3.4, we have the following estimate on m:

$$d_1(m(t), m(s)) \leq c_0(1 + \|D_pH(\cdot, Du)\|_\infty)|t-s|^{1/2} \qquad \forall s,t \in [0,T],$$

where $\|D_pH(\cdot, Du)\|_\infty$ is bounded by a constant C_2 independent of μ, because Du is uniformly bounded due to (2.3.11). Thus, if we choose C_1 in (2.3.8) sufficiently large, then m belongs to $\mathcal{C}$, and the mapping $\Psi : \mu \to m = \Psi(\mu)$ is well-defined from $\mathcal{C}$ into itself.

Let us check that Ψ is a continuous map $\mathcal{C} \to \mathcal{C}$. Let us assume that $\mu_n \to \mu$ in $\mathcal{C}$, and let (u_n, m_n) and (u, m) be the corresponding solutions to (2.3.9)-(2.3.10). Note that

$$F(x, \mu_n(t)) \to (x, \mu(t)) \text{ and } G(x, \mu_n(T) \to G(x, \mu(T),$$

both uniformly, over $\mathbb{T}^d \times [0,T]$ and $\mathbb{T}^d$, respectively, thanks to our continuity assumptions on F and G. Moreover, as the right side of the Hamilton-Jacobi equation for u_n is bounded in $C^{1+\alpha/2,1+\alpha}([0,T] \times \mathbb{T}^n)$, the functions u_n are uniformly bounded in $C^{1+\alpha/2,2+\alpha}([0,T] \times \mathbb{T}^n)$ so that u_n converges in $C^{2,1}([0,T] \times \mathbb{T}^n)$ to the unique solution $u(t,x)$ of the Hamilton-Jacobi equation with the right side $F(x,\mu)$. The measures m_n are then solutions to a linear Fokker-Planck equation with uniformly Hölder continuous coefficients, which provides uniform $C^{1+\alpha/2,2+\alpha}([0,T]\times]\mathbb{T}^n)$ estimates on m_n. Thus, m_n converge, also in $C^{2,1}([0,T]\times]\mathbb{T}^n)$, to the unique solution m of the Fokker-Planck equation associated to $D_pH(x,Du)$. The convergence is then easily proved to be also in $C^0([0,T], \mathcal{P}(\mathbb{T}^d))$. Now, the Schauder fixed point theorem implies that the continuous map $\mu \to m = \Psi(\mu)$ has a fixed point in $\mathcal{C}$: this fixed point (and the corresponding u) is a solution to (2.3.3).

2.3.4 Uniqueness of the solution

Let us assume that, besides the assumptions given at the beginning of the section, the following monotonicity conditions hold:

$$\int_{\mathbb{T}^d} (F(x,m_1) - F(x,m_2))d(m_1 - m_2)(x) \geq 0 \qquad \forall m_1, m_2 \in \mathcal{P}(\mathbb{T}^d) \tag{2.3.12}$$

and

$$\int_{\mathbb{T}^d} (G(x,m_1) - G(x,m_2))d(m_1 - m_2)(x) \geq 0 \qquad \forall m_1, m_2 \in \mathcal{P}(\mathbb{T}^d) \ . \tag{2.3.13}$$

Physically, the monotonicity properties (2.3.12) and (2.3.13) mean that the cost of visiting places with a high density of other agents is higher than staying in sparsely populated regions.

Such monotonicity properties are easier to fulfill for mappings defined on subsets of $\mathcal{P}(Q)$. For instance, if Q is a compact subset of $\mathbb{R}^d$ of positive measure and $\mathcal{P}_{ac}(Q)$ is the set of absolutely continuous measures on Q, with respect to the Lebesgue measure, then

$$F(y,m) = \begin{cases} G(m(y)) & \text{if } m \in \mathcal{P}_{ac}(Q) \\ +\infty & \text{otherwise} \end{cases}$$

satisfies (2.2.23) as soon as $G : \mathbb{R} \to \mathbb{R}$ is continuous and increasing. Here, we denote by $m(y)$ the density of m at y.

If we assume that Q is the closure of a smooth open bounded subset Ω of $\mathbb{R}^d$, another example is given by

$$F(y, m) = \begin{cases} u_m(y) & \text{if } m \in \mathcal{P}_{ac}(Q) \cap L^2(Q) \\ +\infty & \text{otherwise} \end{cases}$$

where u_m is the solution in $H^1(Q)$ of

$$\begin{cases} -\Delta u_m = m & \text{in } \Omega \\ u_m = 0 & \text{on } \partial\Omega \end{cases}$$

Note that in this case the map $y \to F(y, m)$ is continuous.

We also assume that H is uniformly convex with respect to the momentum variable:

$$\frac{1}{C} I_d \leq D^2_{pp} H(x, p) \leq C I_d, \tag{2.3.14}$$

with some $C > 0$.

Theorem 2.3.5 *Under the above conditions, there is a unique classical solution to the mean field equation (2.3.3).*

Proof. Let (u_1, m_1) and (u_2, m_2) be two classical solutions of (2.3.3), and set

$$\bar{u} = u_1 - u_2, \quad \bar{m} = m_1 - m_2,$$

then

$$\begin{aligned} &\frac{d}{dt} \int_{\mathbb{T}^d} \bar{u}\bar{m} dx = \int_{\mathbb{T}^d} [(\partial_t \bar{u})\bar{m} + \bar{u}(\partial_t \bar{m})] dx \\ &= \int_{\mathbb{T}^d} (-\Delta \bar{u} + H(x, Du_1) - H(x, Du_2) - F(x, m_1) + F(x, m_2))\bar{m} dx \\ &+ \int_{\mathbb{T}^d} \bar{u}(\Delta \bar{m} + \operatorname{div}(m_1 D_p H(x, Du_1)) - \operatorname{div}(m_2 D_p H(x, Du_2))) dx. \end{aligned} \tag{2.3.15}$$

Integration by parts shows that

$$\int_{\mathbb{T}^d} -(\Delta \bar{u})\bar{m} + \bar{u}(\Delta \bar{m}) dx = 0,$$

and, from the monotonicity condition on F, we have

$$\int_{\mathbb{T}^d} (-F(x, m_1) + F(x, m_2))\bar{m} dx = \int_{\mathbb{T}^d} (-F(x, m_1) + F(x, m_2))(m_1 - m_2) dx \leq 0.$$

We now rewrite the remaining terms in the right side of (2.3.15) in the following way:

$$\begin{aligned} R &:= \int_{\mathbb{T}^d} [(H(x, Du_1) - H(x, Du_2))\bar{m} - \langle D\bar{u}, m_1 D_p H(x, Du_1) - m_2 D_p H(x, Du_2)\rangle] dx \\ &= -\int_{\mathbb{T}^d} m_1 \left[H(x, Du_2) - H(x, Du_1) - \langle D_p H(x, Du_1), Du_2 - Du_1\rangle\right] dx \\ &- \int_{\mathbb{T}^d} m_2 \left[H(x, Du_1) - H(x, Du_2) - \langle D_p H(x, Du_2), Du_1 - Du_2\rangle\right] dx. \end{aligned}$$

The uniform convexity assumption (2.3.14) on H implies that

$$R \leq -\int_{\mathbb{T}^d} \frac{(m_1+m_2)}{2C}|Du_1 - Du_2|^2 dx \leq 0.$$

Putting the estimates together we get

$$\frac{d}{dt}\int_{\mathbb{T}^d} \bar{u}\bar{m} dx \leq 0. \tag{2.3.16}$$

We integrate this inequality on the time interval $[0,T]$ to obtain

$$\int_{\mathbb{T}^d} \bar{u}(T)\bar{m}(T)dx \leq \int_{\mathbb{T}^d} \bar{u}(0)\bar{m}(0)dx - \int_0^T\int_{\mathbb{T}^d} \frac{(m_1+m_2)}{2C}|Du_1 - Du_2|^2 dx. \tag{2.3.17}$$

Note that $\bar{m}(0) = 0$ while, as $\bar{u}(T) = G(x, m_1(T)) - G(x, m_2(T))$, we have

$$\int_{\mathbb{T}^d} \bar{u}(T)\bar{m}(T)dx = \int_{\mathbb{T}^d} (G(x, m_1(T)) - G(x, m_2(T)))(m_1(T) - m_2(T))dx \geq 0$$

thanks to the monotonicity assumption on G. Now, (2.3.17) implies that

$$\int_{\mathbb{T}^d} \bar{u}(T)\bar{m}(T)dx = 0,$$

but also that

$$Du_1 = Du_2 \text{ in } \{m_1 > 0\}\bigcup\{m_2 > 0\}.$$

As a consequence, m_2 actually solves the same equation as m_1, with the same drift

$$D_pH(x, Du_1) = D_pH(x, Du_2),$$

hence $m_1 = m_2$. This, in turn, implies that u_1 and u_2 solve the same Hamilton-Jacobi equation, so that $u_1 = u_2$. □

2.3.5 An application to games with finitely many players

A single player strategy in an MFG soup

We now discuss the implications of the solutions of the MFG system to games with a large but finite number of players. Let us fix a solution $u(t,x), m(t,x)$ to the mean field system (2.3.3) and investigate the optimal strategy of a generic player who considers the density $m(t,x)$ "of the other players" as given. In other words, the player faces the following minimization problem

$$\inf_{\alpha} \mathcal{J}(\alpha) \qquad \text{where} \qquad \mathcal{J}(\alpha) = \mathbb{E}\left[\int_0^T L(X_s, \alpha_s) + F(X_s, m(s))\ ds + G(X_T, m(T))\right]. \tag{2.3.18}$$

Here, $L(x,v)$ is a kind of Legendre transform of H with respect to the last variable:

$$L(x,v) := \sup_{p\in\mathbb{R}^d}[-\langle p, v\rangle - H(x,p)].$$

Let us stress that the density $m(s,x)$ in (2.3.18) is assumed to be given and be a solution to the MFG system (2.3.3). The process X_t in (2.3.18) is given by

$$X_t = X_0 + \int_0^t \alpha_s ds + \sqrt{2} B_s,$$

with X_0 a fixed random initial condition with the law m_0, independent of B_t. The control α_t is adapted to the filtration $\mathcal{F}_t$ of the $d-$dimensional Brownian motion B_t. In this model, the player does not have any information about the other players, and simply assumes that they all follow the control prescribed by the solution to the MFG system, so that that their density also evolves according to the solution $m(t,x)$ to that system. We claim that then the strategy $\alpha^*(t,x) := -D_pH(x, Du(t,x)$ is optimal for this stochastic control problem for the single player. This confirms that the solution to the MFG system is, in some loose sense, a Nash equilibrium.

Lemma 2.3.6 *Let $\bar{X}_t$ be the solution of the stochastic differential equation*

$$\begin{cases} d\bar{X}_t = \alpha^*(t, \bar{X}_t)dt + \sqrt{2} dB_t \\ \bar{X}_0 = X_0 \end{cases}$$

and set $\bar{\alpha}(t) = \alpha^*(t, X_t)$. Then

$$\inf_\alpha \mathcal{J}(\alpha) = \mathcal{J}(\bar{\alpha}) = \int_{\mathbb{R}^N} u(0,x)\ dm_0(x)\ .$$

Proof. This kind of result is known as a verification Theorem: one has a good candidate for an optimal control, and one checks, using the equation satisfied by the value function $u(t,x)$, that this is indeed the minimum. Let α be an adapted control. We have, by the Itô formula, applied to the function $u(t, X_t)$

$$\begin{aligned} \mathbb{E}[G(X_T, m(T))] &= \mathbb{E}[u(X_T, T)] \\ &= \mathbb{E}\left[u(0, X_0) + \int_0^T (\partial_t u(s, X_s) + \langle \alpha_s, Du(s, X_s)\rangle + \Delta u(s, X_s))\ ds\right] \\ &= \mathbb{E}\left[u(0, X_0) + \int_0^T (H(X_s, Du(s, X_s)) + \langle \alpha_s, Du(s, X_s)\rangle - F(X_s, m(s)))\ ds\right]. \end{aligned}$$

We have used the Hamilton-Jacobi equation satisfied by $u(t,x)$ in the last equality. Thus, by definition of $L(x,v)$, we have

$$\mathbb{E}[G(X_T, m(T))] \geq \mathbb{E}\left[u(0, X_0) + \int_0^T (-L(X_s, \alpha_s) - F(X_s, m(s)))\ ds\right].$$

This shows that

$$\mathbb{E}\left[u(0, X_0)\right] \leq \mathbb{E}\left[\int_0^T (L(X_s, \alpha_s) + F(X_s, m(s)))\ ds + G(X_T, m(T))\right] = J(\alpha) \qquad (2.3.19)$$

for any adapted control α. Let us replace α_s by $\bar{\alpha}_s = -D_pH(\bar{X}_s, Du(s, \bar{X}_s)$ in the above computations. Then, since

$$\begin{aligned} &H(\bar{X}_s, Du(s, \bar{X}_s)) + \langle \bar{\alpha}_s, Du(s, \bar{X}_s)\rangle\ = H(\bar{X}_s, Du(s, \bar{X}_s)) + \langle \alpha^*(\bar{X}_s), Du(s, \bar{X}_s))\rangle \\ &= -L(\bar{X}_s, \alpha^*(\bar{X}_s, Du(s, \bar{X}_s))) = -L(\bar{X}_s, \bar{\alpha}_s)), \end{aligned} \qquad (2.3.20)$$

all the above inequalities become equalities, so that

$$\mathbb{E}\left[u(X_0, 0)\right] = \mathcal{J}(\bar{\alpha}). \tag{2.3.21}$$

This, together with (2.3.19) shows that $\mathcal{J}(\bar{\alpha}) \le J(\alpha)$ for any adapted control α_s. $\square$

A game with a large number of players

We now consider a differential game with a large but finite number of N players and ask if the mean field game model is a good approximation for it in any sense. In this game, each player $i = 1, \dots, N$, is controlling, through the corresponding control α^i, a dynamics of the form

$$dX_t^i = \alpha_t^i dt + \sqrt{2} dB_t^i. \tag{2.3.22}$$

The initial conditions X_0^i for this system are also random and all have the same law m_0. We assume that all X_0^i and all the Brownian motions B_t^i, $i = 1, \dots, N$, are independent. Player i can choose its control α^i adapted to the full filtration $\mathcal{F}_t = \sigma\{X_0^j,\ B_s^j,\ s \le t,\ j = 1, \dots, N\}$, In other words, the players "know about each other". The payoff of the player i is then given by

$$\mathcal{J}_i^N(\alpha^1, \dots, \alpha^N) = \mathbb{E}\left[\int_0^T L(X_s^i, \alpha_s^i) + F(X_s^i, m_{X_s}^{N,i}) ds + G(X_T^i, m_{X_T}^{N,i})\right],$$

where

$$m_{X_s}^{N,i} := \frac{1}{N-1} \sum_{j \ne i} \delta_{X_s^j}$$

is the empirical distribution of the players X^j, with $j \ne i$. Our aim is to explain that the strategy given by the mean field game nearly gives a Nash equilibrium for this problem. More precisely, let $u(t,x)$, $m(t,x)$ be a classical solution to the MFG system (2.3.3) and let us define the control

$$\alpha^*(t,x) := -D_p H(x, Du(t,x)).$$

Given $\alpha^*(t,x)$ we can define the control $\bar{\alpha}^i$ obtained by solving the SDE

$$d\bar{X}_t^i = \alpha^*(t, \bar{X}_t^i) dt + \sqrt{2} dB_t^i \tag{2.3.23}$$

with random initial condition X_0^i and setting $\bar{\alpha}_t^i = \alpha^*(t, \bar{X}_t^i)$. Note that this control is adapted to the filtration $\mathcal{F}_t^i = \sigma(X_0^i,\ B_s^i,\ s \le t\}$, and does not use the information in the full filtration $\mathcal{F}_t$ defined above – the players do not use the precise information about the other players. That is, each player simply uses the control generated by the MFG system and then solves the SDE (2.3.23), completely oblivious to what the other players do. Of course, a strong assumption here is that this is what everyone does. However, as we have seen in Lemma 2.3.6, we do know that if everyone else uses the MFG strategy, then this is also the best strategy for a single player in such soup.

Theorem 2.3.7 *Assume that F and G are Lipschitz continuous in $\mathbb{T}^d \times P(\mathbb{T}^d)$. Then there exists a constant $C > 0$ such that the strategy $(\bar{\alpha}^1, \dots, \bar{\alpha}^N)$ is an $\varepsilon-$Nash equilibrium in the game $\mathcal{J}_1^N, \dots, \mathcal{J}_N^N$ for $\varepsilon := CN^{-1/(d+4)}$: namely*

$$\mathcal{J}_i^N(\bar{\alpha}^1, \dots, \bar{\alpha}^N) \le \mathcal{J}_i^N((\bar{\alpha}^j)_{j \ne i}, \alpha^i) + CN^{-1/(d+4)}$$

for any control α^i adapted to the filtration $(\mathcal{F}_t)$ and any $i \in \{1, \dots, N\}$.

The Lipschitz continuity assumptions on F and G allow to quantify the error. If F and G are just continuous, one can only say that, for any $\varepsilon > 0$, there exists N_0 such that the symmetric strategy $(\bar{\alpha}^1, \dots, \bar{\alpha}^N)$ is an $\varepsilon-$Nash equilibrium in the game $\mathcal{J}_1^N, \dots, \mathcal{J}_N^N$ for any $N \geq N_0$.

Before starting the proof, we need the following result on product measures due to Horowitz and Karandikar (see for instance Rashev and Rüschendorf [258], Theorem 10.2.1).

Lemma 2.3.8 *Assume that Z_i are i.i.d. random variables with a law μ. Then there is a constant C, depending only on d, such that*

$$\mathbb{E}[d_1(m_Z^N, \mu)] \leq CN^{-1/(d+4)}, \qquad \text{where } m_Z^N = \sum_{i=1}^N \delta_{Z_i}.$$

Proof of Theorem 2.3.7. Fix $\varepsilon > 0$. Since the problem is symmetric, it is enough to show that

$$\mathcal{J}_1^N(\bar{\alpha}^1, \dots, \bar{\alpha}^N) \leq \mathcal{J}_1^N((\bar{\alpha}^j)_{j\neq 1}, \alpha) + \varepsilon \tag{2.3.24}$$

for any control α, as soon as N is large enough. Recall that $\bar{X}_t^j$ is the solution of the stochastic differential equation (2.3.23) with the initial condition X_0^j. We note that $\bar{X}_t^j$ are independent and identically distributed with the law $m(t)$ – see Lemma 2.3.3. Therefore, using Lemma 2.3.8, we have for any $t \in [0, T]$,

$$\mathbb{E}\left[d_1(m_{\bar{X}_t}^{N,i}, m(t))\right] \leq CN^{-1/(d+4)}.$$

By the Lipschitz continuity of F and G with respect to the variable m, we have therefore:

$$\mathbb{E}\left[\int_0^T \sup_{x\in\mathbb{T}^d} |F(x, m_{\bar{X}_t}^{N,1}) - F(x, m(t))|dt\right] + \mathbb{E}\left[\sup_{x\in\mathbb{T}^d} |G(x, m_{\bar{X}_T}^{N,1}) - G(x, m(T))|\right] \leq CN^{-1/(d+4)}.$$

Let now α^1 be a control adapted to the filtration $\mathcal{F}_t$ and X_t^1 be the solution to

$$dX_t^1 = \alpha_t^1 dt + \sqrt{2} dB_t^1$$

with a random initial condition X_0^1. We have

$$\begin{aligned}
\mathcal{J}_1^N((\bar{\alpha}^j)_{j\neq 2}, \alpha^1) &= \mathbb{E}\Big[\int_0^T (L(X_s^1, \alpha_s^1) + F(X_s^1, m_{\bar{X}_s}^{N,i}))\, ds + G(X_T^1, m_{\bar{X}_T}^{N,i})\Big] \\
&\geq \mathbb{E}\Big[\int_0^T (L(X_s^1, \alpha_s^1) + F\left(X_s^1, m(s)\right))\, ds + G\left(X_T^1, m(T)\right)\Big] - CN^{-1/(d+4)} \\
&\geq \mathcal{J}_1^N((\bar{\alpha}^j)_{j\neq 1}, \bar{\alpha}^1) - CN^{-1/(d+4)}.
\end{aligned}$$

The last inequality comes from the optimality of $\bar{\alpha}$ in Lemma 2.3.6. This proves the result. □

2.3.6 Extensions

Several other classes of MFG systems have be studied in the literature. We discuss only a few of them, since the number of models has grown exponentially in the last years.

The ergodic MFG system

One may be interested in the large time average of the MFG system (2.3.3) as the horizon T tends to infinity. It turns out that the limit system takes the following form:

$$\begin{array}{ll} (i) & \lambda - \Delta u + H(x, Du) = F(x, m) \qquad \text{in } \mathbb{T}^d, \\ (ii) & -\Delta m - \operatorname{div}(m\ D_pH(x, Du(x))) = 0 \qquad \text{in } \mathbb{T}^d. \end{array} \tag{2.3.25}$$

Here the unknown are now (λ, u, m), where $\lambda \in \mathbb{R}$ is the so-called ergodic constant. The interpretation of the system is the following: each player wants to minimize his ergodic cost

$$\mathcal{J}(x, \alpha) := \limsup_{T\to+\infty} \inf_{\alpha} \mathbb{E}\left[\frac{1}{T}\int_0^T [H^*(X_t, -\alpha_t) + F(X_t, m(t))]dt\right]$$

where X_t in the solution to

$$\begin{aligned} &dX_t = \alpha_t dt + \sqrt{2}dB_t \\ &X_0 = x. \end{aligned} \tag{2.3.26}$$

It turns out that, if (λ, u, m) is a classical solution to (2.3.25), then the optimal strategy of each tiny player is given by the feedback

$$\alpha^*(t, x) := -D_pH(x, Du(x))$$

and, if $\bar{\alpha}$ is the solution to

$$\left\{\begin{array}{l} dX_t = \alpha^*(t, X_t)dt + \sqrt{2}dB_t \\ X_0 = x \end{array}\right. \tag{2.3.27}$$

and if we set $\bar{\alpha}_t := \alpha^*(t, X_t)$, then $\mathcal{J}(x, \bar{\alpha}) = \lambda$ is independent of the initial position. Finally, m is the invariant measure associated with the SDE (2.3.27).

The infinite horizon problem

Another natural model pops up when each player aims at minimizing a infinite horizon cost:

$$\mathcal{J}(x, \alpha) = \inf_{\alpha} \mathbb{E}\left[\int_0^{+\infty} e^{-rt}\left(H^*(X_t, -\alpha_t) + F(X_t, m(t))\right)dt\right]$$

where $r > 0$ is a fixed discount rate. Note that there is no reason for the equilibrium for been given by the initial repartition of the players. This implies that the infinite horizon MFG system is *not* stationary. It is actually a system of evolution equations in infinite horizon, given by:

$$\begin{array}{ll} (i) & -\partial_t u + ru - \Delta u + H(x, Du) = F(x, m(t)) \qquad \text{in } (0, +\infty) \times \mathbb{T}^d \\ (ii) & \partial_t m - \Delta m - \operatorname{div}(m\ D_pH(x, Du(t, x))) = 0 \qquad \text{in } (0, +\infty) \times \mathbb{T}^d \\ (iii) & m(0) = m_0 \text{ in } \mathbb{T}^d, \qquad u(t, x) \text{ bounded.} \end{array} \tag{2.3.28}$$

Note that for the infinite horizon problem the terminal condition for $u(t, x)$ above is replaced by the requirement that $u(t, x)$ is bounded.

2.3.7 Comments

Existence: Existence of solutions for the MFG system can be achieved either by Banach fixed point Theorem (as in the papers by Caines, Huang and Malham [228], under a smallness assumption on the coefficients or on the time interval) or by Schauder arguments (as in Theorem 2.3.1, due to Lasry and Lions [244, 243]). Carmona and Delarue [184] use a stochastic maximum principle to derive an MFG system which takes the form of a system of forward-backward stochastic differential equations of a McKean-Vlasov type.

Uniqueness: Concerning the uniqueness of the solution, one can distinguish two kinds of regimes. Of course the Banach fixed point argument provides directly uniqueness of the solution of the MFG system. However, as explained above, it mostly concerns local in time results. For the large time uniqueness, one can rely on the monotonicity conditions (2.3.12) and (2.3.13). These conditions first appear in Lasry and Lions [244, 243].

Nash equilibria for the $N-$player games: the use of the MFG system to obtain $\varepsilon-$*Nash equilibria* (Theorem 2.3.7) has been initiated—in a slightly different framework—in a series of papers due to Caines, Huang and Malham: see in particular [226] (for linear dynamics) and [228] (for nonlinear dynamics). In these papers, the dependence with respect of the empirical measure of dynamics and payoff occurs through an average, so that the CTL implies that the error term is a order $N^{-1/2}$ (instead of $N^{-1/(d+4)}$ as in Theorem 2.3.7). The genuinely non linear version of the result given above is a variation on a result by Carmon and Delarue [184].

We discuss below the reverse statement: in what extend the MFG system pops up as the limit of Nash equilibria.

Extensions: it is difficult to discuss all the extensions of the MFG systems since the number of papers on this subject has grown exponentially in the last years. We give here only a brief overview.

The ergodic MFG system has been introduced by Lasry and Lions in [245] as the limit, when the number of players tends to infinity, of Nash equilibria in ergodic differential games. As explained in Lions [248], this system also pops up as the limit, as the horizon tends to infinity, of the finite horizon MFG system. We discuss this convergence in the next section, in a slightly simpler setting.

The natural issue of boundary conditions has not been thoroughly investigated up to now. For the PDE approach, the authors have mostly worked with periodic data (as we did above), which completely eliminates this question. In the "probabilistic literature" (as in the work by Caines, Huang and Malham), the natural set-up is the full space. Beside these two extreme cases, little has been written (see however Cirant [188], for Neumann boundary condition in ergodic multi-population MFG systems).

The interesting MFG systems with several populations were introduced in the early paper by Caines, Huang and Malham [228] and revisited by Cirant [188] (for Neuman boundary conditions) and by Kolokoltsov, Li and Yang [235] (for very general diffusions, possibly with jumps).

A very general MFG model for a single population is described in Gomes, Patrizi and

Voskanyan [207] and Gomes and Voskanyan[208], in which the velocity of the population is a nonlocal function of the (repartition of) actions of the players.

2.4 Second order MFG systems with a local coupling

In this section, we consider the MFG system with a local coupling:

$$\begin{array}{ll} (i) \;\; -\partial_t u - \Delta u + H(x,Du) = f(x,m(x,t)), & \text{in } \mathbb{T}^d\times(0,T), \\ (ii) \;\; \partial_t m - \;\Delta m - \operatorname{div}(D_pH(x,Du)m) = 0 & \text{in } \mathbb{T}^d\times(0,T) \\ (iii) \;\; m(0,x) = m_0(x)\ ,\ u(x,T) = G(x). & \end{array} \tag{2.4.1}$$

Here, the Hamiltonian $H:\mathbb{T}^d\times\mathbb{R}^d\to\mathbb{R}$ is as before but the map $f:\mathbb{T}^d\times[0,+\infty)\to\mathbb{R}$ is now a local coupling between the value function of the optimal control problem and the density of the distribution of the players. Our aim is first to show that the problem has a unique solution under suitable assumptions on H and a monotonicity condition on f. Then we explain that the system (2.4.1) can be interpreted as an optimality condition of two optimal control problems of partial differential equations. We complete the section by the analysis of the long time average of the system and its link with the ergodic MFG system.

2.4.1 Existence of a solution

Let us assume that the coupling $f:\mathbb{T}^d\times[0,+\infty)\to\mathbb{R}$ is smooth (say, C^3) and that the initial and terminal conditions m_0 and G are $C^{2+\beta}$.

Theorem 2.4.1 *Under the above assumptions, if*

- *either the Hamiltonian is quadratic:* $H(x,p)=\frac{1}{2}|p|^2$*, and the coupling* f *is bounded,*
- *or* H *is of the class* C^2 *and globally Lipschitz continuous,*

then (2.4.1) has at least one classical solution.

Proof. We first assume that the Hamiltonian is quadratic and f is bounded. Let us mollify f to turn the coupling into a non-local one. We take a smooth nonnegative kernel $\xi(x)$ with compact support such that

$$\int_{\mathbb{R}^d}\xi(x)dx = 1,$$

and define, for any $m\in P(\mathbb{T}^d)$,

$$f^\varepsilon(x,m) = f(x,\xi^\varepsilon\star m),\quad \xi_\varepsilon(s)=\varepsilon^{-d}\xi(s/\varepsilon).$$

Now, f^ε is defined on all $m\in\mathcal{P}(\mathbb{T}^d)$, not only measures absolutely continuous with respect to the Lebesgue measure. As f^ε is regularizing, Theorem 2.3.1 states that the system

$$\begin{array}{ll} (i) \;\; -\partial_t u^\varepsilon - \Delta u^\varepsilon + \frac{1}{2}|Du^\varepsilon|^2 = f^\varepsilon(x,m^\varepsilon), & \text{in } \mathbb{T}^d\times(0,T) \\ (ii) \;\; \partial_t m^\varepsilon - \Delta m^\varepsilon - \operatorname{div}(m^\varepsilon Du^\varepsilon) = 0, & \text{in } \mathbb{T}^d\times(0,T) \\ (iii) \;\; m^\varepsilon(0)=m_0\ ,\ u^\varepsilon(x,T)=G(x), & \end{array} \tag{2.4.2}$$

has at least one classical solution. In order to proceed, one needs estimates on this solution, uniform in $\varepsilon \in (0,1)$. Note that, in view of the boundedness condition on f, the term $f^\varepsilon(x, m^\varepsilon)$ is uniformly bounded. So, by the maximum principle, u^ε are also uniformly bounded:

$$\|u^\varepsilon\|_\infty \leq C.$$

Here, C depends on $\|f\|_\infty$ and T. We now use the Hopf-Cole transform, setting $w^\varepsilon = e^{-u^\varepsilon/2}$. A straightforward computation shows that w^ε solves

$$\begin{aligned} &(i)\ -\partial_t w^\varepsilon - \Delta w^\varepsilon + w^\varepsilon f^\varepsilon(x, m^\varepsilon) = 0, \qquad \text{in } \mathbb{T}^d \times (0,T), \\ &(ii)\ w^\varepsilon(x,T) = e^{-G(x)/2}. \end{aligned} \tag{2.4.3}$$

Since $u^\varepsilon(t,x)$ are uniformly bounded, so are $w^\varepsilon(t,x)$. The standard estimates on the linear equations imply the Hölder bounds on w^ε and Dw^ε:

$$\|w^\varepsilon\|_{C^{\alpha,\alpha/2}} + \|Dw^\varepsilon\|_{C^{\alpha,\alpha/2}} \leq C,$$

where α and C depends only on the L^∞-bound on f and on the $C^{2+\beta}$ regularity of G. As u^ε is bounded, we immediately derive similar estimates for u^ε:

$$\|u^\varepsilon\|_{C^{\alpha,\alpha/2}} + \|Du^\varepsilon\|_{C^{\alpha,\alpha/2}} \leq C. \tag{2.4.4}$$

Next we estimate m^ε: as m^ε solves the linear equation (2.4.2)-(ii) that is in the divergence form, a standard estimate implies that m^ε are bounded in Hölder norm:

$$\|m^\varepsilon\|_{C^{\alpha,\alpha/2}} \leq C.$$

Accordingly the coefficients of (2.4.3) are bounded in $C^{\alpha,\alpha/2}$ because f is smooth in both arguments. Now, we can bootstrap (2.4.4) to obtain a $C^{2+\alpha,1+\alpha/2}$ estimate of the solution w^ε, which can be rewritten as an estimate on u^ε:

$$\|u^\varepsilon\|_{C^{2+\alpha,1+\alpha/2}} \leq C. \tag{2.4.5}$$

In turn, m^ε solve an equation with Hölder continuous coefficients, therefore one has $C^{2+\alpha,1+\alpha/2}$ estimates on m^ε. So we can extract a subsequence of the $(m^\varepsilon, u^\varepsilon)$ which converges in $C^{2,1}$ to (m,u), where (m,u) is a solution to (2.4.1).

Let us now explain the proof when H is of class C^2 and is globally Lipschitz continuous. The idea is basically the same: let $(m^\varepsilon, u^\varepsilon)$ be a solution of the equation with a regularized right side:

$$\begin{aligned} (i)\quad & -\partial_t u^\varepsilon - \Delta u^\varepsilon + H(x, Du^\varepsilon) = f^\varepsilon(x, m^\varepsilon) \qquad \text{in } \mathbb{T}^d \times (0,T) \\ (ii)\quad & \partial_t m^\varepsilon - \Delta m^\varepsilon - \operatorname{div}(m^\varepsilon D_p H(x, Du^\varepsilon)) = 0 \qquad \text{in } \mathbb{T}^d \times (0,T) \\ (iii)\quad & m^\varepsilon(0) = m_0\,,\ u^\varepsilon(x,T) = G(x). \end{aligned} \tag{2.4.6}$$

As $D_p H(x, Du^\varepsilon)$ is globally bounded, m^ε solves a linear equation with bounded coefficients: therefore m^ε is bounded in Hlder norm. Then we come back to (2.4.6)-(i), which has a right-hand side bounded in the Hölder norm: this implies that the solution u^ε is bounded in $C^{2+\alpha,1+\alpha/2}$. One can then conclude as before. □

2.4.2 Uniqueness of a solution

We now discuss uniqueness issues. For doing so, we work in a very general framework and exhibit a structure condition on a coupled Hamiltonian $H : \mathbb{T}^d \times \mathbb{R}^d \times [0, +\infty) \to \mathbb{R}$ for uniqueness of classical solutions $(u, m) : [0, T] \times \mathbb{R}^d \to \mathbb{R}^2$ to the local MFG system:

$$\begin{array}{rl} (i) & -\partial_t u - \Delta u + H(x, Du, m) = 0, \\ (ii) & \partial_t m - \Delta m - \operatorname{div}(m\ D_pH(x, Du, m)) = 0, \\ (iii) & m(0, x) = m_0(x),\ u(T, x) = G(x) \qquad \text{in } \mathbb{T}^d. \end{array} \tag{2.4.7}$$

In the above system, $H = H(x, p, m)$ is a Hamiltonian that is convex in p and depends on the density m, the function $G : \mathbb{T}^d \to \mathbb{T}$ is smooth, and m_0 is a probability density on $\mathbb{R}^d$.

Theorem 2.4.2 *Assume that $H(x, p, m)$ is a C^2 function, such that*

$$\begin{pmatrix} m\partial^2_{pp}H & \frac{1}{2}m\partial^2_{pm}H \\ \frac{1}{2}m(\partial^2_{pm}H)^T & -\partial_m H \end{pmatrix} > 0, \quad \text{for all } (x, p, m) \text{ with } m > 0. \tag{2.4.8}$$

Then the system (2.4.7) has at most one classical solution.

Remark 2.4.3
1. Condition (2.4.8) implies, in particular, that $H(x, p, m)$ is uniformly convex with respect to p and strictly decreasing with respect to m but these conditions are not sufficient for it to hold.
2. For a separate $H(x, p, m)$, of the form $H(x, p, m) = \widetilde{H}(x, p) - f(x, m)$, condition (2.4.8) reduces to $D^2_{pp}\widetilde{H} > 0$ and $D_m f > 0$, so that the above conditions become sufficient.

Before starting the proof of Theorem 2.4.2, let us reformulate condition (2.4.8) in a more convenient way (omitting the x dependence for simplicity):

Lemma 2.4.4 *Condition (2.4.8) implies the inequality*

$$(H(p_2, m_2) - H(p_1, m_1))(m_2 - m_1) - \langle p_2 - p_1, m_2 D_pH(p_2, m_2) - m_1 D_pH(p_1, m_1)\rangle \leq 0, \tag{2.4.9}$$

with equality if and only if $(m_1, p_1) = (m_2, p_2)$.

Remark 2.4.5 In fact the above implication is almost an equivalence, in the sense that, if (2.4.9) holds, then

$$\begin{pmatrix} m\partial^2_{pp}H & \frac{1}{2}m\partial^2_{pm}H \\ \frac{1}{2}m(\partial^2_{pm}H)^T & -\partial_m H \end{pmatrix} \geq 0$$

Proof of Lemma 2.4.4. Set $\widetilde{p} = p_2 - p_1$, $\widetilde{m} = m_2 - m_1$ and, for $\theta \in [0,1]$,

$$p_\theta = p_1 + \theta(p_2 - p_1), \quad m_\theta = m_1 + \theta(m_2 - m_1).$$

Let us consider

$$I(\theta) = (H(p_\theta, m_\theta) - H(p_1, m_1))\widetilde{m} - \langle \widetilde{p}, m_\theta D_p H(p_\theta, m_\theta) - m_1 D_p H(p_1, m_1)\rangle.$$

Then, we have $I(0) = 0$ and

$$\begin{aligned} I'(\theta) &= (\langle D_p H, \widetilde{p}\rangle + \widetilde{m} H_m)\widetilde{m} - \widetilde{m}\langle \widetilde{p}, D_p H\rangle - m_\theta \widetilde{p}_k \widetilde{p}_j \frac{\partial^2 H(p_\theta, m_\theta)}{\partial p_k \partial p_j} - \widetilde{m} m_\theta \widetilde{p}_k \frac{\partial H(p_\theta, m_\theta)}{\partial p_k \partial m} \\ &= -\left(\widetilde{p}^T \ \widetilde{m}\right) \begin{pmatrix} m_\theta \ \partial^2_{pp} H & \frac{1}{2} m_\theta \ \partial^2_{pm} H \\ \frac{1}{2} m_\theta \ (\partial^2_{pm} H)^T & -\partial_m H \end{pmatrix} \begin{pmatrix} \widetilde{p} \\ \widetilde{m} \end{pmatrix}. \end{aligned} \tag{2.4.10}$$

Hence if condition (2.4.8) holds and $(p_1, m_1) \neq (p_2, m_2)$, then

$$0 > I(1) = (H(p_2, m_2) - H(p_1, m_1))(m_2 - m_1) - \langle p_2 - p_1, m_2 D_p H(p_2, m_2) - m_1 D_p H(p_1, m_1)\rangle,$$

finishing the proof. □

Proof of Theorem 2.4.2. Let (u_1, m_1) and (u_2, m_2) be solutions to (2.4.7). Let us set

$$\widetilde{m} = m_2 - m_1, \ \widetilde{u} = u_2 - u_1, \ \widetilde{H} = H(x, Du_2, m_2) - H(x, Du_1, m_1),$$
$$\widetilde{\operatorname{div}} = \operatorname{div}(m_2 D_p H(x, Du_2, m_2)) - \operatorname{div}(m_1 D_p H(x, Du_1, m_1))$$

Then, we have

$$\begin{aligned} &\frac{d}{dt}\int_{\mathbb{T}^d} (u_2(t) - u_1(t))(m_2(t) - m_1(t))dx \\ &= \int_{\mathbb{T}^d} \Big[(\partial_t u_2 - \partial_t u_1)(m_2 - m_1) + (u_2 - u_1)(\partial_t m_2 - \partial_t m_1)\Big]dx \\ &= \int_{\mathbb{T}^d} \Big[(-\nu\Delta\widetilde{u} + \widetilde{H})\widetilde{m} + \widetilde{u}(\nu\Delta\widetilde{m} + \widetilde{\operatorname{div}})\Big]dx \\ &= \int_{\mathbb{T}^d} \Big[\widetilde{H}\widetilde{m} - \langle D\widetilde{u}, m_2 D_p H(Du_2, m_2) - m_1 D_p H(Du_1, m_1)\rangle\Big]dx \ \leq \ 0, \end{aligned} \tag{2.4.11}$$

by condition (2.4.8) and Lemma 2.4.4. Due to the terminal and initial conditions we have both $u_1(T, x) = u_2(T, x) = G(x)$ and $m_1(0, x) = m_2(0, x) = m_0(x$. It follows that

$$0 = \Big[\int_{\mathbb{T}^d} (u_2(t) - u_1(t))(m_2(t) - m_1(t))\Big]_0^T.$$

Thus, integrating (2.4.23) between 0 and T gives

$$\int_0^T \int_{\mathbb{T}^d} \Big(\widetilde{H} \ \widetilde{m} - \langle D\widetilde{u}, m_2 D_p H(Du_2, m_2) - m_1 D_p H(Du_1, m_1)\rangle\Big)dxdt = 0.$$

In view of Lemma 2.4.4, this implies that $D\widetilde{u} = 0$ and $\widetilde{m} = 0$, so that $m_1 = m_2$ and $u_1 = u_2$. □

2.4.3 Optimal control interpretation of an MFG system

Here, we show that the MFG system (2.4.1) can be related to two genuine optimal control problems: the first one is an optimal control of the Hamilton-Jacobi equation and the second one concerns the optimal control of the Fokker-Planck equation.

To motivate this discussion, let us recall that, long before we have introduced the MFG system, we have already encountered a dynamical system that consists of two evolution equations, with the initial condition prescribed for one of them and a terminal condition prescribed for the other, which is a salient feature of the MFG systems. This happened when we considered the optimal control problem

$$v(t,x) = \inf \Big\{ \int_0^t L(\gamma(s), \dot\gamma(s))ds + u_0(\gamma(0)) : \ \gamma(t) = x. \Big\}. \tag{2.4.12}$$

In that case, the optimal trajectory $\bar\gamma(s)$, together with $\bar p(s) = \nabla v(\gamma(s))$ satisfied the Hamiltonian system

$$\begin{aligned} \frac{d\bar\gamma(s)}{ds} &= \nabla_p \widetilde{H}(\bar\gamma(s), \bar p(s)),\\ \frac{d\bar p(s)}{ds} &= -\nabla_x \widetilde{H}(\bar\gamma(s), \bar p(s)), \end{aligned} \tag{2.4.13}$$

with the initial condition $\bar p(0) = \nabla u_0(\bar\gamma(0))$ for the momentum variable and the terminal condition $\bar\gamma(t) = x$ for the position. Recall that the Hamiltonian $\widetilde{H}(x,p)$ that appears in (2.4.13) is the Legendre transform of the Lagrangian $L(x,v)$:

$$H(x,p) = \sup_v \big(\langle p, v\rangle - L(x,v)\big). \tag{2.4.14}$$

The question is if we can find an optimal control problem for which the MFG system (2.4.7)

$$\begin{aligned} &-\partial_t u - \Delta u + H(x, Du, m) = 0,\\ &\partial_t m - \Delta m - \mathrm{div}(m\ D_p H(x, Du, m)) = 0,\\ &m(0,x) = m_0(x),\ u(T,x) = G(x), \end{aligned} \tag{2.4.15}$$

would be the analog of (2.4.13). Of course, such optimal control problem would be infinite-dimensional: the state space would now be not $\mathbb{T}^d$ or $\mathbb{R}^d$ but a space of functions.

We now describe two such constructions. Let us assume, without loss of generality, that $f(x,0) = 0$. Otherwise we can always subtract $f(x,0)$ from both sides of (2.4.1) and add this term to the Hamiltonian $H(x,p,m)$. Let us also define

$$F(x,m) = \int_0^m f(x,\rho)d\rho, \quad m \ge 0,$$

with $F(x,m) = 0$ for $m < 0$. We also assume that the function $f(x,m)$ is nondecreasing with respect to the second variable, so that $F = F(x,m)$ is convex with respect to m. We have already seen that this assumption leads to uniqueness of a solution.

We denote by $F^*(x,\alpha)$ the Legendre transform of $F(x,m)$ in the m-variable:

$$F^*(x,\alpha) = \sup_{m\ge 0} (\alpha m - F(x,m)) \qquad \forall (x,\alpha) \in \mathbb{T}^d \times \mathbb{R}\ . \tag{2.4.16}$$

Note that $F^*(x,\alpha)$ is convex and nondecreasing with respect to α. Convexity of F^* follows immediately from its definition as a supremum of a family of linear functions in α, while monotonicity is a consequence of the fact that the functions under the supremum in (2.4.16) are increasing in α since $m \geq 0$.

We also introduce the Legendre transform $H^*(x,\xi)$ of $H(x,p)$ with respect to the second variable:

$$H^*(x,\xi) = \sup_{p\in\mathbb{R}^d} \left(\langle \xi, p\rangle - H(x,p)\right).$$

We assume throughout this section that F^* and H^* are smooth enough to perform the computations.

The first optimal control problem we consider is the following: the control parameter is a function $\alpha : \mathbb{T}^d \times [0,T] \to \mathbb{R}$, and the state parameter is the function $u(t,x)$. We fix the functions $m_0(x)$ and $G(x)$ and aim at minimizing the functional

$$\mathcal{J}^{HJ}(\alpha) = \int_0^T \int_{\mathbb{T}^d} F^*\left(x, \alpha(t,x)\right) dxdt - \int_{\mathbb{T}^d} u(0,x) dm_0(x), \tag{2.4.17}$$

over Lipschitz continuous maps $\alpha : \mathbb{T}^d \times (0,T) \to \mathbb{R}^d$. Given a control $\alpha(t,x)$, we find $u(t,x)$ as the unique classical solution to the Hamilton-Jacobi equation with the prescribed terminal condition:

$$\begin{aligned} &-\partial_t u - \Delta u + H(x, Du) = \alpha(t,x), \\ &u(T,x) = G(x). \end{aligned} \tag{2.4.18}$$

This gives the value of $u(0,x)$ and defines the terminal cost term in (2.4.17). Alternatively, the optimal control problem can be rewritten as

$$\inf_u \int_0^T \int_{\mathbb{T}^d} F^*\left(x, -\partial_t u(t,x) - \Delta u(t,x) + H(x, Du(t,x))\right) dxdt - \int_{\mathbb{T}^d} u(0,x) dm_0(x), \tag{2.4.19}$$

under the constraint that the function $u(t,x)$ sufficiently smooth and satisfies the terminal condition $u(T,x) = G(x)$. In other words, this is an infinite-dimensional minimization problem, with the path $\gamma(s) = u(s,\cdot)$ taking values in the space of functions, and the Lagrangian defined as

$$\mathcal{L}(u(s,\cdot), \dot{u}(s,\cdot)) = \int_{\mathbb{T}^d} F^*\left(x, -\dot{u}(t,x) - \Delta u(t,x) + H(x, Du(t,x))\right) dx. \tag{2.4.20}$$

As $H(x,p)$ is convex with respect to the last variable and $F^*(x,\alpha)$ is convex and increasing with respect to the last variable, it is clear that the above Lagrangian is convex in u.

The second optimal control problem is related to the Fokker-Planck equation: the control is now a vector valued function $v : [0,T] \times \mathbb{T}^d \to \mathbb{R}^d$ and the state is the function $m(t,x)$. We minimize the functional

$$\mathcal{J}^{FP}(v) = \int_0^T \int_{\mathbb{T}^d} [m(t,x) H^*\left(x, -v(t,x)\right) + F(x, m(t,x))] dxdt + \int_{\mathbb{T}^d} G(x) m(T,x) dx, \tag{2.4.21}$$

where the trajectory $m(t,\cdot)$ is determiend by the control $v(t,\cdot)$ as the solution to the initial value problem for the Fokker-Planck equation

$$\partial_t m - \Delta m(x,t) + \operatorname{div}(mv) = 0 \text{ in } \mathbb{T}^d \times (0,T), \qquad m(0,x) = m_0(x). \tag{2.4.22}$$

This optimal control problem is also convex, up to a change of variables which appears frequently in optimal transportation theory: let us set $w = mv$. Then the problem can be rewritten as

$$\inf_{(m,w)} \int_0^T \int_{\mathbb{T}^d} \Big[m(x,t)H^*\Big(x, -\frac{w(x,t)}{m(x,t)}\Big) + F(x, m(x,t))\Big]dxdt + \int_{\mathbb{T}^d} G(x)m(T,x)dx, \quad (2.4.23)$$

where the pair (m, w) solves the Fokker-Planck equation

$$\partial_t m - \Delta m(x,t) + \operatorname{div}(w) = 0 \text{ in } \mathbb{T}^d \times (0,T), \qquad m(0) = m_0. \quad (2.4.24)$$

This problem is convex because the constraint (2.4.24) is linear and the map

$$\begin{aligned} H_1(m,w) &:= mH^*\Big(x, -\frac{w}{m}\Big) = m \sup_p \Big(-\frac{1}{m}\langle w, p\rangle - H(x,p)\Big) \\ &= \sup_p \big(-\langle w, p\rangle - mH(x,p)\big) \end{aligned} \quad (2.4.25)$$

is convex in m and w as a supremum of linear functions in m and w.

Here is a very elegant observation.

Theorem 2.4.6 *Assume that the functions $\bar{m}(t,x)$ and $\bar{u}(t,x)$ satisfy the initial and terminal conditions: $\bar{m}(0,x) = m_0(x)$, and $\bar{u}(T,x) = G(x)$, and $\bar{m}, \bar{u} \in C^2(\mathbb{T}^d \times [0,T])$. Suppose also that $\bar{m}(x,t) > 0$ for any $(t,x) \in [0,T] \times \mathbb{T}^d$. Then the following statements are equivalent:*

(i) The pair $(\bar{u}, \bar{m})$ is a solution of the MFG system (2.4.1).

(ii) The control $\bar{\alpha}(t,x) := f(x, \bar{m}(t,x))$ is optimal for $\mathcal{J}^{HJ}(\alpha)$ and the corresponding solution to (2.4.18) is given by $\bar{u}(t,x)$. That is, $\bar{u}(t,x)$ is the minimizer in (2.4.19).

(iii) The control $\bar{v}(t,x) := -D_pH(x, D\bar{u}(t,x))$ is optimal for $\mathcal{J}^{FP}$, and $\bar{m}(t,x)$ is the corresponding solution to (2.4.22).

Proof. The proof is by verification. We will show only the equivalence between (i) and (ii), as the equivalence between (i) and (iii) can be established similarly, by using the reformulation in (2.4.23).

Let us first assume that $(\bar{m}, \bar{u})$ is a solution to the MFG system (2.4.1), and set

$$\bar{\alpha}(t,x) := f(x, \bar{m}(t,x)). \quad (2.4.26)$$

Consider any Lipschitz continuous map $\alpha(t,x)$ and the corresponding solution $u(t,x)$ to (2.4.18). Then, by (2.4.18) and the convexity of $F^*(x,\alpha)$ in α, we have

$$\begin{aligned} \mathcal{J}^{HJ}(\alpha) &= \int_0^T \int_{\mathbb{T}^d} F^*(x, \alpha(t,x))dxdt - \int_{\mathbb{T}^d} u(0,x)dm_0(x) \\ &\geq \int_0^T \int_{\mathbb{T}^d} \big(F^*(x, \bar{\alpha}(t,x)) + \partial_\alpha F^*(x, \bar{\alpha}(t,x))(\alpha(t,x) - \bar{\alpha}(t,x)\big)dxdt - \int_{\mathbb{T}^d} u(0,x)dm_0(x). \end{aligned} \quad (2.4.27)$$

Next, we re-write the right side as

$$\begin{aligned}\mathcal{J}^{HJ}(\alpha) = \mathcal{J}^{HJ}(\bar\alpha) + \int_0^T \int_{\mathbb{T}^d} \partial_\alpha F^*(x,\bar\alpha)\left(-\partial_t(u-\bar u) - \Delta(u-\bar u) + H(x,Du) - H(x,D\bar u)\right)dxdt \\ - \int_{\mathbb{T}^d} (u(0,x) - \bar u(0,x))dm_0(x).\end{aligned} \tag{2.4.28}$$

The next step is to use the convexity of $H(x,p)$ in p to bound the right side above as

$$\begin{aligned}\mathcal{J}^{HJ}(\alpha) \geq \mathcal{J}^{HJ}(\bar\alpha) + \int_0^T \int_{\mathbb{T}^d} \partial_\alpha F^*(x,\bar\alpha)\left(-\partial_t(u-\bar u) - \Delta(u-\bar u) + \langle D_pH(x,D\bar u)), D(u-\bar u)\rangle\right)dxdt \\ - \int_{\mathbb{T}^d} (u-\bar u)(0,x)dm_0(x).\end{aligned} \tag{2.4.29}$$

In order to compute the derivative $\partial_\alpha F^*(x,\bar\alpha)$ that appears in the above expression, let us go back to (2.4.16):

$$F^*(x,\alpha) = \sup_{m\geq 0}\left(\alpha m - F(x,m)\right) \qquad \forall (x,\alpha) \in \mathbb{T}^d \times \mathbb{R}\,. \tag{2.4.30}$$

The optimizer $\bar m(x,\alpha)$ in (2.4.30) is determined by

$$\alpha = F_m(x,\bar m(x,\alpha)) = f(x,\bar m(x,\alpha)), \tag{2.4.31}$$

so that

$$F^*(x,\alpha) = \alpha\bar m(x,\alpha) - F(x,\bar m(x,\alpha)). \tag{2.4.32}$$

Differentiating in α and using (2.4.31), we obtain

$$\partial_\alpha F^*(x,\alpha) = \bar m(x,\alpha) + \alpha\partial_\alpha \bar m(x,\alpha) - F_m(x,\bar m(x,\alpha))\partial_\alpha\bar m(x,\alpha) = \bar m(x,\alpha). \tag{2.4.33}$$

Using this expression in (2.4.29) gives

$$\begin{aligned}\mathcal{J}^{HJ}(\alpha) \geq & \mathcal{J}^{HJ}(\bar\alpha) + \int_0^T \int_{\mathbb{T}^d} \bar m\left(-\partial_t(u-\bar u) - \Delta(u-\bar u) + \langle D_pH(x,D\bar u)), D(u-\bar u)\rangle\right)dxdt \\ & - \int_{\mathbb{T}^d} (u-\bar u)(0,x)dm_0(x).\end{aligned} \tag{2.4.34}$$

Integrating by parts we get

$$\begin{aligned}\mathcal{J}^{HJ}(\alpha) \geq & \mathcal{J}^{HJ}(\bar\alpha) + \int_0^T \int_{\mathbb{T}^d} (u-\bar u)\left(\partial_t\bar m - \Delta\bar m - \operatorname{div}(\bar m D_pH(x,D\bar u))\right)dxdt \\ & + \int_{\mathbb{T}^d} (u(T,x) - \bar u(T,x))m(T,x)dx \geq \mathcal{J}^{HJ}(\bar\alpha).\end{aligned} \tag{2.4.35}$$

The last inequality comes from the Fokker-Planck equation satisfied by $\bar m(t,x)$ and since

$$u(T,x) = \bar u(T,x) = G(x).$$

Thus, we have proved that the control $\bar{\alpha}(t,x) = f(x, \bar{m}(t,x))$ is optimal for the minimization problem for $\mathcal{J}^{HJ}(\alpha)$.

Conversely, let us assume that the control $\bar{\alpha}(t,x)$ is optimal for the functional $\mathcal{J}^{HJ}(\alpha)$, and $\bar{u}(t,x)$ be the corresponding solution to (2.4.18):

$$\begin{aligned} & -\partial_t \bar{u} - \Delta \bar{u} + H(x, D\bar{u}) = \bar{\alpha}(t,x), \qquad \text{in } \mathbb{T}^d \times (0,T), \\ & \bar{u}(x,T) = G(x). \end{aligned} \tag{2.4.36}$$

We also set

$$\bar{m}(t,x) = \partial_\alpha F^*(x, \bar{\alpha}(t,x)), \tag{2.4.37}$$

so that

$$\bar{\alpha}(t,x) := f(x, \bar{m}(t,x)), \tag{2.4.38}$$

as follows from (2.4.31)-(2.4.33). In particular, this shows that $\bar{u}(t,x)$ is a solution to the Hamilton-Jacobi equation (2.4.1)-(i) in the MFG system.

The main issue is to show that the function $\bar{m}(t,x)$ defined by (2.4.37) satisfies the Fokker-Planck equation (2.4.1)-(ii) in the MFG system:

$$\bar{m}_t - \Delta \bar{m} - \operatorname{div}(m D_p H(x, D\bar{u})) = 0, \tag{2.4.39}$$

as well as the initial condition $\bar{m}(0,x) = m_0(x)$. To this end, take a smooth function $a(t,x)$ and, for $h \neq 0$, let $u_h(t,x)$ be the solution to (2.4.18) associated to the control

$$\alpha(t,x) = \bar{\alpha}(t,x) + ha(t,x). \tag{2.4.40}$$

That is, $u_h(t,x)$ solves

$$\begin{aligned} & -\partial_t u - \Delta u + H(x, Du) = \alpha(t,x), \\ & u(T,x) = G(x). \end{aligned} \tag{2.4.41}$$

Then, the difference ratios

$$w_h = \frac{u_h - \bar{u}}{h}, \tag{2.4.42}$$

converge, as $h \to 0$, to some $w(t,x)$ that solves the linearized equation

$$\begin{aligned} & -\partial_t w - \Delta w + \langle D_p H(x, D\bar{u}), Dw \rangle = a(t,x), \\ & w(T,x) = 0. \end{aligned} \tag{2.4.43}$$

Using the optimality of $\bar{\alpha}$ and (2.4.37), we obtain

$$0 = \frac{d\mathcal{J}^{HJ}(\bar{\alpha} + ha)}{h}\Big|_{h=0} = \int_0^T \int_{\mathbb{T}^d} \bar{m} \left(-\partial_t w - \Delta w + \langle D_p H(x, D\bar{u})), Dw \rangle \right) - \int_{\mathbb{T}^d} w(0,x) dm_0(x).$$

We integrate by parts to get, as $w(T,x) = 0$, that

$$0 = \int_0^T \int_{\mathbb{T}^d} w \left(\partial_t \bar{m} - \Delta \bar{m} - \operatorname{div}(\bar{m} D_p H(x, D\bar{u})) \right) - \int_{\mathbb{T}^d} w(0,x)(m_0(x) - \bar{m}(0,x)) dx. \tag{2.4.44}$$

Note that if one fixes $w \in C^3$ such that $w(T,x) = 0$, we can always define $a(t,x)$ by (2.4.43). This implies that relation (2.4.44) holds for any $w \in C^3$ such that $w(T,x) = 0$. Therefore, the function $\bar{m}(t,x)$ is a weak solution of (2.4.1)-(ii) with the initial condition $\bar{m}(0,x) = m_0(x)$. This finishes the proof. □

2.4.4 The long time averages

In this section, we study the long time average of solutions to the MFG system (2.4.1). Let us first recall the results on the long time behavior of the solutions to the initial value for the Hamilton-Jacobi equations of the form

$$\begin{aligned} &\widetilde{u}_t + H(x, \nabla \widetilde{u}) = 0, \\ &\widetilde{u}(0,x) = u_0(x). \end{aligned} \tag{2.4.45}$$

Then, we have shown that there exists $\lambda \in \mathbb{R}$ and a function $\bar{v}(x)$ such that

$$H(x, \bar{v}(x)) + \lambda = 0, \tag{2.4.46}$$

and

$$|\widetilde{u}(t,x) - \lambda t - \bar{v}(x)| \to 0, \quad \text{as } t \to +\infty. \tag{2.4.47}$$

If we rephrase (2.4.45) as a terminal value problem, setting $u(t,x) = \widetilde{u}(T-t,x)$, then (2.4.47) becomes

$$|u(T-t,x) - \lambda t - \bar{v}(x)| \to 0, \quad \text{as } t \to +\infty. \tag{2.4.48}$$

Here, $T > 0$ is fixed and $t \to +\infty$. In the MFG situation, we will be interested in the regime where T is large, while the evaluation time $T - t$ is also large but far away from the terminal time T. A way to reformulate a statement such as (2.4.48), if we are only interested in times $1 \lll t \lll T$, as, again, will be the case for the MFG problem, is to set $t = Ts$ and consider the function

$$v^T(s,x) = u(Ts,x), \tag{2.4.49}$$

with $0 \leq s \leq 1$. Then, (2.4.48) should "morally" become a statement of the form

$$|v^T(1-s,x) - \lambda Ts - \bar{v}(x)| \to 0, \quad \text{as } T \to +\infty, \tag{2.4.50}$$

for $0 < s < 1$ fixed, which is equivalent to

$$|v^T(s,x) - \lambda T(1-s) - \bar{v}(x)| \to 0, \quad \text{as } T \to +\infty, \tag{2.4.51}$$

A weaker version of (2.4.51) is

$$\left| \frac{1}{T} v^T(s,x) - \lambda(1-s) \right| \to 0, \quad \text{as } T \to +\infty. \tag{2.4.52}$$

Our goal in this section is to obtain a similar result to (2.4.52) for the MFG system

$$\begin{aligned} &(i) \; -\partial_t u - \Delta u + H(x, Du) = f(x, m(x,t)), \\ &(ii) \; \partial_t m - \Delta m - \operatorname{div}(mDu) = 0, \\ &(iii) \; m(0,x) = m_0(x), \quad u(T,x) = G(x). \end{aligned} \tag{2.4.53}$$

We concentrate on the simple case $H(x,p) = |p|^2/2$. We also suppose that the coupling $f(x,m)$ is bounded and strictly increasing with respect to the last variable:

$$\frac{\partial f}{\partial m}(x,m) > 0. \tag{2.4.54}$$

As in the previous section, we suppose, without loss of generality, that f is non negative. Moreover, we assume that the initial density $m_0(x)$ is positive and smooth. In particular, there exists $c_0 > 0$ so that

$$m_0(x) \geq c_0 > 0, \quad \text{for all } x \in \mathbb{T}^d. \tag{2.4.55}$$

To emphasize that we are interested in the behavior of the solution as the horizon T tends to $+\infty$, we denote by (u^T, m^T) the solution to

$$\begin{aligned} &(i)\ -\partial_t u - \Delta u + \frac{1}{2}|Du|^2 = f(x, m(x,t)), \\ &(ii)\ \partial_t m - \Delta m - \operatorname{div}(mDu) = 0, \\ &(iii)\ m(0,x) = m_0(x), \quad u(T,x) = G(x). \end{aligned} \tag{2.4.56}$$

This system is still considered on the torus $\mathbb{T}^n$. The analog of the steady problem (2.4.46) for the Hamilton-Jacobi equation, is, in the present situation, the system

$$\begin{aligned} &(i)\quad \bar\lambda - \Delta\bar u + \frac{1}{2}|D\bar u|^2 = f(x,\bar m) \\ &(ii)\quad -\Delta\bar m - \operatorname{div}(\bar m D\bar u) = 0 \\ &(iii)\quad \int_{\mathbb{T}^d} \bar u\, dx = 0\,, \quad \int_{\mathbb{T}^d} \bar m\, dx = 1. \end{aligned} \tag{2.4.57}$$

The unknowns here are $(\bar\lambda, \bar u, \bar m)$, similarly to (2.4.46) where the unknowns are both λ and u. Let us first remark that the above system has a unique solution.

Proposition 2.4.7 *Under the assumptions of this section, the system (2.4.57) on* $\mathbb{T},^n$ *has a unique classical solution* $(\bar\lambda, \bar u, \bar m)$, *and*

$$\bar m(x) = e^{-\bar u(x)} \left(\int_{\mathbb{T}^d} e^{-\bar u(y)} dy \right)^{-1} > 0. \tag{2.4.58}$$

Note that relation (2.4.58) is an immediate consequence of (2.4.57)-(ii) and the normalization for $\bar m(x)$ in (2.4.57)-(iii). Thus, (2.4.57)-(i) is an equation for $\bar u(x)$ and $\bar\lambda$ alone. Their existence essentially follows from the Lions-Papanicolaou-Varadhan theorem.

In order to understand to what extent the solution $(\bar\lambda, \bar u, \bar m)$ to (2.4.57) drives the behavior of (u^T, m^T), let us take $s \in [0,1]$ and consider the scaled functions

$$\upsilon^T(s,x) := u^T(sT,x), \quad \mu^T(s,x) := m^T(sT,x). \tag{2.4.59}$$

Theorem 2.4.8 *We have the convergence*

$$\frac{\upsilon^T(s,x)}{T} \to (1-s)\bar\lambda, \quad \text{in } L^2(\mathbb{T}^d \times (0,1)), \text{ as } T \to +\infty. \tag{2.4.60}$$

Remark 2.4.9 With more estimates than presented here, one can show that $\mu^T(s,x)$ converges to $\bar m(x)$ in $L^p(\mathbb{T}^d \times (0,1))$, for any $p < \frac{d+2}{d}$.

The proof of Theorem 2.4.8 requires several intermediate steps. The starting point is the usual estimate, which is crucial in establishing the uniqueness of the solution to (2.4.56).

Lemma 2.4.10 *For any* $0 \le t_1 < t_2 \le T$ *we have*

$$\int_{\mathbb{T}^d} (u^T - \bar{u})(m^T - \bar{m})dx\Big|_{t_1}^{t_2}$$
$$+ \int_{t_1}^{t_2} \int_{\mathbb{T}^d} \frac{(m^T + \bar{m})}{2} |Du^T - D\bar{u}|^2 + (f(x, m^T) - f(x, \bar{m}))(m^T - \bar{m})\; dxdt = 0$$

Proof. Since T is fixed, we simply write m and u instead of m^T and u^T. We first integrate over $\mathbb{T}^d \times (t_1, t_2)$ the equation satisfied by $(u - \bar{u})$ multiplied by $(m - \bar{m})$. Since

$$\int_{\mathbb{T}^d} (m(t,x) - \bar{m}(x))dx = 0,$$

and $\bar{u}$ does not depend on time, we get, after integration by parts:

$$\begin{aligned}&\int_{t_1}^{t_2} \int_{\mathbb{T}^d} \Big[(-\partial_t u)(m - \bar{m}) + \langle D(m - \bar{m}), D(u - \bar{u})\rangle + \frac{1}{2}(m - \bar{m})(|Du|^2 - |D\bar{u}|^2)\Big] dxdt \\ &= \int_{t_1}^{t_2} \int_{\mathbb{T}^d} (f(x, m) - f(x, \bar{m}))(m - \bar{m})dxdt.\end{aligned} \tag{2.4.61}$$

In the same way, we integrate over $\mathbb{T}^d \times (t_1, t_2)$ the equation satisfied by $(m - \bar{m})$ multiplied by $(u - \bar{u})$:

$$\int_{t_1}^{t_2} \Big(\int_{\mathbb{T}^d} (u - \bar{u})\partial_t m + \langle D(m - \bar{m}), D(u - \bar{u})\rangle + \langle mDu - \bar{m}D\bar{u}, D(u - \bar{u})\rangle \Big) dxdt = 0\,. \tag{2.4.62}$$

We now compute the difference between these two equations:

$$\begin{aligned}&\int_{t_1}^{t_2} \int_{\mathbb{T}^d} \Big[\partial_t[(u - \bar{u})(m - \bar{m})] + \langle mDu - \bar{m}D\bar{u}, D(u - \bar{u})\rangle \\ &- \frac{1}{2}(m - \bar{m})(|Du|^2 - |D\bar{u}|^2) + (f(x, m) - f(x, \bar{m}))(m - \bar{m})\Big] dxdt = 0\,.\end{aligned} \tag{2.4.63}$$

To complete the proof we just note that

$$\langle mDu - \bar{m}D\bar{u}, D(u - \bar{u})\rangle - \frac{1}{2}(m - \bar{m})(|Du|^2 - |D\bar{u}|^2) = \frac{(m + \bar{m})}{2}|Du - D\bar{u}|^2\,.$$

This finishes the proof. □

Another crucial point is given by the following lemma, which exploits the fact that (2.4.56) has a Hamiltonian structure. Note that this is directly related to the optimal control interpretation of the MFG as explained in the previous section.

Lemma 2.4.11 *There exists a constant* M^T *such that for all* $0 \le t \le T$ *we have*

$$\frac{1}{2}\int_{\mathbb{T}^d} m^T(t,x)\,|Du^T(t,x)|^2\,dx + \int_{\mathbb{T}^d} \langle Du^T(t,x), Dm^T(t,x)\rangle\; dx - \int_{\mathbb{T}^d} F(x, m^T(t))\,dx = M^T, \tag{2.4.64}$$

where $F(x,m) = \int_0^m f(x,\rho)\,d\rho$.

Proof. We multiply (2.4.56)-(i) by $\partial_t m^T(t)$ and (2.4.56)-(ii) by $\partial_t u^T(t)$. Summing the two equations we get

$$-(\Delta u^T)\partial_t m^T + \frac{1}{2}|Du^T|^2\,\partial_t m^T - f(x,m^T)\partial_t m^T = (\Delta m^T)\partial_t u^T + \operatorname{div}(m^T Du^T)\partial_t u^T \quad (2.4.65)$$

Integrating with respect to x gives:

$$\int_{\mathbb{T}^d}\left(\langle Du^T,\partial_t Dm^T\rangle + \langle Dm^T,\partial_t Du^T\rangle\right)\,dx + \int_{\mathbb{T}^d}\left[\frac{1}{2}|Du^T|^2\,\partial_t m^T + m^T\,\langle Du^T,\partial_t Du^T\rangle\right]\,dx$$
$$-\int_{\mathbb{T}^d} f(x,m^T)\partial_t m^T\,dx = 0.$$

This means that

$$\frac{d}{dt}\left\{\int_{\mathbb{T}^d}\langle Du^T,Dm^T\rangle\,dx + \frac{1}{2}\int_{\mathbb{T}^d} m^T|Du^T|^2\,dx - \int_{\mathbb{T}^d} F(x,m^T)\,dx\right\} = 0\,,$$

Thus, (2.4.64) holds. □

The next step is to show the following.

Lemma 2.4.12 *There exists a constant $C>0$ that depends on $m_0(x)$ and $G(x)$, so that for all $T>0$ we have*

$$|M^T| + \int_{\mathbb{T}^d}|Du(0,x)|^2dx \le C. \quad (2.4.66)$$

Proof. For an upper bound on M^T, note that we have, since $f \ge 0$ and $u(T)=G(x)$, that

$$\begin{aligned} M^T &= \int_{\mathbb{T}^d}\langle Du(T,x),Dm(T,x)\rangle\,dx + \frac{1}{2}\int_{\mathbb{T}^d} m(T,x)|Du(T,x)|^2\,dx - \int_{\mathbb{T}^d} F(x,m(T))\,dx \\ &\le -\int_{\mathbb{T}^d}(\Delta u(T,x))m(T,x)\,dx + \frac{1}{2}\int_{\mathbb{T}^d} m(T,x)|Du(T,x)|^2\,dx \\ &\le (\|\Delta G\|_{L^\infty} + \|DG\|_{L^\infty}^2)\|m(T)\|_{L^1(\mathbb{T}^d)} = C\,. \end{aligned} \quad (2.4.67)$$

On the other hand, we also have

$$M^T = \int_{\mathbb{T}^d}\langle Du(0,x),Dm_0(x)\rangle\,dx + \frac{1}{2}\int_{\mathbb{T}^d} m_0(x)|Du(0,x)|^2\,dx - \int_{\mathbb{T}^d} F(x,m_0)\,dx. \quad (2.4.68)$$

However, since $m_0>0$, we can write

$$\left|\int_{\mathbb{T}^d}\langle Du(0,x),Dm_0(x)\rangle dx\right| \le \frac{1}{4}\int_{\mathbb{T}^d} m_0(x)|Du(0,x)|^2\,dx + \int_{\mathbb{T}^d}\frac{|Dm_0(x)|^2}{m_0(x)}\,dx\,. \quad (2.4.69)$$

Note that the last integral in the right side above is finite because of the assumption (2.4.55) on the positivity and smoothness of the initial condition $m_0(x)$. We deduce from (2.4.68) and (2.4.69) that

$$M^T \ge \frac{1}{4}\int_{\mathbb{T}^d} m_0(x)|Du(0,x)|^2\,dx - C\,. \quad (2.4.70)$$

It follows from (2.4.67) and (2.4.70) that M^T is bounded both from above and from below. We also deduce from (2.4.70) and the boundedness of M^T that

$$\int_{\mathbb{T}^d} |Du(0,x)|^2\, dx \le C \;, \tag{2.4.71}$$

finishing the proof of (2.4.66). □

Combining Lemma 2.4.12 with Lemma 2.4.10 we get:

Lemma 2.4.13 *There exists $C > 0$ so that*

$$\int_0^T \int_{\mathbb{T}^d} \Big[\frac{(m^T + \bar m)}{2}|Du^T - D\bar u|^2 + (f(x, m^T) - f(x, \bar m))(m^T - \bar m)\Big] dxdt \le C \tag{2.4.72}$$

Proof. Using Lemma 2.4.10, we have

$$\begin{aligned} &\int_0^T \int_{\mathbb{T}^d} \frac{(m^T + \bar m)}{2}|Du^T - D\bar u|^2 + (f(x, m^T) - f(x, \bar m))(m^T - \bar m)\; dxdt \\ &= \int_{\mathbb{T}^d} (u^T(0) - \bar u)(m_0 - \bar m)dx - \int_{\mathbb{T}^d} (u^T(T) - \bar u)(m^T(T) - \bar m)dx. \end{aligned} \tag{2.4.73}$$

Recalling that $u^T(T, x) = G(x)$ and the bounds assumed on G, as well as the mass conservation property of $m^T(t, x)$, we see that the last term in the right side above is bounded. If we set

$$\widetilde{u}^T(t) = \int_{\mathbb{T}^d} u^T(t, x)dx, \tag{2.4.74}$$

we can use the same mass conservation property for $m(t, x)$ to write

$$\begin{aligned} \int_{\mathbb{T}^d} u^T(0, x)(m_0(x) - \bar m(x))dx &= \int_{\mathbb{T}^d} (u^T(0, x) - \widetilde{u}^T(0))(m_0(x) - \bar m(x))dx \\ &\le C\left(\|m_0\|_\infty + \|\bar m\|_\infty\right) \|Du^T(0, \cdot)\|_{L^2(\mathbb{T}^d)}. \end{aligned} \tag{2.4.75}$$

Corollary 2.4.12 implies that this term is bounded. This gives (2.4.72). □

Rewriting Lemma 2.4.13 in terms of υ^T and μ^T we obtain:

Corollary 2.4.14 *The map $D\upsilon^T(s, x)$ converges to $D\bar u(x)$ in $L^2(\mathbb{T}^d \times (0, 1))$ as $T \to +\infty$, while $f(x, \mu^T(s, x))$ converges to $f(x, \bar m(x))$, in $L^1(\mathbb{T}^d \times (0, 1))$ as $T \to +\infty$.*

Proof. Since $\bar m(x)$ and $m_0(x)$ are bounded from below by a positive constant, Lemma 2.4.13 implies that

$$\int_0^1 \int_{\mathbb{T}^d} |D\upsilon^T(s, x) - D\bar u(x)|^2\; dxds \le \frac{C}{T}.$$

This implies the convergence of $D\upsilon^T$.

To prove the convergence of $f(x, \mu^T(s, x))$, note that due to the strict monotonicity of $f(x, m)$ in m, see assumption (2.4.54), there exists $\delta > 0$ such that

$$\frac{\partial f}{\partial m}(x, m) \ge \delta, \quad \text{for all } (x, m) \in \mathbb{T}^d \times [0, 2\|\bar m\|_\infty].$$

Thus, it follows from (2.4.72) and uniform bounds on both $m^T(t,x)$ and $\bar{m}(x)$ that

$$\begin{aligned}\frac{C}{T} \geq &\int_0^1 \int_{\mathbb{T}^d} (f(x,\mu^T(s,x)) - f(x,\bar{m}(x)))(\mu^T(s,x) - \bar{m}(x))\, dxds \\ \geq & C \iint_{\{\mu^T(s,x)\geq 2\|\bar{m}\|_\infty\}} \left|f(x,\mu^T(s,x)) - f(x,\bar{m}(x))\right| dxds \\ &+ \delta \iint_{\{\mu^T(s,x)< 2\|\bar{m}\|_\infty\}} \left|\mu^T(s,x) - \bar{m}(x)\right| dxds.\end{aligned} \tag{2.4.76}$$

Therefore, we have

$$\begin{aligned}&\|f(\cdot,\mu^T) - f(\cdot,\bar{m})\|_{L^1} \leq \iint_{\{\mu^T(s,x)\geq 2\|\bar{m}\|_\infty\}} \left|f(x,\mu^T(s,x)) - f(x,\bar{m}(x))\right| dxds \\ &+ \iint_{\{\mu^T(s,x)< 2\|\bar{m}\|_\infty\}} \left|f(x,\mu^T(s,x)) - f(x,\bar{m}(x))\right| dxds \\ &\leq \iint_{\{\mu^T(s,x)\geq 2\|\bar{m}\|_\infty\}} \left|f(x,\mu^T(s,x)) - f(x,\bar{m}(x))\right| dxds \\ &+ \sup_{0\leq m\leq 2\|\bar{m}\|_\infty} \left|\frac{\partial f}{\partial m}\right| \iint_{\{\mu^T(s,x)< 2\|\bar{m}\|_\infty\}} \left|\mu^T(s,x) - \bar{m}(x)\right| dxds \leq \frac{C}{T}\left(1+\frac{1}{\delta}\right),\end{aligned} \tag{2.4.77}$$

which implies the convergence of $f(x,\mu^T(s,x))$ to $f(x,\bar{m}(x))$ in L^1. □

Proof of Theorem 2.4.8. We now prove the convergence of $\upsilon^T(sT,x)/T$ to $\bar{\lambda}(1-s)$. Let us integrate the equation satisfied by υ^T on $\mathbb{T}^d \times (t,1)$:

$$\frac{1}{T}\left(\int_{\mathbb{T}^d} \upsilon^T(s,x)dx - \int_{\mathbb{T}^d} G(x)dx\right) + \frac{1}{2}\int_s^1 \int_{\mathbb{T}^d} |D\upsilon^T(\tau,x)|^2 dxd\tau = \int_s^1 \int_{\mathbb{T}^d} f(x,\mu^T(\tau,x))dxd\tau. \tag{2.4.78}$$

We know from Corollary 2.4.14 that $D\upsilon^T(s,x) \to D\bar{u}(x)$ in L^2 and $f(x,\mu^T(s,x)) \to f(x,\bar{m}(x))$ in L^1. We deduce that

$$\lim_{T\to+\infty} \frac{1}{T}\int_{\mathbb{T}^d} \upsilon^T(s,x)dx = (1-s)\int_{\mathbb{T}^d}\left[-\frac{1}{2}|D\bar{u}|^2 + f(x,\bar{m})\right] dx = (1-s)\bar{\lambda}\,, \tag{2.4.79}$$

the last equality being obtained by integrating over $\mathbb{T}^d$ equation (2.4.57)-(i). Let us set

$$\langle \upsilon^T\rangle(s) = \int_{\mathbb{T}^d} \upsilon^T(s,x)dx, \quad \widetilde{\upsilon}^T(s,x) = \upsilon^T(s,x) - \langle \upsilon^T\rangle(s).$$

Using the Poincaré inequality, we get

$$\int_0^1 \int_{\mathbb{T}^d} \left|\widetilde{\upsilon}^T(s,x) - \bar{u}(x)\right|^2 \leq C\int_0^1 \int_{\mathbb{T}^d} |D\upsilon^T(s,x) - D\bar{u}(x))|^2 \to 0. \tag{2.4.80}$$

This shows the convergence of $\widetilde{\upsilon}^T(s,x)$ to $\bar{u}(x)$ in L^2, as $T\to+\infty$. In addition, (2.4.79) implies the convergence

$$\frac{\langle \upsilon^T\rangle(s)}{T} \to (1-s)\bar{\lambda}. \tag{2.4.81}$$

Together, these imply the convergence in L^2 of $\upsilon^T(s,x)/T$ to $(1-s)\bar{\lambda}$. □

Exercise 2.4.15 Recall that we have already seen a similar situation in the construction in the proof of the Lions-Papanicolaou-Varadhan theorem: the solution becomes large but its gradient stays bounded. Make this connection more precise.

2.4.5 Comments

Other existence results of classical solutions of second order MFG systems with local coupling can be found in Cardaliaguet, Lasry, Lions and Porretta [180] (for quadratic Hamiltonian, without conditions on the coupling f) and for more general Hamiltonians under various structure conditions on the coupling in a series of papers by Gomes, Pires and Sanchez-Morgado [206, 209, 211] and Gomes and Pimentel [213]. The case of MFG system with congestion is considered in Gomes and Mitake [214].

Even for some data, it is not known if there always exists a classical solution to the MFG system. To overcome this issue, concepts of weak solutions have been introduced in Lasry and Lions [243] and in Porretta [257].

The general uniqueness criterium given in Theorem 2.4.2 has been introduced by Lions [248], who explains the sharpness of the condition.

The fact that the MFG system with local coupling possesses a variational structure is pointed out in Lasry and Lions in [243]. This plays a key role for the first order MFG system with local coupling, since this allows to build solutions in that setting.

Finally the long time behavior of the MFG system is described in section 2.4.4 has been first discussed by Lions in [248] and sharpened in Cardaliaguet, Lasry, Lions and Porretta [181]. Other results in that direction can be found in Gomes, Mohr and Suza [203] (for discrete MFG systems), Cardaliaguet, Lasry, Lions and Porretta [180] (for MFG system with a nonlocal coupling), Cardaliaguet [177] (for the first order MFG with a nonlocal coupling) and in Cardaliaguet and Graber [176] (for the first order MFG with local coupling). For second order MFG systems, the rate of this convergence is exponential (see [180, 181]).

2.5 The space of probability measures

We have already seen the important role of the space of probability measures in the mean field game theory. It is now time to investigate the basic properties of this space more thoghroughly.

The first two parts of this section are dedicated to metric aspects of probability measures spaces. The results are given mostly without proofs, which can be found, for instance, in Villani's monographs [263, 264] or in the monograph by Ambrosio, Gigli and Savar [152].

2.5.1 The Monge-Kantorovich distances

Let X be a Polish space (i.e., separable metric space) and $\mathcal{P}(X)$ be the set of Borel probability measures on X. A sequence of measures μ_n is narrowly convergent to a measure $\mu \in \mathcal{P}(X)$ if

$$\lim_{n\to+\infty} \int_X f(x)d\mu_n(x) = \int_X f(x)d\mu(x) \qquad \forall f \in \mathcal{C}^0_b(X) \ ,$$

where $\mathcal{C}^0_b(X)$ is the set of continuous, bounded maps on X. The Prokhorov Theorem states that a subset $\mathcal{K}$ of $\mathcal{P}(X)$ is relatively compact in $\mathcal{P}(X)$ if and only if it is tight: for all $\varepsilon > 0$ there exists a compact set $X_\varepsilon \subset X$ such that $\mu(X\backslash X_\varepsilon) \leq \varepsilon$ for all $\mu \in \mathcal{K}$. In particular, for any $\mu \in \mathcal{P}(X)$ and any $\varepsilon > 0$, there is a compact subset X_ε of X with $\mu(X\backslash X_\varepsilon) \leq \varepsilon$ (Ulam's Lemma).

There are several ways to metrize the topology of narrow convergence, at least on some subsets of $\mathcal{P}(X)$. Let us denote by d the distance on X and, for $p \in [1, +\infty)$, by $\mathcal{P}_p(X)$ the set of probability measures m such that

$$\int_X d^p(x_0, x)dm(x) < +\infty \qquad \text{for some (and hence for all) point } x_0 \in X.$$

The Monge-Kantorowich distance on $\mathcal{P}_p(X)$ is given by

$$d_p(m, m') = \inf_{\gamma \in \Pi(m,m')} \left[\int_{X^2} d(x, y)^p d\gamma(x, y) \right]^{1/p} \tag{2.5.1}$$

where $\Pi(\mu, \nu)$ is the set of Borel probability measures on X such that $\gamma(A \times X) = \mu(A)$ and $\gamma(X \times A) = \nu(A)$ for any Borel set $A \subset X$. In other words, a Borel probability measure γ on $X \times X$ belongs to $\Pi(m, m')$ if and only if

$$\int_{X^2} \varphi(x)d\gamma(x, y) = \int_X \varphi(x)dm(x) \qquad \text{and} \qquad \int_{X^2} \varphi(y)d\gamma(x, y) = \int_X \varphi(y)dm'(y) ,$$

for any Borel and bounded measurable map $\varphi : X \to \mathbb{R}$. Note that $\Pi(\mu, \nu)$ is non-empty, because for instance $\mu \otimes \nu$ always belongs to $\Pi(\mu, \nu)$. Moreover, by the Hölder inequality, $\mathcal{P}_p(X) \subset \mathcal{P}_{p'}(X)$ for any $1 \leq p' \leq p$ and

$$d_p(m, m') \leq d_{p'}(m, m') \qquad \forall m, m' \in \mathcal{P}_p(X) .$$

We now explain that there exists at least an optimal measure in (2.5.1). This optimal measure is often refered to as *an optimal transport plan* from m to m'.

Lemma 2.5.1 (Existence of an optimal transport plan) *For any $m, m' \in \mathcal{P}_p(X)$, there is at least one measure $\bar{\gamma} \in \Pi(m, m')$ with*

$$d_p(m, m') = \left[\int_{X^2} d(x, y)^p d\bar{\gamma}(x, y) \right]^{1/p} .$$

Proof. We first show that $\Pi(\mu, \nu)$ is tight. For any $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset X$ such that $\mu(K_\varepsilon) \geq 1 - \varepsilon/2$ and $\nu(K_\varepsilon) \geq 1 - \varepsilon/2$. Then, for any $\gamma \in \Pi(\mu, \nu)$, we have

$$\gamma(K_\varepsilon \times K_\varepsilon) \geq \gamma(K_\varepsilon \times X) - \gamma(K_\varepsilon \times (X \backslash K_\varepsilon)) \geq \mu(K_\varepsilon) - \gamma(X \times (X \backslash K_\varepsilon)) \geq 1 - \varepsilon/2 - \nu(X \backslash K_\varepsilon) \geq 1 - \varepsilon.$$

This means that $\Pi(\mu, \nu)$ is tight. It is also closed for the weak-* convergence. Since the map

$$\gamma \to \int_{X^2} d(x, y)^p d\gamma(x, y)$$

is lower semi-continuous for the weak-* convergence, it has a minimum on $\Pi(m, m')$. $\square$

Let us now check that d_p is a distance.

Lemma 2.5.2 *For any $p \geq 1$, d_p is a distance on $\mathcal{P}_p$.*

Proof. Only the triangle inequality presents some difficulty. Let $m, m', m'' \in \mathcal{P}_p$ and γ, γ' be optimal transport plans from m to m' and from m' to m'' respectively. We desintegrate the measures γ and γ' with respect to m': $d\gamma(x,y) = d\gamma_y(x)dm'(y)$ and $d\gamma'(y,z) = d\gamma'_y(z)dm'(y)$ and we defined the measure π on $X \times X$ by

$$\int_{X\times X} \varphi(x,z)d\pi(x,z) = \int_{X\times X\times X} \varphi(x,z)d\gamma_y(x)d\gamma'_y(z)dm'(y) \ .$$

Then one easily checks that $\pi \in \Pi(m, m'')$ and we have, by Hlder inequality,

$$\begin{aligned}\left[\int_{X\times X} d^p(x,z)d\pi(x,z)\right]^{1/p} &\leq \left[\int_{X\times X\times X} (d(x,y)+d(y,z))^p d\gamma_y(x)d\gamma'_y(z)dm'(y)\right]^{1/p} \\ &\leq \left[\int_{X\times X} d^p(x,y)d\gamma_y(x)dm'(y)\right]^{1/p} + \left[\int_{X\times X} d^p(y,z)d\gamma_y(z)dm'(y)\right]^{1/p} \\ &= \mathbf{d}_p(m,m') + \mathbf{d}_p(m',m'')\end{aligned}$$

So $\mathbf{d}_p(m, m'') \leq \mathbf{d}_p(m, m') + \mathbf{d}_p(m', m'')$. ∎

We now prove that the distance $\mathbf{d}_p$ metricize the weak-* convergence of measures.

Proposition 2.5.3 *If a sequence of measures (m_n) of $\mathcal{P}_p(X)$ converges to m for $\mathbf{d}_p$, then (m_n) weakly converges to m.*

"Conversely", if the (m_n) are concentrated on a fixed compact subset of X and weakly converge to m, then the (m_n) converge to m in $\mathbf{d}_p$.

Remark 2.5.4 The sharpest statement can be found in [263]: a sequence of measures (m_n) of $\mathcal{P}_p(X)$ converges to m for $\mathbf{d}_p$ if and only if (m_n) weakly converges to m and

$$\lim_{n\to+\infty} \int_X d^p(x,x_0)dm_n(x) = \int_X d^p(x,x_0)dm(x) \qquad \text{for some (and thus any) } x_0 \in X \ .$$

Proof. In a first step, we only show now that, if (m_n) converges to m for $\mathbf{d}_p$, then

$$\lim_{n\to+\infty} \int_X \varphi(x)dm_n(x) = \int_X \varphi(x)dm(x) \tag{2.5.2}$$

for any $\varphi \in \mathcal{C}^0_b(X)$. The proof of the converse statement is explained after Theorem 2.5.5.

We first prove that (2.5.2) holds for Lipschitz continuous maps: indeed, if φ is Lipschitz continuous for some Lipschitz constant L, then, for any optimal transport plan $\gamma_n \in \Pi(m_n, m)$ from m_n to m, we have

$$\begin{aligned}\left|\int_X \varphi(x)dm_n(x) - \int_X \varphi(x)dm(x)\right| &= \left|\int_X (\varphi(x)-\varphi(y))d\gamma_n(x)\right| \\ &\leq L\int_X d(x,y)d\gamma_n(x) \ \leq \ L\mathbf{d}_p(m_n,m) \ .\end{aligned}$$

So (2.5.2) holds for any Lipschitz continuous φ.

If now $\varphi \in \mathcal{C}^0_b(X)$, we approximate φ by the Lipschitz continuous map

$$\varphi^\varepsilon(x) = \inf_{y\in X} \left\{ \varphi(y) - \frac{1}{\varepsilon} d(x,y) \right\} \qquad \forall x \in X \ .$$

Then it is an easy exercise to show that $\varphi^\varepsilon(x) \to \varphi(x)$ as $\varepsilon \to 0$. Moreover φ^ε is $(1/\varepsilon)-$Lipschitz continuous, bounded by $\|\varphi\|_\infty$ and satisfies $\varphi^\varepsilon \geq \varphi$. In particular, from Lebesgue Theorem,

$$\lim_{\varepsilon\to 0} \int_X \varphi^\varepsilon(x) dm(x) = \int_X \varphi(x) dm(x) \ .$$

Applying (2.5.2) to the Lipschitz continuous map φ^ε we have

$$\limsup_{n\to+\infty} \int_X \varphi(x) dm_n(x) \leq \limsup_{n\to+\infty} \int_X \varphi^\varepsilon(x) dm_n(x) = \int_X \varphi^\varepsilon(x) dm(x) \ .$$

Then, letting $\varepsilon \to 0$, we get

$$\limsup_{n\to+\infty} \int_X \varphi(x) dm_n(x) \leq \int_X \varphi(x) dm(x) \ .$$

Applying the above inequality to $-\varphi$ also gives

$$\liminf_{n\to+\infty} \int_X \varphi(x) dm_n(x) \geq \int_X \varphi(x) dm(x) \ .$$

So (2.5.2) holds for any $\varphi \in \mathcal{C}^0_b(X)$. ⊐

In these notes, we are mainly interested in two Monge-Kantorovich distances, $\mathbf{d}_1$ and $\mathbf{d}_2$. The distance $\mathbf{d}_2$, which is often called the Wasserstein distance, is particularly usefull when X is a Euclidean or a Hilbert space. Its analysis will be the object of the next subsection.

As for the distance $\mathbf{d}_1$, which often takes the name of the Kantorovich-Rubinstein distance, we have already encountered it several times. Let us point out a very important equivalent representation:

Theorem 2.5.5 (Kantorovich-Rubinstein Theorem) *For any $m, m' \in \mathcal{P}_1(X)$,*

$$\mathbf{d}_1(m, m') = \sup \left\{ \int_X f(x) dm(x) - \int_X f(x) dm'(x) \right\}$$

where the supremum is taken over the set of all $1-$Lipschitz continuous maps $f : X \to \mathbb{R}$.

Remark 2.5.6 In fact the above "Kantorovich duality result" holds for much more general costs (i.e., it is not necessary to minimize the power of a distance). The typical assertion in this framework is, for any lower semicontinuous map $c : X \times X \to \mathbb{R}_+ \cup \{+\infty\}$, the following equality holds:

$$\inf_{\gamma\in\Pi(m,m')} \int_{X\times X} c(x,y) d\gamma(x,y) = \sup_{f,g} \int_X f(x) dm(x) + \int_X g(y) dm'(y) \ ,$$

where the supremum is taken over the maps $f \in L^1_m(X)$, $g \in L^1_{m'}(X)$ such that

$$f(x) + g(y) \leq c(x,y) \qquad \text{for } dm-\text{almost all } x \text{ and } dm'-\text{almost all } y.$$

Proof. [Ideas of proof of Theorem 2.5.5.] The complete proof of this result exceeds the scope of these note and can be found in several text books (see [263] for instance). First note that, if f is $1-$Lipschitz continuous, then

$$f(x) - f(y) \le d(x, y) \qquad (x, y) \in X \times X \ .$$

Integrating this inequality over any measure $\gamma \in \Pi(m, m')$ gives

$$\int_X f(x)dm(x) - \int_X f(y)dm'(y) \le \int_{X\times X} d(x, y)d\gamma(x, y) \ ,$$

so that, taking the infimum over γ and the supremum of f gives

$$\sup \left\{ \int_X f(x)dm(x) - \int_X f(x)dm'(x) \right\} \le \mathbf{d}_1(m, m') \ .$$

The opposite inequality is much more subtle. We now assume that X is compact and denote by $\mathcal{M}_+(X^2)$ the set of all nonnegative Borel measures on $X \times X$. We first note that, for any $\gamma \in \mathcal{M}_+(X^2)$

$$\sup_{f,g\in\mathcal{C}^0(X)} \int_X f(x)dm(x) + \int_X g(y)dm'(y) - \int_{X\times X} (f(x) + g(y))d\gamma(x, y) = \begin{cases} 0 & \text{if } \gamma \in \Pi(m, m') \\ +\infty & \text{otherwise} \end{cases}$$

So

$$\mathbf{d}_1(m, m') = \inf_{\gamma\in\mathcal{M}(X^2)} \sup_{f,g\in\mathcal{C}^0(X)} \int_{X\times X} (d(x, y) - f(x) - g(y))d\gamma(x, y) + \int_X f(x)dm(x) + \int_X g(y)dm'(y)$$

If we could use the min-max Theorem, then we would have

$$\mathbf{d}_1(m, m') = \sup_{f,g\in\mathcal{C}^0(X)} \inf_{\gamma\in\mathcal{M}(X^2)} \int_{X\times X} (d(x, y) - f(x) - g(y))d\gamma(x, y) + \int_X f(x)dm(x) + \int_X g(y)dm'(y)$$

where

$$\inf_{\gamma\in\mathcal{M}(X^2)} \int_{X\times X} (d(x, y) - f(x) - g(y))d\gamma(x, y) = \begin{cases} 0 & \text{if } f(x) + g(y) \le d(x, y) \ \forall x, y \in X \\ -\infty & \text{otherwise} \end{cases}$$

So

$$\mathbf{d}_1(m, m') = \sup_{f,g} \int_X f(x)dm(x) + \int_X g(y)dm'(y)$$

where the supremum is taken over the maps $f, g \in \mathcal{C}^0(X)$ such that $f(x) + g(y) \le d(x, y)$ holds for any $x, y \in X$. Let us fix $f, g \in \mathcal{C}^0(X)$ satisfying this inequality and set $\widetilde{f}(x) = \min_{y\in X}[d(x, y) - g(y)]$ for any $x \in X$. Then, by definition, $\widetilde{f}$ is $1-$Lipschitz continuous, $\widetilde{f} \ge f$ and $\widetilde{f}(x) + g(y) \le d(x, y)$. So

$$\int_X f(x)dm(x) + \int_X g(y)dm'(y) \le \int_X \widetilde{f}(x)dm(x) + \int_X g(y)dm'(y) \ .$$

We can play the same game by replacing g by $\widetilde{g}(y) = \min_{x\in X} d(x,y) - \widetilde{f}(x)$, which is also $1-$Lipschitz continuous and satisfies $\widetilde{g} \geq g$ and $\widetilde{f}(x) + \widetilde{g}(y) \leq d(x,y)$. But one easily checks that $\widetilde{g}(y) = -\widetilde{f}(y)$. So

$$\int_X f(x)dm(x) + \int_X g(y)dm'(y) \leq \int_X \widetilde{f}(x)dm(x) - \int_X \widetilde{f}(y)dm'(y) \ .$$

Hence

$$\mathbf{d}_1(m,m') \leq \sup_{\widetilde{f}} \int_X \widetilde{f}(x)dm(x) - \int_X \widetilde{f}(y)dm'(y)$$

where the supremum is taken over the $1-$Lipschitz continuous maps $\widetilde{f}$. This completes the formal proof of the result. ◻

Proof. [End of the proof of Proposition 2.5.3.] It remains to show that, if the (m_n) are concentrated on a fixed compact subset K of X and weakly converge to m, then the (m_n) converge to m in $\mathbf{d}_p$. Note that $m(K) = 1$, so that m is also concentrated on K.

We now show that it is enough to do the proof for $p = 1$: indeed, if $\gamma \in \Pi(m_n, m)$, then $\gamma(K \times K) = 1$ because m_n and m are concentrated on K. Therefore

$$\int_{X\times X} d^p(x,y)d\gamma(x,y) = \int_{K\times K} d^p(x,y)d\gamma(x,y) \leq [\text{diam}(K)]^{p-1} \int_{K\times K} d(x,y)d\gamma(x,y)$$

where $\text{diam}(K)$ denotes the diameter of K, i.e., $\text{diam}(K) = \max_{x,y\in K} d(x,y)$, which is bounded since K is compact. Setting $C = [\text{diam}(K)]^{(p-1)/p}$, we get

$$\mathbf{d}_p(m_n,m) \leq \inf_{\gamma\in\Pi(m_n,m)} C \left[\int_{K\times K} d(x,y)d\gamma(x,y)\right]^{1/p} \leq C[\mathbf{d}_1(m_n,m)]^{1/p}$$

and it is clearly enough to show that the right-hand side has a limit.

In order to prove that (m_n) converge to m in $\mathbf{d}_1$, we use Theorem 2.5.5 which implies that we just have to show that

$$\lim_{n\to+\infty} \sup_{\text{Lip}(f)\leq 1} \int_K f(x)d(m_n - m)(x) = 0 \ .$$

Note that can can take the supremum over the set of maps f such that $f(x_0) = 0$ (for some fixed point $x_0 \in K$). Now, by Ascoli Theorem, the set F of maps f such that $f(x_0) = 0$ and $\text{Lip}(f) \leq 1$ is compact. In particular, for any n, there is some $f_n \in F$ such that $\mathbf{d}_1(m_n,m) = \int_K f_n(x)d(m_n - m)(x)$. Let $f \in F$ be a limit of a subsequence of the (f_n) (still denoted (f_n)). Then, by uniform convergence of (f_n) to f and weak convergence of (m_n) to m, we have

$$\limsup_n \mathbf{d}_1(m_n,m) = \limsup_n \int_K f_n(x)d(m_n - m)(x) = 0 \ ,$$

which proves that, for any converging subsequence of the precompact family (f_n) there is a subsequence of the $(\mathbf{d}_1(m_n,m))$ which converges to 0. This implies that the full sequence $(\mathbf{d}_1(m_n,m))$ converges to 0. ◻

In the case where $X = \mathbb{R}^d$, we repeatedly use the following compactness criterium:

Lemma 2.5.7 *Let $r \geq p > 0$ and $\mathcal{K} \subset \mathcal{P}_p$ be such that*

$$\sup_{\mu \in \mathcal{K}} \int_{\mathbb{R}^d} |x|^r d\mu(x) < +\infty \ .$$

Then the set $\mathcal{K}$ is tight. If moreover $r > p$, then $\mathcal{K}$ is relatively compact for the $\mathbf{d}_p$ distance.

Note carefully that bounded subsets of $\mathcal{P}_p$ are not relatively compact for the $\mathbf{d}_k$ distance. For instance, in dimension $d = 1$ and for $p = 2$, the sequence of measures $\mu_n = \frac{n-1}{n}\delta_0 + \frac{1}{n}\delta_{n^2}$ satisfies $\mathbf{d}_2(\mu_n, \delta_0) = 1$ for any $n \geq 1$ but μ_n narrowly converges to δ_0.

Proof. [Proof of Lemma 2.5.7.] Let $\varepsilon > 0$ and $R > 0$ sufficiently large. We have for any $\mu \in \mathcal{K}$:

$$\mu(\mathbb{R}^d \backslash B_R(0)) \leq \int_{\mathbb{R}^d \backslash B_R(0)} \frac{|x|^r}{R^r} d\mu(x) \leq \frac{C}{R^r} < \varepsilon \ ,$$

where $C = \sup_{\mu \in \mathcal{K}} \int_{\mathbb{R}^d} |x|^r d\mu(x) < +\infty$. So $\mathcal{K}$ is tight.

Let now (μ_n) be a sequence in $\mathcal{K}$. From the previous step we know that (μ_n) is tight and therefore there is a subsequence, again denoted (μ_n), which narrowly converges to some μ. Let us prove that the convergence holds for the distance $\mathbf{d}_p$. Let $R > 0$ be large and let us set $\mu_n^R := \Pi_{B_R(0)} \sharp \mu_n$ and $\mu^R := \Pi_{B_R(0)} \sharp \mu$, where $\Pi_{B_R(0)}$ denotes the projection onto $B_R(0)$. Note that

$$\begin{aligned} \mathbf{d}_p^p(\mu_n^R, \mu_n) \ &\leq \ \int_{\mathbb{R}^d} |\Pi_{B_R(0)}(x) - x|^p d\mu_n(x) \leq \int_{(B_R(0))^c} |x|^p d\mu_n(x) \\ &\leq \ \frac{1}{R^{r-p}} \int_{(B_R(0))^c} |x|^r d\mu_n(x) \leq \frac{C}{R^{r-p}} \ . \end{aligned}$$

In the same way, ${}^{p}_{p}(\mu^R, \mu) \leq \frac{C}{R^{r-p}}$. Let us fix $\varepsilon > 0$ and let us choose R such that $\frac{C}{R^{r-p}} \leq (\varepsilon/3)^p$. Since the μ_n^R have a support in the compact set $B_R(0)$ and weakly converge to μ^R, Proposition 2.5.3 states that the sequence (μ_n^R) converges to μ^R for the distance $\mathbf{d}_p$. So we can choose n_0 large enough such that $\mathbf{d}_p(\mu_n^R, \mu^R) \leq \varepsilon/3$ for $n \geq n_0$. Then

$$\mathbf{d}_p(\mu^n, \mu) \leq \mathbf{d}_p(\mu_n^R, \mu_n) + \mathbf{d}_p(\mu_n^R, \mu^R) + \mathbf{d}_p(\mu^R, \mu) \leq \varepsilon \qquad \forall n \geq n_0 \ .$$

□

2.5.2 The Wasserstein space of probability measures on $\mathbb{R}^d$

From now on we work in $X = \mathbb{R}^d$. Let $\mathcal{P}_2 = \mathcal{P}_2(\mathbb{R}^d)$ be the set of Borel probability measures on $\mathbb{R}^d$ with a second ordre moment: m belongs to $\mathcal{P}_2$ if m is a Borel probability on $\mathbb{R}^d$ with $\int_{\mathbb{R}^d} |x|^2 m(dx) < +\infty$. The Wasserstein distance is just the Monge-Kankorovich distance when $p = 2$:

$$\mathbf{d}_2(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \left[\int_{\mathbb{R}^{2d}} |x - y|^2 d\gamma(x, y) \right]^{1/2} \tag{2.5.3}$$

where $\Pi(\mu, \nu)$ is the set of Borel probability measures on $\mathbb{R}^{2d}$ such that $\gamma(A \times \mathbb{R}^d) = \mu(A)$ and $\gamma(\mathbb{R}^d \times A) = \nu(A)$ for any Borel set $A \subset \mathbb{R}^d$.

An important point, that we shall use sometimes, is the fact that the optimal transport plan can be realized as *an optimal transport map* whenever μ is absolutely continuous.

Theorem 2.5.8 (Existence of an optimal transport map) *If $\mu \in \mathcal{P}_2$ is absolutely continuous, then, for any $\nu \in \mathcal{P}_2$, there exists a convex map $\Phi : \mathbb{R}^N \to \mathbb{R}$ such that the measure $(id_{\mathbb{R}^d}, D\Phi)\sharp\mu$ is optimal for $\mathbf{d}_2(\mu,\nu)$. In particular $\nu = D\Phi\sharp\mu$.*

Conversely, if the convex map $\Phi : \mathbb{R}^N \to \mathbb{R}$ satisfies $\nu = D\Phi\sharp\mu$, then the measure $(id_{\mathbb{R}^d}, D\Phi)\sharp\mu$ is optimal for $\mathbf{d}_2(\mu,\nu)$.

The proof of this result, due to Y. Brenier [166], exceeds the scope of these notes. It can be found in various places, such as [263].

2.5.3 Polynomials on $\mathcal{P}(Q)$

Let Q be a compact metric space and let us denote as usual by $\mathcal{P}(Q)$ the set of probability measures on Q. We say that a map $P \in \mathcal{C}^0(\mathcal{P}(Q))$ is a monomial of degree k if there are k real-valued continuous maps $\phi_i : Q \to \mathbb{R}$ $(i = 1, \dots, k)$ such that

$$P(m) = \prod_{i=1}^{k} \int_Q \phi_i(x)dm(x) \qquad \forall m \in \mathcal{P}(Q) \; .$$

If Q is a compact subset of $\mathbb{R}^d$, it is usual convenient to also assume that the maps ϕ_i are $\mathcal{C}^\infty$.

Note that the product of two monomials is still a monomial. Hence the set of polynomials, i.e., the set of finite linear combinations of monomials, is subalgebra of $\mathcal{C}^0(\mathcal{P}(Q))$. It contains the unity: $P(m) = 1$ for all $m \in \mathcal{P}(Q)$ (choose $\phi = 1$). It also separates points: indeed, if $m_1, m_2 \in \mathcal{P}(Q)$ are distinct, then there is some smooth map $\phi : \mathbb{R}^d \to \mathbb{R}$ with compact support such that $\int_Q \phi(x)dm_1(x) \neq \int_Q \phi(x)dm_2(x)$. Then the monomial $P(m) = \int_Q \phi(x)dm(x)$ separates m_1 and m_2. Using Stone-Weierstrass Theorem we have proved the following:

Proposition 2.5.9 *The set of polynomials is dense in $\mathcal{C}^0(\mathcal{P}(Q))$.*

2.5.4 Hewitt and Savage Theorem

We now investigate here the asymptotic behavior of symmetric measures of a large number of variables. Let us fix a compact probability metric space. We say that a measure μ on Q^k (where $k \in \mathbb{N}^*$) is symmetric if, for any permutation σ on $\{1, \dots, k\}$, $\pi_\sigma\sharp\mu = \mu$, where $\pi_\sigma(x_1, \dots, x_k) = (x_{\sigma(1)}, \dots, x_{\sigma(k)})$.

For any $k \geq 1$, Let m^k be a symmetric measure on Q^k and let us set, for any $n < k$,

$$m^k_n = \int_{Q^{n-k}} dm^k(x_{n+1}, \dots, x_n) \; .$$

Then, from a diagonal argument, we can find a subsequence $k' \to +\infty$ such that $(m^{k'}_n)$ has a limit m_n as $k' \to +\infty$ for any $n \geq 0$. Note that the m_n are still symmetric and satisfies $\int_Q dm_{n+1}(x_{n+1}) = m_n$ for any $n \geq 1$. Hewitt and Savage describes the structure of such sequence of measures.

Theorem 2.5.10 (Hewitt and Savage) *Let (m_n) be a sequence of symmetric probability measures on Q^n such that $\int_Q dm_{n+1}(x_{n+1}) = m_n$ for any $n \geq 1$. Then there is a probability measure μ on $\mathcal{P}(Q)$ such that, for any continuous map $f \in \mathcal{C}^0(\mathcal{P}(Q))$,*

$$\lim_{n\to+\infty} \int_{Q^n} f\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right) dm_n(x_1,\dots,x_n) = \int_{\mathcal{P}(Q)} f(m)d\mu(m)\ . \tag{2.5.4}$$

Moreover

$$m_n(A_1\times\cdots\times A_n) = \int_{\mathcal{P}(Q)} m(A_1)\dots m(A_n)d\mu(m) \tag{2.5.5}$$

for any $n \in \mathbb{N}^*$ and any Borel sets $A_1,\dots,A_n \subset Q$.

Remark 2.5.11 An important case is when the measure $m_n = \otimes_{i=1}^n m_0$, where $m_0 \in \mathcal{P}(Q)$. Then, because of (2.5.5), the limit measure has to be δ_{m_0}. In particular, for any continuous map $f \in \mathcal{C}^0(\mathcal{P}(Q))$, (2.5.4) becomes

$$\lim_{n\to+\infty} \int_{Q^n} f\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right) dm_n(x_1,\dots,x_n) = f(m_0)\ .$$

In particular, if $\mathbf{d}_1$ is the Kantorovich-Rubinstein distance on $\mathcal{P}(Q)$, then

$$\lim_{n\to+\infty} \int_{Q^n} \mathbf{d}_1\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, m_0\right) dm_n(x_1,\dots,x_n) \ = \ 0\ .$$

Remark 2.5.12 (Probabilistic interpretation of the Hewitt and Savage Theorem) The above result is strongly related with De Finetti's Theorem (see for instance [234]). Let $(\Omega, \mathcal{A}, \P)$ be a probability space and (X_k) a sequence of random variables with values in Q. The sequence (X_k) is said to be exchangeable if for all $n \in \mathbb{N}^*$, the law of $(X_{\sigma(1)},\dots,X_{\sigma(n)})$ is the same as the law of $(X_1,\dots,X_n)$ for any permutation σ of $\{1,\dots,n\}$. For instance, if the (X_n) are iid, then the sequence is exchangeable.

De Finetti's Theorem states that there is a $\sigma-$algebra $\mathcal{F}_\infty$ conditional to which the (X_i) are iid: namely

$$\P\left[X_1 \in A_1,\dots,X_n \in A_n \mid \mathcal{F}_\infty\right] = \prod_{i=1}^n \P\left[X_i \in A_i \mid \mathcal{F}_\infty\right]$$

for any $n \in \mathbb{N}^*$ and any Borel sets $A_1,\dots,A_n \subset Q$.

Proof. [Proof of Theorem 2.5.10.] For any $n \geq 1$ let us define the linear functional $L_n \in (\mathcal{C}^0(\mathcal{P}(Q)))^*$ by

$$L_n(P) = \int_{Q^n} P(\frac{1}{n}\sum_{i=1}^n \delta_{y_i})m_n(dy_1,\dots,dy_n) \qquad \forall P \in \mathcal{C}^0(\mathcal{P}(Q))\ .$$

We want to show that L_n has a limit as $n \to +\infty$. Since the L_n are obviously uniformly bounded, it is enough to show that $L_n(P)$ has a limit for any map P of the form

$$P(m) = \int_{Q^j} \phi(x_1, \dots, x_j) dm(x_1) \dots dm(x_j) \tag{2.5.6}$$

where $\phi : Q^j \to \mathbb{R}$ is continuous, because such class of functions contain the monomials defined in subsection 2.5.3, and the set of resulting polynomials is dense in $\mathcal{C}^0(\mathcal{P}(Q))$. Note that, for any $n \geq j$ and any $y_1, \dots, y_n \in Q$,

$$P(\frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}) = \frac{1}{n^j} \sum_{(i_1,\dots,i_j)} \phi(y_{i_1}, \dots, y_{i_j})$$

where the sum is taken over the $(i_1, \dots, i_j) \in \{1, \dots, n\}^j$. So

$$L_n(P) = \frac{1}{n^j} \sum_{i_1,\dots,i_j} \int_{Q^n} \phi(y_{i_1}, \dots, y_{i_j}) m_n(dy_1, \dots, dy_n)$$

Since m_n is symmetric and satisfies $\int_{Q^{n-j}} dm_n(x_{j+1}, \dots, x_n) = m_j$, if $i_1, \dots, i_j$ are distinct we have

$$\int_{Q^n} \phi(y_{i_1}, \dots, y_{i_j}) m_n(dy_1, \dots, dy_n) = \int_{Q^j} \phi(y_1, \dots, y_j) dm_j(x_1, \dots, x_j) \ .$$

On another hand

$$\sharp\{(i_1, \dots, i_j) \ , \ i_1, \dots, i_j \text{ distinct}\} = \frac{n!}{(n-j)!} \sim_{n\to+\infty} n^j \ ,$$

so that

$$\lim_{n\to+\infty} L_n(P) = \int_{Q^j} \phi(y_1, \dots, y_j) dm_j(x_1, \dots, x_j) \ .$$

This prove the existence of a limit L of L_n as $n \to +\infty$. Note that $L \in (\mathcal{C}^0(\mathcal{P}(Q)))^*$, that L is nonegative and that $L(1) = 1$. By Riesz representation Theorem there is a unique Borel measure μ on $\mathcal{P}(Q)$ such that $L(P) = \int_{\mathcal{P}(Q)} P(m) d\mu(m)$.

It remains to show that the measure μ satisfies relation (2.5.5). Let P be again defined by (2.5.6). We have already proved that

$$L(P) = \int_{Q_j} \phi(y_1, \dots, y_j) dm_j(x_1, \dots, x_j) = \int_{\mathcal{P}(Q)} P(m)\mu(dm)$$

where

$$\int_{\mathcal{P}(Q)} P(m)\mu(dm) = \int_{\mathcal{P}(Q)} \left(\int_{Q^j} \phi(x_1, \dots, x_j) dm(x_1) \dots dm(x_j) \right) \mu(dm)$$

Let now $A_1, \dots, A_j$ be closed subsets of Q. We can find a nonincreasing sequence (ϕ_k) of continuous functions on $\mathbb{R}^j$ which converges to $\mathbf{1}_{A_1}(x_1) \dots \mathbf{1}_{A_j}(x_j)$. This gives (2.5.5) for any closed subsets $A_1, \dots, A_j$ of Q, and therefore for any Borel measurable subset of $A_1, \dots, A_j$ of Q. $\square$

The fact that we are working on a compact set plays little role and this assumption can be removed, as we show in a particular case.

Corollary 2.5.13 *Let m_0 be probability measure on a Polish space X with a first order moment (i.e., $m_0 \in \mathcal{P}_1(X)$) and let $m_n = \otimes_{i=1}^n m_0$ be the law on X^N of n iid random variables with law m_0. Then, for any Lipschitz continuous map $f \in \mathcal{C}^0(\mathcal{P}_1(X))$,*

$$\lim_{n\to+\infty} \int_{X^n} f\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right) dm_n(x_1,\dots,x_n) = f(m_0)\ .$$

Proof. For $\varepsilon > 0$ let K_ε be a compact subset of X such that $\mu_0(K_\varepsilon) \geq 1-\varepsilon$. We also choose K_ε in such a way that, for some fixed $\bar{x} \in X$, $\int_{X\setminus K_\varepsilon} d(x,\bar{x})dm_0(x) \leq \varepsilon$. Without loss of generality we can suppose that $\bar{x} \in K_\varepsilon$. Let us now denote by π the map defined by $\pi(x) = x$ if $x \in K_\varepsilon$, $\pi(x) = \bar{x}$ otherwise, and set $m_\varepsilon = \pi\sharp m_0$ and $m_n^\varepsilon = \otimes_{i=1}^n m_\varepsilon$. Note that by definition $m_n^\varepsilon = (\pi,\dots,\pi)\sharp m_n$. Since m_ε is concentrated on a compact set, we have, from Theorem 2.5.10,

$$\lim_{n\to+\infty} \int_{X^n} f\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right) dm_n^\varepsilon(x_1,\dots,x_n) = f(m_\varepsilon)\ .$$

On the other hand, using the Lipschitz continuity of f, one has for any n:

$$\begin{aligned}
\left|\int_{X^n} f\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right) d(m_n^\varepsilon - m_n)\right| &\leq \int_{X^n}\left|f\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right) - f\left(\frac{1}{n}\sum_{i=1}^n \delta_{\pi(x_i)}\right)\right| dm_n \\
&\leq Lip(f)\int_{X^n} \mathbf{d}_1\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \frac{1}{n}\sum_{i=1}^n \delta_{\pi(x_i)}\right)\ dm_n \\
&\leq Lip(f)\int_{X\setminus K_\varepsilon} d(x,\bar{x})\ dm_0(x)\ \leq\ Lip(f)\varepsilon
\end{aligned}$$

In the same way,

$$|f(m_0) - f(m_\varepsilon)| \leq Lip(f)\mathbf{d}_1(m_0, m_\varepsilon) \leq Lip(f)\varepsilon\ .$$

Combining the above inequalities easy gives the result. ◻

Another consequence of the Hewitt and Savage Theorem is:

Theorem 2.5.14 *Let Q be compact and $u_n : Q^n \to \mathbb{R}$ be symmetric and converge to $U : \mathcal{P}(Q) \to \mathbb{R}$ in the sense of Theorem 2.2.8:*

$$\lim_{n\to+\infty} \sup_{X\in Q^n} |u_n(X) - U(m_X^n)| = 0$$

and (m_n) be a sequence of symmetric probability measures on Q^n such that $\int_Q dm_{n+1}(x_{n+1}) = m_n$ for all n and μ be the associate probability measure on $\mathcal{P}(Q)$ as in the Hewitt and Savage Theorem. Then

$$\lim_{n\to+\infty} \int_{Q^n} u_n(x_1,\dots,x_n)dm_n(x_1,\dots,x_n) = \int_{\mathcal{P}(Q)} U(m)d\mu(m)\ .$$

Proof. From the convergence of u_n to U we have

$$\lim_{n\to+\infty} \left|\int_{Q^n} u_n(x_1,\dots,x_n)dm_n(x_1,\dots,x_n) - \int_{Q^n} U(m_{x_1,\dots,x_n}^n)dm_n(x_1,\dots,x_n)\right| = 0\ ,$$

while, since U is continuous, Hewitt and Savage Theorem states that

$$\lim_{n\to+\infty}\int_{Q^n} U(m^n_{x_1,\ldots,x_n})dm_n(x_1,\ldots,x_n) = \int_{\mathcal{P}(Q)} U(m)d\mu(m)\ .$$

Combining these two relations gives the result. □

2.5.5 Comments

The study of optimal transport and Monge-Kantorovitch distances is probably one of the most dynamic areas in analysis in these last two decades. The applications of this analysis are numerous, from probability theory to P.D.Es and from to geometry. The first two subsections of this part rely on Villani's monographs [263, 264] or in the monograph by Ambrosio, Gigli and Savar [152]. The definition of polynomials on $\mathcal{P}(Q)$ comes from [248], as well as the proof of the Hewitt and Savage Theorem (see also the original reference by Hewitt and Savage [223] and Kingman [234] for a survey on exchangeability).

2.6 Derivatives in the space of measures

We are now interested in the analysis of Hamilton-Jacobi equations in the space of measures. As we shall see later, such equations provide the right framework for the study of limits of large systems of Hamilton-Jacobi equations in finite dimensional spaces. The first part of this section is devoted to the notion of derivative in the Wasserstein space.

2.6.1 Derivatives in the $L^2(\mathbb{R}^d)$ sense

Definition 2.6.1 *We say that $U : \mathcal{P}_1 \to \mathbb{R}^k$ is C^1 if there exists a continuous map*

$$\frac{\delta U}{\delta m} : \mathcal{P}_1 \times \mathbb{T}^d \to \mathbb{R}^k,$$

such that, for any $m, m' \in \mathcal{P}_1$,

$$U(m') - U(m) = \int_0^1 \int_{\mathbb{T}^d} \frac{\delta U}{\delta m}((1-s)m + sm', y)\ d(m' - m)(y)$$

Note that $\delta U/\delta m$ is defined up to an additive constant. To fix the idea, we always assume that

$$\int_{\mathbb{T}^d} \frac{\delta U}{\delta m}(m, y) dy = 0.$$

If $\delta U/\delta m$ is Lipschitz continuous with respect to the second variable, with a Lipschitz constant bounded independently of m, then U is Lipschitz continuous and

$$\begin{aligned} |U(m') - U(m)| &\leq d_1(m, m') \int_0^1 \left\| \frac{\delta U}{\delta m}((1-s)m + sm', \cdot) \right\|_{Lip} ds \\ &\leq d_1(m, m') \sup_{m''} \left\| \frac{\delta U}{\delta m}(m'', \cdot) \right\|_{Lip}. \end{aligned}$$

This leads us to define, if $\dfrac{\delta U}{\delta m}$ is of class C^1 with respect to the second variable with a C^1 norm bounded independently of m, $D_m U : \mathcal{P}_1 \times \mathbb{T}^d \to \mathbb{R}^d$ by

$$D_m U(m, y) := D_y \frac{\delta U}{\delta m}(m, y)$$

Given a C^1 map $U : \mathcal{P}_1 \to \mathbb{R}$, let us define

$$\mathrm{Lip}_n(\frac{\delta U}{\delta m}) := \sup_{m_1 \neq m_2} (d_1(m_1, m_2))^{-1} \left\| \frac{\delta U}{\delta m}(\cdot, m_1, \cdot) - \frac{\delta U}{\delta m}(\cdot, m_2, \cdot) \right\|_{C^{n+2\alpha} \times C^{n-1+2\alpha}}$$

The following Lemma explains that $D_m U$ is also a kind of derivative of U. We state the result in a quantified way, since it is used under that form in the sequel.

Proposition 2.6.2 *Assume that $U : \mathbb{T}^d \times \mathcal{P}_1 \to \mathbb{R}$ is C^1 and that*

$$\left\| \frac{\delta U}{\delta m}(\cdot, m, \cdot) \right\|_{C^{n+2\alpha} \times C^{n-1+2\alpha}} + \mathrm{Lip}_n\left(\frac{\delta U}{\delta m}\right) \leq C_n.$$

Fix $m \in \mathcal{P}_1$ and let $\phi \in L^2(m, \mathbb{R}^d)$ be a vector field. Then

$$\left\| U(\cdot, (id+\phi)\sharp m) - U(\cdot, m) - \int_{\mathbb{T}^d} \langle D_m U(\cdot, m, y), \phi(y)\rangle dm(y) \right\|_{C^{n+2\alpha}} \leq (C_n + 1)\|\phi\|^2_{L^2(m)}$$

Proof. Let us first check that, for any $m, m' \in \mathcal{P}_1$, we have

$$\left\| U(\cdot, m') - U(\cdot, m) - \int_{\mathbb{T}^d} \frac{\delta U}{\partial m}(\cdot, m, y) d(m'-m)(y) \right\|_{C^{n+2\alpha}} \leq C_n \mathbf{d}_1^2(m, m')$$

Indeed, for any $l \in \mathbb{N}^d$ with $|l| \leq n$ and any $x, x' \in \mathbb{T}^d$, we have

$$\begin{aligned}
&\Big| D_x^l U(x, m') - D_x^l U(x, m) - \int_{\mathbb{T}^d} D_x^l \frac{\delta U}{\partial m}(x, m, y) d(m'-m)(y)) \\
&\qquad -(D_x^l U(x', m') - D_x^l U(x', m) - \int_{\mathbb{T}^d} D_x^l \frac{\delta U}{\partial m}(x', m, y) d(m'-m)(y)) \Big| \\
&\quad \leq \int_0^1 \Big| \int_{\mathbb{T}^d} \Big(D_x^l \frac{\delta U}{\delta m}(x, (1-s)m + sm', y) - D_x^l \frac{\delta U}{\partial m}(x, m, y) \\
&\qquad - (D_x^l \frac{\delta U}{\delta m}(x', (1-s)m + sm', y) - D_x^l \frac{\delta U}{\partial m}(x', m, y)) \Big) d(m'-m)(y) \Big| ds \\
&\quad \leq \sup_{s,y} \Big| D_y D_x^l \frac{\delta U}{\delta m}(x, (1-s)m + sm', y) - D_y D_x^l \frac{\delta U}{\partial m}(x, m, y) \\
&\qquad - (D_y D_x^l \frac{\delta U}{\delta m}(x', (1-s)m + sm', y) - D_y D_x^l \frac{\delta U}{\partial m}(x', m, y)) \Big| \mathbf{d}_1(m, m') \\
&\quad \leq \mathrm{Lip}_n \left(\frac{\delta U}{\delta m} \right) |x - x'|^{2\alpha} \mathbf{d}_1^2(m, m')
\end{aligned}$$

This proves our claim.

Using this claim we obtain

$$\begin{aligned}
&\left\| U(\cdot, (id+\phi)\sharp m) - U(\cdot, m) - \int_{\mathbb{T}^d} \frac{\delta U}{\partial m}(\cdot, m, y) d((id+\phi)\sharp m - m)(y) \right\|_{C^{n+2\alpha}} \\
&\qquad \leq C_n \mathbf{d}_1^2(m, (id+\phi)\sharp m).
\end{aligned}$$

Using once more the regularity of U, we obtain (omitting the dependence with respect to (x, m) for simplicity):

$$\begin{aligned}
&\left| \int_{\mathbb{T}^d} \frac{\delta U}{\partial m}(y) d(id+\phi)\sharp m(y) - \int_{\mathbb{T}^d} \frac{\delta U}{\partial m}(y) dm(y) - \int_{\mathbb{T}^d} \langle D_m U(y), \phi(y)\rangle dm(y) \right| \\
&\quad = \left| \int_{\mathbb{T}^d} \frac{\delta U}{\partial m}(y + \phi(y)) dm(y) - \int_{\mathbb{T}^d} \frac{\delta U}{\partial m}(y) dm(y) - \int_{\mathbb{T}^d} \langle D_m U(y), \phi(y)\rangle dm(y) \right| \\
&\quad \leq \left| \int_0^1 \int_{\mathbb{T}^d} \langle D_y \frac{\delta U}{\partial m}(y + s\phi(y)), \phi(y)\rangle dm(y) ds - \int_{\mathbb{T}^d} \langle D_m U(y), \phi(y)\rangle dm(y) \right| \\
&\quad \leq C_n \|\phi\|^2_{L^2(m)}
\end{aligned}$$

As

$$\mathbf{d}_1^2(m, (id+\phi)\sharp m) \leq \|\phi\|^2_{L^1(m)},$$

the result is proved. □

2.7 The Master equation

In this section we investigate the well-posedness of the (first order) master equation:

$$\begin{cases} -\partial_t U - \Delta_x U + H(x, D_x U) - \displaystyle\int_{\mathbb{T}^d} \mathrm{div}_y D_m U(t,x,m,y)\, dm(y) \\ \qquad + \displaystyle\int_{\mathbb{T}^d} \langle D_m U(t,x,m,y), D_p H(x, D_x U)\rangle dm = F(x,m) \\ \qquad\qquad \text{in } (0,T)\times \mathbb{T}^d \times \mathcal{P}_2 \\ U(T,x,m) = G(x,m) \qquad \text{in } \mathbb{T}^d \times \mathcal{P}_2 \end{cases} \tag{2.7.1}$$

2.7.1 Wellposedness of the equation

Definition 2.7.1 *We say that a map $V : [0,T]\times\mathbb{T}^d\times\mathcal{P}_2 \to \mathbb{R}$ is a classical solution to the Master equation if*

- *V is continuous in all its arguments (for the $\mathbf{d}_1$ distance on $\mathcal{P}_2$), is of class C^2 in x and C^1 in time,*
- *V is of class C^1 with respect to m with a derivative $\frac{\delta V}{\delta m} = \frac{\delta V}{\delta m}(t,x,m,y)$ with globally continuous first and second order derivatives with respect to the space variables.*
- *The following relation holds for any $(t,x,m) \in (0,T)\times\mathbb{T}^d\times\mathcal{P}_2$,*

$$\begin{cases} -\partial_t V(t,x,m) - \Delta_x V(t,x,m) + H(x, D_x V(t,x,m)) - \displaystyle\int_{\mathbb{T}^d} \mathrm{div}_y D_m V(t,x,m,y)\, dm(y) \\ \qquad + \displaystyle\int_{\mathbb{T}^d} \langle D_m V(t,x,m,y), D_p H(x, D_x V(t,x,m))\rangle dm = F(x,m) \end{cases} \tag{2.7.2}$$

and $V(T,x,m) = G(x,m)$ in $\mathbb{T}^d\times\mathcal{P}_2$.

Throughout the section, $H : \mathbb{T}^d \times \mathbb{R}^d \to \mathbb{R}$ is smooth, globally Lipschitz continuous and satisfies the coercivity condition:

$$C^{-1}\frac{I_d}{1+|p|} \le D^2_{pp}H(x,p) \le CI_d \qquad \text{for } (x,p)\in\mathbb{T}^d\times\mathbb{R}^d. \tag{2.7.3}$$

We also always assume that the maps $F, G : \mathbb{T}^d \times \mathcal{P}_1 \to \mathbb{R}$ are globally Lipschitz continuous and monotone:

$$F \text{ and } G \text{ are monotone.} \tag{2.7.4}$$

Note that assumption (2.7.4) implies that $\frac{\delta F}{\delta m}$ and $\frac{\delta G}{\delta m}$ satisfy the following monotonicity property (explained for F):

$$\int_{\mathbb{T}^d}\int_{\mathbb{T}^d} \frac{\delta F}{\delta m}(x,m,y)\mu(x)\mu(y)dxdy \ge 0$$

for any smooth map $\mu : \mathbb{T}^d \to \mathbb{R}$ with $\int_{\mathbb{T}^d}\mu = 0$.

Let us fix $n \in \mathbb{N}^*$ and $\alpha \in (0,1/2)$. We set

$$\mathrm{Lip}_n(\frac{\delta F}{\delta m}) := \sup_{m_1\neq m_2} (\mathbf{d}_1(m_1,m_2))^{-1} \left\| \frac{\delta F}{\delta m}(\cdot, m_1,\cdot) - \frac{\delta F}{\delta m}(\cdot, m_2, \cdot)\right\|_{C^{n+2\alpha}\times C^{n-1+2\alpha}}$$

and use the symmetric notation for G. We call **(HF(n))** the following regularity conditions on F:

$$(\mathbf{HF}(\mathbf{n})) \qquad \sup_{m\in\mathcal{P}_1}\left(\|F(\cdot,m)\|_{C^{n+2\alpha}}+\left\|\frac{\delta F(\cdot,m,\cdot)}{\delta m}\right\|_{C^{n+2\alpha}\times C^{n+2\alpha}}\right)+\mathrm{Lip}_n(\frac{\delta F}{\delta m}) \;<\; \infty.$$

and **(HG(n))** the symmetric condition on G:

$$(\mathbf{HG}(\mathbf{n})) \qquad \sup_{m\in\mathcal{P}_1}\left(\|G(\cdot,m)\|_{C^{n+2\alpha}}+\left\|\frac{\delta G(\cdot,m,\cdot)}{\delta m}\right\|_{C^{n+2\alpha}\times C^{n+2\alpha}}\right)+\mathrm{Lip}_n(\frac{\delta G}{\delta m}) \;<\; \infty.$$

In order to explain the existence of a solution to the master equation, we need to introduce the solution of the MFG system: for any $(t_0,m_0)\in[0,T)\times\mathcal{P}_2$, let (u,m) be the solution to:

$$\left\{\begin{array}{l} -\partial_t u-\Delta u+H(x,Du)=F(x,m(t))\\ \partial_t m-\Delta m-\mathrm{div}(mD_pH(x,Du))=0\\ u(T,x)=G(x,m(T)),\;\; m(t_0,\cdot)=m_0 \end{array}\right. \tag{2.7.5}$$

Thanks to the monotony condition (2.7.4), we know that the system admits a unique solution. We set

$$U(t_0,x,m_0):=m(t_0,x) \tag{2.7.6}$$

Theorem 2.7.2 *Assume that* **(HF(n))** *and* **(HG(n))** *hold for some $n\geq 4$. Then the map U defined by (2.7.6) is the unique classical solution to the master equation (2.7.2). Moreover, U is globally Lipschitz continuous in the sense that*

$$\|U(t_0,\cdot,m_0)-U(t_0,\cdot,m_1)\|_{C^{n+\alpha}}\leq C_n\mathbf{d}_1(m_0,m_1) \tag{2.7.7}$$

with Lipschitz continuous derivatives:

$$\|D_mU(t_0,\cdot,m_0,\cdot)-D_mU(t_0,\cdot,m_0,\cdot)\|_{C^{n+\alpha}\times C^{n+\alpha}}\leq C_n\mathbf{d}_1(m_0,m_1) \tag{2.7.8}$$

for any $t_0\in[0,T]$, $m_0,m_1\in\mathcal{P}_1$.

The main point in the proof of Theorem 2.7.2 is to check that the map U defined by (2.7.6) satisfies (2.7.7), (2.7.8). This exceeds the scope of these notes and we refer the reader to [**?**]. Once we know that U is quite smooth, the conclusion follows easily:

Proof. [Proof of Theorem 2.7.2 (existence).] Let $m_0\in C^1$, $m_0>0$. Let $t_0>0$, (u,m) be the solution of the MFG system (2.7.5) starting from m_0 at time t_0. Then

$$\frac{U(t_0+h,x,m_0)-U(t_0,x,m_0)}{h} = \frac{U(t_0+h,x,m_0)-U(t_0+h,x,m(t_0+h))}{h} +\frac{U(t_0+h,x,m(t_0+h))-U(t_0,x,m_0)}{h}.$$

As

$$\partial_t m-\mathrm{div}[m(D(\ln(m))+D_pH(x,Du))]=0,$$

Lemma 2.7.3 below says that

$$\mathbf{d}_1(m(t_0+h),(id-h\Phi)\sharp m_0)=o(h)$$

where

$$\Phi(x) := D(\ln(m_0(x))) + D_pH(x, Du(t_0, x))$$

and $o(h)/h \to 0$ as $h \to 0$. So, by Lipschitz continuity of U and then differentiability of U,

$$\begin{aligned} U(t_0+h, x, m(t_0+h)) &= U(t_0+h, x, (id - h\Phi)\sharp m_0) + o(h) \\ &= U(t_0+h, x, m_0) - h\int_{\mathbb{T}^d} \langle D_mU(t_0+h, x, m_0, y), \Phi(u)\rangle m_0(y)dy + o(h), \end{aligned}$$

and therefore, by continuity of U and D_mU,

$$\lim_{h\to 0} \frac{U(t_0+h, x, m(t_0+h)) - U(t_0+h, x, m_0)}{h}$$
$$= -\int_{\mathbb{T}^d} \left(\langle D_mU(t_0, x, m_0)(y), D(\ln(m_0)) + D_pH(x, Du(t_0))\rangle\right) m_0(y)dy.$$

On the other hand, for $h > 0$,

$$U(t_0+h, x, m(t_0+h)) - U(t_0, x, m_0) = u(t_0+h, x) - u(t_0, x) = h\partial_t u(t_0, x) + o(h),$$

so that

$$\lim_{h\to 0^+} \frac{U(t_0+h, x, m(t_0+h)) - U(t_0, x, m_0)}{h} = \partial_t u(t_0, x).$$

Therefore $\partial_t U(t_0, x, m_0)$ exists and is equal to

$$\begin{aligned} \partial_t U(t_0, x, m_0) &= \int_{\mathbb{T}^d} (\langle D_mU(t_0, x, m_0, y), D(\ln(m_0)) + D_pH(x, Du(t_0))\rangle) m_0(y)dy + \partial_t u(t_0, x) \\ &= -\int_{\mathbb{T}^d} \mathrm{div}_y D_mU(t_0, x, m_0, y)m_0(y)dy \\ &\qquad + \int_{\mathbb{T}^d} \langle D_mU(t_0, x, m_0, y), D_pH(x, Du(t_0))\rangle m_0(y)dy \\ &\qquad -\Delta u(t_0, x) + H(x, Du(t_0, x)) - F(x, m_0) \\ &= -\int_{\mathbb{T}^d} \mathrm{div}_y D_mU(t_0, x, m_0, y)m_0(y)dy \\ &\qquad + \int_{\mathbb{T}^d} \langle D_mU(t_0, x, m_0, y), D_pH(x, D_xU(t_0, y, m_0))\rangle m_0(y)dy \\ &\qquad -\Delta_{xx}U(t_0, x, m_0) + H(x, D_xU(t_0, x, m_0)) - F(x, m_0) \end{aligned}$$

This means that U satisfies (2.7.2) at (t_0, x, m_0). By continuity, U satisfies the equation everywhere. ⊓⊔

Lemma 2.7.3 *Let $V = V(t, x)$ be a C^1 vector field, $m_0 \in \mathcal{P}_2$ and m be the weak solution to*

$$\left\{ \begin{array}{l} \partial_t m + \mathrm{div}(mV) = 0 \\ m(0) = m_0 \end{array} \right.$$

Then

$$\lim_{h\to 0^+} \mathbf{d}_1(m(h), (id + hV(0, \cdot))\sharp m_0)/h = 0.$$

Proof. Recall that $m(h) = X^{\cdot}(h)\sharp m_0$, where $X^x(h)$ is the solution to the ODE

$$\left\{ \begin{array}{l} \frac{d}{dt}X^x(t) = V(t, X^x(t)) \\ X^x(0) = x \end{array} \right.$$

Let ϕ be a Lipschitz test function. Then

$$\begin{array}{rcl} \int_{\mathbb{T}^d} \phi(m(h) - (id + hV(0,\cdot))\sharp m_0) & = & \int_{\mathbb{T}^d} (\phi(X^x(h)) - \phi(x + hV(0,x)))m_0 dx \\ & \leq & \|D\phi\|_\infty \int_{\mathbb{T}^d} |X^x(h) - x - hV(0,x)| dm_0(x) = \|D\phi\|_\infty o(h), \end{array}$$

which proves that $\mathbf{d}_1(m(h), (id + hV(0,\cdot))\sharp m_0) = o(h)$. ☐

Proof. [Proof of Theorem 2.7.2 (uniqueness).] We use a technique introduced by in [248], consisting at looking at the MFG system (2.7.5) as a system of characteristics for the master equation (2.7.2). We reproduce it here for sake of completeness. Let V be another solution to the master equation. The main point is that, by definition of solution $D^2_{xy}\frac{\delta V}{\delta m}$ is bounded, and therefore $D_x V$ is Lipschitz continuous with respect to the measure variable.

Let us fix (t_0, m_0). In view of the Lipschitz continuity of $D_x V$, one can easily uniquely solve the PDE (by standard fixed point argument):

$$\left\{ \begin{array}{l} \partial_t \widetilde{m} - \Delta \widetilde{m} - \mathrm{div}(\widetilde{m} D_p H(x, D_x V(t, x, \widetilde{m})) = 0 \\ \widetilde{m}(t_0) = m_0 \end{array} \right.$$

Then let us set $\widetilde{u}(t,x) = V(t, x, \widetilde{m}(t))$. By the regularity properties of V, $\widetilde{u}$ is at least of class $C^{2,1}$ with

$$\begin{array}{rcl} \partial_t \widetilde{u}(t,x) & = & \partial_t V(t, x, \widetilde{m}(t)) + \langle \frac{\delta V}{\delta m}(t, x, \widetilde{m}(t), \cdot), \partial_t \widetilde{m}(t) \rangle_{C^2, (C^2)'} \\ & = & \partial_t V(t, x, \widetilde{m}(t)) + \langle \frac{\delta V}{\delta m}(t, x, \widetilde{m}(t), \cdot), \Delta \widetilde{m} + \mathrm{div}(\widetilde{m} D_p H(x, D_x V(t, x, \widetilde{m})) \rangle_{C^2, (C^2)'} \\ & = & \partial_t V(t, x, \widetilde{m}(t)) + \int_{\mathbb{T}^d} \mathrm{div}_y D_m V(t, x, \widetilde{m}(t), y) \; d\widetilde{m}(t)(y) \\ & & - \int_{\mathbb{T}^d} \langle D_m V(t, x, \widetilde{m}(t), y), D_p H(x, D_x V(t, x, \widetilde{m})) \rangle \; d\widetilde{m}(t)(y) \end{array}$$

Recalling that V satisfies the master equation:

$$\begin{array}{rcl} \partial_t \widetilde{u}(t,x) & = & -\Delta_x V(t, x, \widetilde{m}(t)) + H(x, D_x V(t, x, \widetilde{m}(t))) - F(x, \widetilde{m}(t)) \\ & = & -\Delta \widetilde{u}(t,x) + H(x, D\widetilde{u}(t,x)) - F(x, \widetilde{m}(t)) \end{array}$$

with terminal condition $\widetilde{u}(T,x) = V(T, x, \widetilde{m}(T)) = G(x, \widetilde{m}(T))$. Therefore the pair $(\widetilde{u}, \widetilde{m})$ is a solution of the MFG system (2.7.5). As the solution of this system is unique, we get that $V(t_0, x, m_0) = U(t_0, x, m_0)$ is uniquely defined. ☐

2.7.2 Comment

Most formal properties of the Master equation have been introduced and presented by Lions in [248]. Nothing on the actual existence of solutions was really known beforeIn [169], where a master equation is studied without the coupling term ($F = G = 0$). [202] analyzes the first order master equation in short time. [191] obtains the existence and uniqueness for the master equation under convexity conditions on H with respect to the space variable. The result presented in these notes is a very simplified version of [?], where the existence and uniqueness of solutions for the master equation with common noise is established under the monotonicity condition on the coupling terms.

2.8 Convergence of the Nash system

In this section, we consider a classical *symmetrical* solution $(v^{N,i})$ of the Nash system with a common noise:

$$\begin{cases} -\partial_t v^{N,i} - \sum_j \Delta_{x_j} v^{N,i} + H(x_i, D_{x_i} v^{N,i}) \\ \qquad\qquad + \sum_{j\neq i} \langle D_p H(x_j, D_{x_j} v^{N,j}), D_{x_j} v^{N,i}\rangle = F(x_i, m_X^{N,i}) \qquad \text{in } (0,T)\times \mathbb{T}^{Nd} \\ v^{N,i}(T,x) = G(x_i, m_X^{N,i}) \qquad \text{in } \mathbb{T}^{Nd} \end{cases} \tag{2.8.1}$$

where we set, for $X = (x_1, \ldots, x_N) \in (\mathbb{T}^d)^N$, $m_X^{N,i} = \dfrac{1}{N-1}\sum_{j\neq i} \delta_{x_j}$. Our aim is to show that the solution $(v^{N,i})$ converges, in a suitable sense, to the solution of the master equation with a common noise.

Throughout this part we denote by $U = U(t,x,m)$ the solution of the master equation built in Theorem 2.7.2 which satisfies (2.7.7) and (2.7.8). It solves

$$\begin{cases} -\partial_t U - \Delta_x U + H(x, D_x U) - \displaystyle\int_{\mathbb{T}^d} \text{div}_y D_m U \; dm(y) \\ + \displaystyle\int_{\mathbb{T}^d} \langle D_m U(t,x,m,y), D_p H(y, D_x U(t,y,m))\rangle dm(y) = F(x,m) \\ \qquad\qquad \text{in } (0,T)\times\mathbb{T}^d\times\mathcal{P}_2 \\ U(T,x,m) = G(x,m) \qquad \text{in } \mathbb{T}^d \times \mathcal{P}_2 \end{cases} \tag{2.8.2}$$

Throughout the section, we suppose that the assumptions of the previous section are in force.

2.8.1 The Nash system

In this section we recall a classical interpretation of the Nash system:

$$\begin{cases} -\partial_t v^{N,i} - \sum_j \Delta_{x_j} v^{N,i} + H(x_i, D_{x_i} v^{N,i}) \\ \qquad\qquad + \sum_{j\neq i} \langle D_p H(x_j, D_{x_j} v^{N,j}), D_{x_j} v^{N,i}\rangle = F(x_i, m_X^{N,i}) \qquad \text{in } (0,T)\times \mathbb{T}^{Nd} \\ v^{N,i}(T,x) = G(x_i, m_X^{N,i}) \qquad \text{in } \mathbb{T}^{Nd} \end{cases}$$

where we set, for $X = (x_1, \ldots, x_N) \in (\mathbb{T}^d)^N$, $m_X^{N,i} = \dfrac{1}{N-1}\sum_{j\neq i} \delta_{x_j}$.

The game consists, for each player $i = 1, \dots, N$ and for any initial position $x_0 = (x_0^1, \dots, x_0^N)$, in minimizing

$$J_i(t_0, x_0, (\alpha^j)) = \mathbb{E}\left[\int_{t_0}^T L(X_t^i, \alpha_t^i) + F(X_t^i, m_{X_t}^{N,i})\ dt + G(X_t^i, m_{X_t}^{N,i})\right]$$

where, for each $i = 1, \dots, N$,

$$dX_t^i = \alpha_t^i dt + \sqrt{2} dB_t^i, \qquad X_{t_0}^i = x_0^i$$

We have set $X_t = (X_t^1, \dots, X_t^N)$. The Brownian motions (B_t^i) are independent, but the controls (α^i) are supposed to depend on the filtration $\mathcal{F}$ generated by all the Brownian motions.

Proposition 2.8.1 (Verification Theorem) *Let $(v^{N,i})$ be a classical solution to the above system. Then the $N-$uple of maps $(\alpha^{i,*})_{i=1,\dots,d} := (-D_pH(x_i, D_{x_i}v^{N,1})_{i=1,\dots,d}$ is a Nash equilibrium in feedback form of the game: for any $i = 1, \dots, d$, for any initial condition $(t_0, x_0) \in [0,T] \times \mathbb{T}^{Nd}$, for any control α^i adapted to the whole filtration $\mathcal{F}$, one has*

$$J_i(t_0, x_0, (\alpha^{j,*})) \leq J_i(t_0, x_0, \alpha^i, (\alpha^{j,*})_{j\neq i})$$

Proof. The proof—which is standard—relies on verification argument and is left to the reader. □

2.8.2 Finite dimensional projections of U

Let U be the solution to the master equation (2.8.2). For $N \geq 2$ and $i \in \{1, \dots, N\}$ we set

$$u^{N,i}(t, X) = U(t, x_i, m_X^{N,i}) \quad \text{where } X = (x_1, \dots, x_N) \in (\mathbb{T}^d)^N,\ m_X^{N,i} = \frac{1}{N-1}\sum_{j\neq i} \delta_{x_j}.$$

Note that the $u^{N,i}$ are at least C^2 with respect to the x_i variable because so is U. Moreover, $\partial_t u^{N,i}$ exists and is continuous because of the equation satisfied by U. The next statement says that $u^{N,i}$ is actually globally $C^{1,1}$ in the space variables:

Proposition 2.8.2 *For any $N \geq 2$, $i \in \{1, \dots, N\}$, $u^{N,i}$ is of class $C^{1,1}$ in the space variables, with*

$$D_{x_j} u^{N,i}(t, X) = \frac{1}{N-1} D_m U(t, x_i, m_X^N, x_j) \qquad (j \neq i)$$

and

$$\left\| D_{x_k, x_j} u^{N,i}(t, \cdot)\right\|_\infty \leq \frac{C}{N} \qquad (k \neq,\ j \neq i).$$

Proof. Let $X = (x_j)$ be such that $x_j \neq x_k$ for any $j \neq k$. Let $\varepsilon := \min_{j\neq k} |x_j - x_k|$. For $V = (v_j) \in (\mathbb{R}^d)^N$ with $v_i = 0$, we consider a smooth vector field ϕ such that

$$\phi(x) = v_j \qquad \text{if } x \in B(x_j, \varepsilon/4).$$

Then, as U satisfies (2.7.7), (2.7.8), we can apply Proposition 2.6.2 which says that, (omitting the dependence with respect to t for simplicity)

$$\begin{aligned} u^{N,i}(X+V) - u^{N,i}(X) &= U((id+\phi)\sharp m_X^{N,i}) - U(m_X^{N,i}) \\ &= \int_{\mathbb{T}^d} \langle D_m U(m_X^{N,i}, y), \phi(y)\rangle dm_X^{N,i}(y) + O(\|\phi\|^2_{L^2(m_X^{N,i})}) \\ &= \frac{1}{N-1}\sum_{j\neq i} \langle D_m U(m_X^{N,i}, x_j), v_j\rangle + O(\sum_{j\neq i} |v_j|^2) \end{aligned}$$

This shows that $u^{N,i}$ has a first order expansion at X with respect to the variables $(x_j)_{j\neq i}$ and that

$$D_{x_j} u^{N,i}(t,X) = \frac{1}{N-1} D_m U(t, x_i, m_X^N, x_j) \qquad (j \neq i).$$

As $D_m U$ is continuous with respect to all its variables, $u^{N,i}$ is C^1 with respect to the space variables in $[0,T]\times \mathbb{T}^{Nd}$. ◻

We now show that $(u^{N,i})$ is "almost" a solution to the Nash system (2.8.1):

Proposition 2.8.3 *One has, for any $i \in \{1,\dots,N\}$,*

$$\left\{ \begin{array}{l} -\partial_t u^{N,i} - \sum_j \Delta_{x_j} u^{N,i} + H(x_i, D_{x_i} u^{N,i}) \\ \qquad + \sum_{j\neq i} \langle D_{x_j} u^{N,i}(t,X), D_p H(x_j, D_{x_j} u^{N,j}(t,X))\rangle = F(x_i, m_X^{N,i}) + r^{N,i}(t,X) \\ \qquad\qquad\qquad\qquad \text{in } (0,T)\times\mathbb{T}^{Nd} \\ u^{N,i}(T,X) = G(x_i, m_X^{N,i}) \qquad \text{in } \mathbb{T}^{Nd} \end{array} \right. \tag{2.8.3}$$

where $r^{N,i} \in L^\infty((0,T)\times\mathbb{T}^{dN})$ with

$$\|r^{N,i}\|_\infty \leq \frac{C}{N}.$$

Proof. As U solves (2.8.2), one has at a point $(t, x_i, m_X^{N,i})$:

$$\begin{aligned} -\partial_t U - \Delta_x U + H(x_i, D_x U) - \int_{\mathbb{T}^d} \text{div}_y D_m U(t, x_i, m_X^{N,i}, y)\ dm_X^{N,i}(y) \\ + \int_{\mathbb{T}^d} \langle D_m U(t, x_i, m_X^{N,i}, y), D_p H(y, D_x U(t, y, m_X^{N,i}))\rangle dm_X^{N,i}(y) = F(x_i, m_X^{N,i}) \end{aligned}$$

So $u^{N,i}$ satisfies:

$$\begin{aligned} -\partial_t u^{N,i} - \Delta_{x_i} u^{N,i} + H(x_i, D_{x_i} u^{N,i}) - \frac{1}{N-1}\sum_{j\neq i} \text{div}_y D_m U(t, x_i, m_X^{N,i}, y_j) \\ + \frac{1}{N-1}\sum_{j\neq i} \langle D_{x_j} u^{N,i}(t,X), D_p H(x_j, D_x U(t, x_j, m_X^{N,i}))\rangle = F(x_i, m_X^{N,i}) \end{aligned}$$

By the Lipschitz continuity of $D_x U$ with respect to m, we have

$$\left| D_x U(t, x_j, m_X^{N,i}) - D_x U(t, x_j, m_X^{N,j}) \right| \leq C\mathbf{d}_1(m_X^{N,i}, m_X^{N,j}) \leq \frac{C}{N-1},$$

so that, by Proposition 2.8.2,

$$\left| \frac{1}{N-1} D_x U(t, x_j, m_X^{N,i}) - D_{x_j} u^{N,j}(t, X) \right| \le \frac{C}{N^2}$$

and

$$\begin{aligned} \frac{1}{N-1} \sum_{j \neq i} &\langle D_{x_j} u^{N,i}(t, X), D_p H(x_j, D_x U(t, x_j, m_X^{N,i})) \rangle \\ &= \sum_{j \neq i} \langle D_{x_j} u^{N,i}(t, X), D_p H(x_j, D_{x_j} u^{N,j}(t, X)) \rangle + O(1/N). \end{aligned}$$

On the other hand,

$$\sum_j \Delta_{x_j} u^{N,i} = \Delta_{x_i} u^{N,i} + \sum_{j \neq i} \Delta_{x_j} u^{N,i}$$

where, using Proposition 2.8.2 and Lipschitz continuity of $D_m U$ with respect to m,

$$\sum_{j \neq i} \Delta_{x_j} u^{N,i} = \int_{\mathbb{T}^d} \operatorname{div}_y D_m U(t, x_i, m_X^{N,i}, y) dm_X^{N,i}(y) + O(1/N) \qquad \text{a.e.}$$

Therefore

$$\begin{aligned} -\partial_t u^{N,i} - \sum_j \Delta_{x_j} u^{N,i} + H(x_i, D_{x_i} u^{N,i}) \\ + \sum_{j \neq i} \langle D_{x_j} u^{N,i}(t, X), D_p H(x_j, D_{x_j} u^{N,j}(t, X)) \rangle + O(1/N) = F(x_i, m_X^{N,i}) \end{aligned}$$

which shows the result. □

2.8.3 Convergence

We now consider a classical *symmetrical* solution $(v^{N,i})$ of the Nash system:

$$\begin{cases} -\partial_t v^{N,i} - \sum_j \Delta_{x_j} v^{N,i} + H(x_i, D_{x_i} v^{N,i}) \\ \qquad + \sum_{j \neq i} \langle D_p H(x_j, D_{x_j} v^{N,j}), D_{x_j} v^{N,i} \rangle = F(x_i, m_X^{N,i}) \qquad \text{in } (0,T) \times \mathbb{T}^{Nd} \\ v^{N,i}(T, x) = G(x_i, m_X^{N,i}) \qquad \text{in } \mathbb{T}^{Nd} \end{cases} \tag{2.8.4}$$

By symmetrical, we mean that, for any $X = (x_l) \in \mathbb{T}^{Nd}$ and for any indices $j \neq k$, if $\widetilde{X} = (\widetilde{x}_l)$ is the N-vector obtained from X by permuting the j and k vectors (i.e., $\widetilde{x}_l = x_l$ for $l \neq j, k$, $\widetilde{x}_j = x_k$, $\widetilde{x}_k = x_j$), then

$$v^{N,i}(t, \widetilde{X}) = v^{N,i}(t, X) \text{ if } i \neq j, k, \text{ while } v^{N,i}(t, \widetilde{X}) = v^{N,k}(t, X) \text{ if } i = j.$$

Note in particular that the $u^{N,i}$ are symmetrical.

Our main result says that the $v^{N,i}$ "converges" to U as $N \to +\infty$. More precisely:

Theorem 2.8.4 *Let $(v^{N,i})$ be a symmetrical solution to (2.8.4). Fix $N \geq 1$ and $i \in \{1, \ldots, N\}$. Given $(t_0, x, m_0) \in [0,T] \times \mathbb{T}^d \times \mathcal{P}_1$, let us set*

$$v^{N,i}(t_0, x, m_0) := \int_{\mathbb{T}^d} \cdots \int_{\mathbb{T}^d} v^{N,i}(t_0, X) \prod_{j \neq i} m_0(dx_j) \qquad \text{where } X = (x_1, \ldots, x_N).$$

Then

$$\left\| v^{N,i}(t_0, \cdot, m_0) - U(t_0, \cdot, m_0) \right\|_{L^1(m_0)} \leq C N^{-1/(d+4)}.$$

where C does not depend on t_0, m_0, i nor N.

The idea of the proof consists in comparing "optimal trajectories" for $v^{N,i}$ and for $u^{N,i}$. For this, let us fix $t_0 \in [0,T)$, $m_0 \in \mathcal{P}_2$ and let (z_i) be an i.i.d family of N random variables of law m_0. We set $Z = (z_i)$. Let also (B^i) be a family of N independent B.M. which is also independent of (z_i). We consider the systems of SDEs with variables $(X = (x_i)_{i \in \{1,\ldots,N\}})$ and $(Y = (y_i)_{i \in \{1,\ldots,N\}})$:

$$\begin{cases} dx_{i,t} = -D_p H(x_{i,t}, D_{x_i} u^{N,i}(t, X_t))dt + \sqrt{2} dB_t^i \\ x_{i,t_0} = z_i \end{cases}$$

and

$$\begin{cases} dy_{i,t} = -D_p H(y_{i,t}, D_{x_i} v^{N,i}(t, Y_t))dt + \sqrt{2} dB_t^i \\ y_{i,t_0} = z_i \end{cases} \tag{2.8.5}$$

Note that, since the $u^{N,i}$ are symmetrical, all the (x_i) have the same law. The same holds for the (y_i).

Theorem 2.8.5 *We have*

$$\mathbb{E}\left[\sup_{t \in [t_0, T]} |y_{i,t} - x_{i,t}| \right] \leq C N^{-1/2}$$

and

$$\mathbb{E}\left[\sup_{t \in [t_0,T]} \left| u^{N,i}(t, X_t) - v^{N,i}(t, Y_t) \right| + \int_{t_0}^T |D_{x_i} v^{N,i}(t, Y_t) - D_{x_i} u^{N,i}(t, X_t)| dt \right] \leq C N^{-1/2}. \tag{2.8.6}$$

where C does not depend on t_0, m_0 and N.

Proof. [Proof of Theorem 2.8.5.] For simplicity, we work with $t_0 = 0$. Let us introduce a few notations: for $X = (x_j)_{j \in \{1,\ldots,N\}} \in \mathbb{T}^{Nd}$, $z \in \mathbb{T}^d$ and $k \in \{1, \ldots N\}$, let us denote by $u^{N,i}(t, z, X^i)$ the value of $u^{N,i}(t, \cdot)$ evaluated at the point $\widetilde{X} = (\widetilde{x}_j)_{j \in \{1,\ldots,N\}}$ obtained from X by replacing x_i by z (i.e, $\widetilde{x}_j = x_j$ if $j \neq i$ and $\widetilde{x}_i = z$).

As $(u^{N,i})$ satisfies (2.8.3), we have by standard computation

$$\begin{aligned} \mathbb{E}[u^{N,i}(0, Z)] \; = \; & \mathbb{E}\left[\int_0^T (-H(x_{i,t}, D_{x_i} u^{N,i}) + \langle D_p H(x_{i,t}, D_{x_i} u^{N,i}), D_{x_i} u^{N,i} \rangle \right. \\ & \left. + F(x_{i,t}, m_{X_t}^{N,i}) + r^{N,i}) \; dt + G(x_{i,T}, m_{X_T}^{N,i}) \right] \end{aligned} \tag{2.8.7}$$

(where $u^{N,i}$ is always evaluated at (t, X_t)). We now compute the variation of $v^{N,i}(t, x_{i,t}, Y^i_t)$. Since the individual B.M. driving $x_{i,t}$ and those driving the $(y_{j,\cdot})_{j\neq i}$ are independent, we have, using equation (2.8.4) satisfied by v,

$$\begin{aligned}
&dv^{N,i}(t, x_{i,t}, Y^i_t) \\
&\quad = \Big(\partial_t v^{N,i} + \sum_j \Delta_{x_j} v^{N,i} - \sum_{j\neq i} \langle D_{x_j} v^{N,i}, D_p H(y_{j,t}, D_{x_i} v^{N,i}(t, Y_t))\rangle \\
&\qquad -\langle D_{x_i} v^{N,i}, D_p H(x_{i,t}, D_{x_i} u^{N,i}(t, X_t))\rangle\Big)\, dt + \sqrt{2}\sum_j \langle D_{x_j} v^{N,i}, dB^j_t\rangle \\
&\quad = (H(x_{i,t}, D_{x_i} v^{N,i}) - \langle D_{x_i} v^{N,i}, D_p H(x_{i,t}, D_{x_i} u^{N,i}(t, X_t))\rangle)dt \\
&\qquad -F(x_{i,t}, m^{N,i}_{Y_t})dt + \sqrt{2}\sum_j \langle D_{x_j} v^{N,i}, dB^j_t\rangle
\end{aligned}$$

where $v^{N,i}$ is evaluated at $(t, x_{i,t}, Y^i_t)$. Taking the expectation, integrating in time and using the terminal condition of $v^{N,i}$ gives:

$$\begin{aligned}
\mathbb{E}\left[v^{N,i}(0, Z)\right] \;=\; \mathbb{E}\Big[\int_0^T &(-H(x_{i,t}, D_{x_i} v^{N,i}) + \langle D_{x_i} v^{N,i}, D_p H(x_{i,t}, D_{x_i} u^{N,i}(t, X_t))\rangle \\
&+F(x_{i,t}, m^{N,i}_{Y_t}))\ dt + G(x_{i,T}, m^{N,i}_{Y_T})\Big]
\end{aligned}$$

So

$$\begin{aligned}
&\mathbb{E}[u^{N,i}(0, Z) - v^{N,i}(0, Z)] \\
&\quad = \mathbb{E}\Big[\int_0^T H(x_{i,t}, D_{x_i} v^{N,i}) - H(x_{i,t}, D_{x_i} u^{N,i}) - \langle D_p H(x_{i,t}, D_{x_i} u^{N,i}), D_{x_i} v^{N,i} - D_{x_i} u^{N,i}\rangle \\
&\qquad +(F(x_{i,t}, m^{N,i}_{X_t}) - F(x_{i,t}, m^{N,i}_{Y_t})) + r^{N,i}\ dt + G(x_{i,T}, m^{N,i}_{X_T}) - G(x_{i,T}, m^{N,i}_{Y_T})\Big]
\end{aligned}$$

Let us set, for $z \geq 0$,

$$\Psi(z) = \begin{cases} z^2 & \text{if } z \in [0, 1] \\ 2z - 1 & \text{if } z \geq 1 \end{cases}$$

Note for later use that Ψ is increasing and convex on $[0, +\infty)$. By assumption (2.3.14) on H, we have, for any $C_0 > 0$, for any $x \in \mathbb{T}^d$ and $p, q \in \mathbb{R}^d$ with $\min\{|p|, |q|\} \leq C_0$,

$$H(x, q) - H(x, p) - \langle D_p H(x, p), q - p\rangle \;\geq C^{-1}\Psi(|p - q|)$$

where C depends only on the constant in (2.3.14) and on C_0. Therefore, as $\|D_{x_i} u^{N,i}\|_\infty$ is bounded independently of N by a constant C_0 and using the estimate $\|r^{N,i}|_\infty \leq CN^{-1}$,

$$\begin{aligned}
&\mathbb{E}[u^{N,i}(0, Z) - v^{N,i}(0, Z)] \\
&\quad \geq \mathbb{E}\Big[\int_0^T (1/C)\Psi(|D_{x_i} v^{N,i}(t, x_{i,t}, Y^i_t) - Du^{N,i}(t, X_t)|) \\
&\qquad +(F(x_{i,t}, m^{N,i}_{X_t}) - F(x_{i,t}, m^{N,i}_{Y_t}))\ dt + G(x_{i,T}, m^{N,i}_{X_T}) - G(x_{i,T}, m^{N,i}_{Y_T})\Big] - CN^{-1}
\end{aligned}$$

Computing in the same way the variation of the terms $-u^{N,i}(t, y_{i,t}, X^i_t) + v^{N,i}(t, Y_t)$, we get

$$\begin{aligned}
&\mathbb{E}[-u^{N,i}(0, Z) + v^{N,i}(0, Z)] \\
&\quad \geq \mathbb{E}\Big[\int_0^T (1/C)\Psi(|D_{x_i} v^{N,i}(t, Y_t) - D_{x_i} u^{N,i}(t, y_{i,t}, X^i_t)|) \\
&\qquad +(F(y_{i,t}, m^{N,i}_{Y_t}) - F(y_{i,t}, m^{N,i}_{X_t}))\ dt + G(y_{i,T}, m^{N,i}_{Y_T}) - G(y_{i,T}, m^{N,i}_{X_T})\Big] - CN^{-1}
\end{aligned}$$

Therefore

$$\begin{aligned}CN^{-1} \geq \mathbb{E}\left[\int_0^T (1/C)\Psi(|D_{x_i}v^{N,i}(t,x_{i,t},Y_t^i) - D_{x_i}u^{N,i}(t,X_t)|)dt\right] \\ +\mathbb{E}\left[\int_0^T (1/C)\Psi(|D_{x_i}v^{N,i}(t,Y_t) - D_{x_i}u^{N,i}(t,y_{i,t},X_t^i)|)dt\right] \\ +\mathbb{E}\left[\int_0^T F(x_{i,t},m_{X_t}^{N,i}) - F(x_{i,t},m_{Y_t}^{N,i}) - F(y_{i,t},m_{X_t}^{N,i}) + F(y_{i,t},m_{Y_t}^{N,i}))dt\right] \\ +\mathbb{E}\left[G(x_{i,T},m_{X_T}^{N,i}) - G(x_{i,T},m_{Y_T}^{N,i}) - G(y_{i,T},m_{X_T}^{N,i}) + G(y_{i,T},m_{Y_T}^{N,i})\right].\end{aligned}$$

Let us set $m_X^N = \frac{1}{N}\sum_j \delta_{x_j}$. We note that $\mathbf{d}_1(m_X^{N,i}, m_X^N) \leq CN^{-1}$. Hence

$$\begin{aligned}CN^{-1} \geq \mathbb{E}\left[\int_0^T (1/C)\Psi(|D_{x_i}v^{N,i}(t,x_{i,t},Y_t^i) - D_{x_i}u^{N,i}(t,X_t)|)dt\right] \\ +\mathbb{E}\left[\int_0^T (1/C)\Psi(|D_{x_i}v^{N,i}(t,Y_t) - D_{x_i}u^{N,i}(t,y_{i,t},X_t^i)|)dt\right] \\ +\mathbb{E}\left[\int_0^T F(x_{i,t},m_{X_t}^{N}) - F(x_{i,t},m_{Y_t}^{N}) - F(y_{i,t},m_{X_t}^{N}) + F(y_{i,t},m_{Y_t}^{N}))dt\right] \\ +\mathbb{E}\left[G(x_{i,T},m_{X_T}^{N}) - G(x_{i,T},m_{Y_T}^{N}) - G(y_{i,T},m_{X_T}^{N}) + G(y_{i,T},m_{Y_T}^{N})\right].\end{aligned}$$

We now sum these expressions over i. Since

$$\begin{aligned}\sum_i G(x_{i,T},m_{X_T}^{N}) - G(x_{i,T},m_{Y_T}^{N}) - G(y_{i,T},m_{X_T}^{N}) + G(y_{i,T},m_{Y_T}^{N}) \\ = N\int_{\mathbb{T}^d}(G(x,m_{X_T}^N) - G(x,m_{Y_T}^N))d(m_{X_T}^N - m_{Y_T}^N)) \ \geq \ 0,\end{aligned}$$

and the same holds for the terms involving F, we obtain:

$$\begin{aligned}C \geq \sum_i \mathbb{E}\left[\int_0^T \Psi(|D_{x_i}v^{N,i}(t,x_{i,t},Y_t^i) - D_{x_i}u^{N,i}(t,X_t)|)dt\right] \\ +\sum_i \mathbb{E}\left[\int_0^T \Psi(|D_{x_i}v^{N,i}(t,Y_t) - D_{x_i}u^{N,i}(t,y_{i,t},X_t^i)|)dt\right]\end{aligned}$$

By symmetry of the $(u^{N,i})$ and of the $(v^{N,i})$, the random variables

$$D_{x_i}v^{N,i}(t,x_{i,t},Y_t^i) - D_{x_i}u^{N,i}(t,X_t)$$

have the same law for any i. We have therefore

$$\mathbb{E}\left[\int_0^T \Psi(|D_{x_i}v^{N,i}(t,x_{i,t},Y_t^i) - D_{x_i}u^{N,i}(t,X_t)|)dt\right] \leq CN^{-1} \qquad \forall i \in \{1,\dots,N\}$$

and, in the same way

$$\mathbb{E}\left[\int_0^T \Psi(|D_{x_i}v^{N,i}(t,Y_t) - D_{x_i}u^{N,i}(t,y_{i,t},X_t^i)|)dt\right] \leq CN^{-1} \qquad \forall i \in \{1,\dots,N\}. \qquad (2.8.8)$$

We are now ready to estimate the difference $x_{i,t} - y_{i,t}$. In view of the equation satisfied by x_i and by y_i, we have

$$\begin{aligned}
|x_{i,t} - y_{i,t}| \ &\leq \ \int_0^t |- D_pH(x_{i,s}, D_{x_i}u^{N,i}(s, X_s)) + D_pH(y_{i,s}, D_{x_i}u^{N,i}(s, y_{i,s}, X_s^i))| \ ds \\
&\quad + \int_0^t |- D_pH(y_{i,s}, D_{x_i}u^{N,i}(s, y_{i,s}, X_s^i)) + D_pH(y_{i,s}, D_{x_i}v^{N,i}(s, Y_s)) \ ds \\
&\leq \ C\int_0^t |x_{i,s} - y_{i,s}| \ ds + C\int_0^T |D_{x_i}u^{N,i}(s, y_{i,s}, X_s^i) - D_{x_i}v^{N,i}(s, Y_s)| \ ds
\end{aligned}$$

where we have used the uniform Lipschitz bound of $D_{x_i}u^{N,i}$ in the variable x_i. Note that by convexity of Ψ and (2.8.8),

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T |D_{x_i}u^{N,i}(s, y_{i,s}, X_s^i) - D_{x_i}v^{N,i}(s, Y_s)| \ ds\right] \\
&\qquad \leq \Psi^{-1}\left(\mathbb{E}\left[\int_0^T \Psi(|D_{x_i}u^{N,i}(s, y_{i,s}, X_s^i) - D_{x_i}v^{N,i}(s, Y_s)|) \ ds\right]\right) \leq \Psi^{-1}(CN^{-1}) \leq CN^{-1/2}.
\end{aligned}$$

So, by Gronwall inequality, we obtain

$$\mathbb{E}\left[\sup_{t\in[0,T]} |x_{i,t} - y_{i,t}|\right] \leq CN^{-1/2} \qquad \forall i \in \{1, \ldots, N\}. \tag{2.8.9}$$

We now estimate the difference between $u^{N,i}$ and $v^{N,i}$: recall first that, for any $t \in [0, T]$,

$$\begin{aligned}
u^{N,i}(t, X_t) \ = \ \mathbb{E}^{Z,t}\Bigg[&\int_t^T (-H(x_{i,s}, D_{x_i}u^{N,i}) + \langle D_pH(x_{i,s}, D_{x_i}u^{N,i}), D_{x_i}u^{N,i}\rangle \\
&+F(x_{i,s}, m_{X_s}^{N,i}) + r^{N,i}) + r^{N,i} \ ds + G(x_{i,T}, m_{X_T}^{N,i})\Bigg]
\end{aligned}$$

where $E^{Z,t} = \mathbb{E}[\cdot \mid Z, \ \mathcal{F}_t]$. The symmetrical expression holds for $v^{N,i}(t, Y_t)$. Hence

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]} \left|u^{N,i}(t, X_t) - v^{N,i}(t, Y_t)\right|\right] \\
&\quad \leq \mathbb{E}\Bigg[\int_0^T C\left[|x_{i,t} - y_{i,t}| + |D_{x_i}v^{N,i}(t, Y_t) - D_{x_i}u^{N,i}(t, X_t)|\right] \\
&\qquad +|F(x_{i,t}, m_{X_t}^{N,i}) - F(y_{i,t}, m_{Y_t}^{N,i})| \ dt + |G(x_{i,T}, m_{X_T}^{N,i}) - G(y_{i,t}, m_{Y_t}^{N,i})|\Bigg] + CN^{-1}
\end{aligned}$$

We estimate the various terms in the above inequality. The difference $|x_{i,t} - y_{i,t}|$ is bounded by (2.8.9). The second expression can be treated as above:

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T |D_{x_i}v^{N,i}(t, Y_t) - D_{x_i}u^{N,i}(t, X_t)|\right] \\
&\quad \leq \mathbb{E}\left[\int_0^T |D_{x_i}v^{N,i}(t, Y_t) - D_{x_i}u^{N,i}(t, x_{i,t}, Y_t^i)| + |D_{x_i}v^{N,i}(t, x_{i,t}, Y_t^i) - D_{x_i}u^{N,i}(t, X_t)|\right] \\
&\quad \leq CN^{-1/2}
\end{aligned}$$

where the last inequality is given by (2.8.8). In particular this proves the second half of inequality (2.8.6). Let us proceed with the first half. As $\mathbf{d}_1(m_X^{N,i}, m_Y^{N,i}) \leq \frac{1}{N-1}\sum_{j\neq i}|x_j - y_j|$, we get, by the Lipschitz continuity of F and G,

$$\mathbb{E}\left[\int_0^T |F(x_{i,t}, m_{X_t}^{N,i}) - F(y_{i,t}, m_{Y_t}^{N,i})|\ dt + |G(x_{i,T}, m_{X_T}^{N,i}) - G(y_{i,t}, m_{Y_t}^{N,i})|\right]$$
$$\leq C\mathbb{E}\left[\sup_{t\in[0,T]}\left(|x_{i,t} - y_{i,t}| + N^{-1}\sum_{j\neq i}|x_{j,t} - y_{j,t}|\right)\right] \ \leq\ CN^{-1/2}.$$

Putting the above estimates together:

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left|u^{N,i}(t, X_t) - v^{N,i}(t, Y_t)\right|\right] \leq CN^{-1/2}.$$

□

Proof. [Proof of Theorem 2.8.4.] From the Lipschitz continuity of U and Horowitz and Karandikar Lemma (Lemma 2.3.8), we have, for any $x_i \in \mathbb{T}^d$,

$$\int_{\mathbb{T}^{d(N-1)}} |u^{N,i}(t, X) - U(t, x_i, m_0)| \prod_{j\neq i} m_0(dx_j)$$
$$= \int_{\mathbb{T}^{d(N-1)}} |U(t, x_i, m_X^{N,i}) - U(t, x_i, m_0)| \prod_{j\neq i} m_0(dx_j)$$
$$\leq C\int_{\mathbb{T}^{d(N-1)}} \mathbf{d}_1(m_X^{N,i}, m_0) \prod_{j\neq i} m_0(dx_j)\ \leq\ CN^{-1/(d+4)}.$$

Combining Theorem 2.8.5 with the above inequality, we obtain therefore

$$\int_{\mathbb{T}^d}\left|\int_{\mathbb{T}^{d(N-1)}} v^{N,i}(t, (x_j)) \prod_{j\neq i} m_0(dx_j) - U(t, x_i, m_0)\right|\ dm_0(x_i)$$
$$\leq \mathbb{E}\left[|v^{N,i}(t, Z) - u^{N,i}(t, Z)|\right] + \int_{\mathbb{T}^{dN}} |u^{N,i}(t, X) - U(t, x_i, m_0)| \prod_j m_0(dx_j)$$
$$\leq CN^{-1/2} + CN^{-1/(d+4)}\ \leq\ CN^{-1/(d+4)}.$$

□

2.8.4 Comment

The convergence of the Nash system as the number of players tends to infinity was known in only two cases: for ergodic mean field games (Larsy-Lions [245]), because in this case the Nash equilibrium system reduces to a coupled system of N equations in $\mathbb{T}^d$ (instead of N equations in $\mathbb{T}^{Nd}$ as (**??**)); or in short time (Lions [248]), where the estimates on the derivatives of the $v^{N,i}$ propagate from the initial condition. The result presented here is a simplified version of [**?**].

Bibliography

[1] S. Agmon, On positivity and decay of solutions of second order elliptic equations on Riemannian manifolds, pp. 19-52, in: Methods of Functional Analysis and Theory of Elliptic Equations, D. Greco, ed., Liguori Ed.Napoli, 1983.

[2] G. Alberti, L. Ambrosio and X. Cabré, On a long-standing conjecture of E. De Giorgi: symmetry in 3D for general nonlinearities and a local minimality property. Acta Appl. Math. **65**, 2001, 9–33.

[3] L. Ambrosio and X. Cabré, Entire solutions of semilinear elliptic equations in $\mathbb{R}^3$ and a conjecture of de Giorgi, Jour. Amer. Math. Soc. **13**, 2000, 725–739.

[4] D.G. Aronson, H.F. Weinberger, *Multidimensional nonlinear diffusion arising in population genetics*, Adv. Math. **30** (1978), 33–76.

[5] H. Bahouri, J.-Y. Chemin, *Équations de transport relatives á des champs de vecteurs nonLipschitziens et mécanique des uides. (French) [Transport equations for non-Lipschitz vector Fields and fluid mechanics],* Arch. Rational Mech. Anal., **127** (1994), no. 2, 159–181

[6] I.J. Bakelman, *Convex analysis and nonlinear geometric elliptic equations*, Springer-Verlag, Berlin, 1994

[7] G. Barles, *Solutions de viscosité des équations de Hamilton-Jacobi*, Mathématiques & Applications, **17** Springer-Verlag, Paris, 1994.

[8] R.F. Bass, *Probabilistic Techniques in Analysis*, Springer, 1995.

[9] R.F. Bass, *Diffusions and Elliptic Operators*, Springer, 1998.

[10] H. Bateman, *Some recent researches on the motion of fluids,* Monthly Weather Review **43** (1915), 163-170

[11] J.T. Beale, T. Kato and A. Majda, *Remarks on the breakdown of smooth solutions for the 3D Euler equations,* Commun. Math. Phys. **94** (1984), 61–66

[12] A. Bensoussan, J.L. Lions and G. Papanicoalou, *Asymptotic Analysis for Periodic Structures*, AMS, 2011.

[13] H. Berestycki, L. Caffarelli and L. Nirenberg, Monotonicity for elliptic equations in unbounded Lipschitz domains, Comm. Pure Appl. Math., **50**, 1997, 1089–1111.

[14] H. Berestycki and F. Hamel, *Reaction-diffusion Equations And Propagation Phenomena*, Springer, to appear.

[15] H. Berestycki, F. Hamel and R. Monneau, One-dimensional symmetry of bounded entire solutions of some elliptic equations, Duke Math. Jour., **103**, 2000, 375-396.

[16] H. Berestycki, F. Hamel and N. Nadirashvili, Elliptic eigenvalue problems with large drift and applications to nonlinear propagation phenomena. Comm. Math. Phys. **253**, 2005, 451–480.

[17] H. Berestycki, F. Hamel and G. Nadin, Asymptotic spreading in heterogeneous diffusive excitable media. J. Funct. Anal. **255**, 2008, 2146–2189.

[18] H. Berestycki, F. Hamel and L. Roques, *Analysis of the periodically fragmented environment model : I - Species persistence,* J. Math. Biol. 51 (2005), 75–113.

[19] H. Berestycki, F. Hamel and L. Rossi, Liouville-type results for semilinear elliptic equations in unbounded domains, Ann. Mat. Pura Appl. (4) **186**, 2007, 469–507.

[20] H. Berestycki, A. Kiselev, A. Novikov and L. Ryzhik, The explosion problem in a flow, Jour. d'Anal. Mathématique, **110**, 2010, 31–65.

[21] H. Berestycki and L. Nirenberg, On the method of moving planes and the sliding method, Bol. Soc. Brasil. Mat. **22**, 1991, 1–37.

[22] H. Berestycki and L. Nirenberg, Traveling fronts in cylinders, Annales de l'IHP, Analyse non linéare, **9**, 1992, 497-572.

[23] H. Berestycki, L. Nirenberg and S.R.S Varadhan, The principal eigenvalue and maximum principle for second-order elliptic operators in general domains, Comm. Pure Appl. Math., **47**, 1994, 47–92.

[24] H. Berestycki and L. Rossi, On the principal eigenvalue of elliptic operators in $\mathbb{R}^N$ and applications. J. Eur. Math. Soc. (JEMS) **8**, 2006, 195–215.

[25] J.M. Burgers, *A mathematical model illustrating the theory of turbulence,* Adv. Appl. Mech **1** (1948), 171–199

[26] X. Cabre, Partial differential equations, geometry and stochastic control (in Catalan), Butl. Soc. Catalana Mat. **15**, 2000, 7–27.

[27] X. Cabre, Elliptic PDEs in probability and geometry: symmetry and regularity of solutions, Discr. Cont. Sys. B, **20**, 2008, 425–457.

[28] X. Cabre, A.-C. Coulon and J.-M. Roquejoffre, Propagation in Fisher-KPP type equations with fractional diffusion in periodic media, C. R. Acad. Sci. Paris, Ser. I **350**, (2012) 885–890

[29] L. Caroll, Alice's Adventures in Wonderland, Macmillan, 1866.

[30] A.-C. Coulon and J.-M. Roquejoffre, Transition between linear and exponential propagation in Fisher-KPP type reaction-diffusion equations. Comm. Partial Differential Equations, **37** (2012), 2029–2049.

[31] L. Caffarelli and A. Vasseur. *Drift diffusion equations with fractional diffusion and the quasi-geostrophic equation*, Ann. of Math. (2) **171** (2010), 1903–1930

[32] A.J. Chorin and J.E. Marsden, *A Mathematical Introduction to Fluid Mechanics,* Springer, 1993

[33] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations,* New York: McGraw-Hill, 1955

[34] P. Constantin, A. Kiselev, L. Ryzhik and A. Zlatos, *Diffusion and mixing in fluid flow,* Ann. of Math., **168**, 2008, 643–674.

[35] P. Constantin, A. Majda and E. Tabak. *Formation of strong fronts in the 2D quasi-geostrophic thermal active scalar*. Nonlinearity, 7, (1994), 1495–1533

[36] P. Constantin and V. Vicol, *Nonlinear maximum principles for dissipative linear nonlocal operators and applications,* Geometric and Functional Analysis, **22** (2012), 1289-1321

[37] P. Constantin, A. Tarfulea, and V. Vicol, *Long time dynamics of forced critical SQG,* Comm. Math. Phys. **335** (2015), no. 1, 93–141

[38] P. Constantin and C. Foias, *Navier-Stokes Equations,* University of Chicago Press, 1988

[39] P. Constantin, Q. Nie, and N. Schorghofer, *Nonsingular surface-quasi-geostrophic flow,* Phys. Lett. A, **24** (1998), 168–172

[40] P. Constantin and J. Wu, *Regularity of Hölder continuous solutions of the supercritical quasi-geostrophic equation,* Ann. Inst. H. Poincare Anal. Non Linearie, **25** (2008), 1103–1110

[41] D. Cordoba, *Nonexistence of simple hyperbolic blow up for the quasi-geostrophic equation,* Ann. of Math., **148**, (1998), 1135–1152

[42] D. Cordoba and C. Fefferman, *Growth of solutions for QG and 2D Euler equations,* Journal of the AMS, **15**, 665–670

[43] A. Cordoba and D. Cordoba, *A maximum principle applied to to quasi-geostrophic equations,* Commun. Math. Phys. **249** (2004), 511–528

[44] M.G. Crandall and P.-L. Lions, Viscosity solutions of Hamilton-Jacobi equations, Trans. Amer. Math. Soc., **277**, 1983, 1–42.

[45] D'Alembert, *Recherches sur la courbe que forme une corde tenduë mise en vibration,* Histoire de l'académie royale des sciences et belles lettres de Berlin, **3** (1747), 214–219

[46] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology*, Vol. 3: Spectral Theory and Applications, Springer, 2000.

[47] M. del Pino, M. Kowalczyk, and J. Wei, On a conjecture by De Giorgi in dimensions 9 and higher. Symmetry for elliptic PDEs, 115–137, Contemp. Math., 528, Amer. Math. Soc., Providence, RI, 2010.

[48] M. Dabkowski, A. Kiselev, L. Silvestre and V. Vicol, *Global well-posedness of slightly supercritical active scalar equations,* Anal. PDE **7** (2014), no. 1, 43–72

[49] S. Denisov, *Infinite superlinear growth of the gradient for the two-dimensional Euler equation,* Discrete Contin. Dyn. Syst. A, **23** (2009), 755–764

[50] S. Denisov, *Double-exponential growth of the vorticity gradient for the two-dimensional Euler equation,* Proc. Amer. Math. Soc. **143** (2015), no. 3, 1199–1210

[51] S. Denisov, *The sharp corner formation in 2D Euler dynamics of patches: infinite double exponential rate of merging,* Arch. Ration. Mech. Anal. **215** (2015), no. 2, 675–705

[52] C.R. Doering and J.D. Gibbon, *Applied Analysis of the Navier-Stokes Equations,* Cambridge Texts in Applied Mathematics (Book 12), Cambridge University Press, 1995

[53] H. Dong and D. Du, *Global well-posedness and a decay estimate for the critical quasi-geostrophic equation,* Discrete Contin. Dyn. Syst., **21** (2008) no. 4, 1095–1101

[54] H. Dong, D. Du and D. Li, *Finite time singularities and global well-posedness for fractal Burgers' equation,* Indiana Univ. Math. J., **58** (2009), 807-

[55] L. Euler, *Principes généraux du mouvement des fluides,* Mémoires de L'Académie Royale des Sciences et des Belles-Lettres de Berlin **11** (4 September 1755, printed 1757), 217–273; reprinted in Opera Omnia ser. 2 **12**, 219–250

[56] L.C. Evans, Partial Differential Equations, Second Edition, AMS, 2010.

[57] L.C. Evans and R. Gariepy, *Measure Theory and Fine Properties of Funcitons*, CRC Press, 1992.

[58] L.C. Evans and P. Souganidis, A PDE approach to certain large deviation problems for systems of parabolic equations. Analyse non linéaire (Perpignan, 1987). Ann. Inst. H. Poincaré Anal. Non Linéaire **6** (1989), suppl., 229–258.

[59] L.C. Evans and P. Souganidis, A PDE approach to geometric optics for certain semilinear parabolic equations. Indiana Univ. Math. J. **38**, 1989, 141–172.

[60] E.B. Fabes and D.W. Stroock, A new proof of Moser's parabolic Harnack inequality using the old ideas of Nash, Arch. Rational Mech. Anal. **96**, 1986, 327–338.

[61] A. Fathi, *Sur la convergence du semi-groupe de Lax-Oleinik*, C. R. Acad. Sci. Paris Sér. I Math. **327**, 1998, 267–270.

[62] A. Fathi, *Regularity of C^1 solutions of the Hamilton-Jacobi equation*, Ann. Fac. Sci. Toulouse, **12** (2003), 479–516.

[63] A. Fathi, *Weak KAM Theory*, Cambridge University Press.

[64] P.C. Fife, J.B. McLeod, *The approach of solutions of nonlinear diffusion equations to travelling front solutions,* Arch. Rat. Mech. Anal., **65** (1977), 335–361.

[65] R.A. Fisher, The wave of advance of advantageous genes, Ann. Eugenics, **7**, 1937, 355-369.

[66] A.R. Forsyth, *Theory of Differential Equations. Part 4. Partial Differential Equations* (Vol. 5-6), 1906

[67] B. Franke, Integral inequalities for the fundamental solutions of diffusions on manifolds with divergence-free drift. (English summary) Math. Z. **246**, 2004, n 373–403.

[68] M. Freidlin and J. Gártner, The propagation of concentration waves in periodic and random media. (Russian) Dokl. Akad. Nauk SSSR **249**, 1979, 521–525.

[69] M. Freidlin, Functional integration and partial differential equations. Annals of Mathematics Studies, 109. Princeton University Press, Princeton, NJ, 1985.

[70] U. Frisch, *Turbulence,* Cambridge University Press, 1999.

[71] T. Gallay, E. Risler, *A variational proof of global stability for bistable travelling waves,* Differential Integral Equations, 20.8 (2007), 901–926.

[72] P. Garabedian, *Partial Differential Equations,* AMS Publishing, 1998

[73] D. Gilbarg and N. Trudinger, *Elliptic Partial Differential Equations of the Second Order*, Springer-Verlag, Berlin, Heidelberg, 1983.

[74] N. Ghoussoub and C. Gui, On the De Giorgi conjecture in dimensions 4 and 5, Ann. Math., **157**, 2003, 313-334.

[75] Q. Han, *A Basic Course in Partial Differential Equations*, AMS, 2011.

[76] Q. Han and F.Lin, *Elliptic Partial Differential equations*, AMS, 1997.

[77] I. Held, R. Pierrehumbert, S. Garner and K. Swanson. *Surface quasi-geostrophic dynamics*, J. Fluid Mech., 282, (1995), 1–20

[78] E. Hölder, *Über die unbeschränkte Fortsetzbarkeit einer stetigen ebenen Bewegung in einer unbegrenzten inkompressiblen Flüssigkeit,* Math. Z. **37** (1933), 727–738

[79] F. John, *Partial Differential Equations,* Springer.

[80] T. Kato, *Remarks on the Euler and Navier-Stokes equations in* $\mathbb{R}^2$*,* Proc. Symp. Pure Math. **45** (1986), 1–7

[81] Y. Katznelson, *An Introduction to Harmonic Analysis*, Cambridge University Press, 2004.

[82] A. Kiselev, F. Nazarov and R. Shterenberg, *On blow up and regularity in dissipative Burgers equation,* Dynamics of PDEs, **5** (2008), 211–240

[83] A. Kiselev, F. Nazarov and A. Volberg, *Global well-posedness for the critical* 2*D dissipative quasi-geostrophic equation,* Inventiones Math. **167** (2007) 445–453

[84] A. Kiselev, *Some recent results on the critical surface quasi-geostrophic equation: a review. Hyperbolic problems: theory, numerics and applications,* 105–122, Proc. Sympos. Appl. Math., 67, Part 1, Amer. Math. Soc., Providence, RI, 2009

[85] A. Kiselev and F. Nazarov, *A variation on a theme of Caffarelli and Vasseur,* Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **370** (2009)

[86] A. Kiselev and V. Sverak, *Small scale creation for solutions of the incompressible two dimensional Euler equation,* Annals of Math. **180** (2014), 1205-1220

[87] H. Koch, *Transport and instability for perfect fluids,* Math. Ann. **323** (2002), 491–523

[88] A.N. Kolmogorov, I.G. Petrovskii and N.S. Piskunov, Étude de l'équation de la chaleur avec augmentation de la quantité de matière et son application à un problème biologique, Bull. Moskov. Gos. Univ. Mat. Mekh. **1** (1937), 1-25. (see [117] pp. 105-130 for an English transl.)

[89] O.A. Ladyzhenskaya, N.N Uraltseva, V.A. Solonnikov, *Linear and quasilinear equations of parabolic type,*Translations of Mathematical Monographs, AMS, 1968.

[90] J.- M. Lasry, P.- L. Lions, A Remark on Regularization in Hilbert Spaces, Israel Math. J. **55**, 1996, 257–266.

[91] M. Ledoux, *The concentration of measure phenomenon.* Mathematical Surveys and Monographs, 89. American Mathematical Society, Providence, RI, 2001.

[92] P. G. LeFloch, Hyperbolic systems of conservation laws. The theory of classical and nonclassical shock waves.Lectures in Mathematics ETH Zu?rich, Birkh?auser Verlag, Basel, 2002.

[93] P.-L. Lions, *Generalized solutions of Hamilton-Jacobi Equations,* Research Notes in Mathematics, Pitman, 1983.

[94] P.-L. Lions, G. Papanicolaou and S.R.S. Varadhan, Homogenization of Hamilton-Jacobi equations, Preprint.

[95] A. Majda and A. Bertozzi, *Vorticity and Incompressible Flow,* Cambridge University Press, 2002

[96] A. Majda, *Introduction to PDEs and Waves for the Atmoshpere and Ocean,* Courant Lecture Notes in Mathematics, AMS 2003

[97] R. Mancinelli, D. Vergni and A. Vulpiani *Front propagation in reactive systems with anomalous diffusion,*Phys. D, **185** (2003), 175–195.

[98] C. Marchioro and M. Pulvirenti, *Mathematical Theory of Incompressible Nonviscous Fluids,* Applied Mathematical Sciences Series (Springer-Verlag, New York), **96**, 1994

[99] V. Maz'ja, *Sobolev Spaces,* Springer-Verlag, Berlin Heidelberg, 1985

[100] A. Mellet, J. Nolen, L. Ryzhik, J.-M. Roquejoffre, *Stability of generalized transitions fronts,* Communications in PDE, **34**, 2009, 521–552

[101] A. Morgulis, A. Shnirelman and V. Yudovich, *Loss of smoothness and inherent instability of 2D inviscid fluid flows,* Comm. Partial Differential Equations **33** (2008), no. 4-6, 943-968

[102] S. Mirrahimi, J.-M. Roquejoffre, *A class of Hamilton-Jacobi equations with constraint: uniqueness and constructive approach,* Journal of differential equations, **250**, 2016, 4717–4738.

[103] J. Moser, *On the volume elements on a manifold,* Trans. Amer. Math. Soc., **120**, 1965, 286–294.

[104] J.D. Murray, Mathematical biology. I. An introduction. Third edition. Interdisciplinary Applied Mathematics, 17. Springer-Verlag, New York, 2002.

[105] J.D. Murray, Mathematical biology. II. Spatial models and biomedical applications. Third edition. Interdisciplinary Applied Mathematics, 18. Springer-Verlag, New York, 2003

[106] N.S. Nadirashvili, *Wandering solutions of the two-dimensional Euler equation (Russian),* Funktsional. Anal. i Prilozhen. **25** (1991), 70–71; translation in Funct. Anal. Appl. **25** (1991), 220–221 (1992)

[107] G. Namah, J.-M. Roquejoffre, *Remarks on the time-asymptotic behaviour of the solutions of Hamilton-Jacobi equations*, Comm. Partial Diff. Eq., **24**, 1999, 883–889.

[108] J. Nash, Continuity of solutions of parabolic and elliptic equations, Amer. Jour. Math., **80**, 1958, 931-954.

[109] J.R. Norris, Long-time behaviour of heat flow: global estimates and exact asymptotics. Arch. Rational Mech. Anal. **140**, 1997, 161–195

[110] R. Nussbaum and Y. Pinchover, On variational principles for the generalized principal eigenvalue of second order elliptic operators and some applications, J. Anal. Math. **59**, 1992, 161-177

[111] G. Namah and J.-M. Roquejoffre, Convergence to periodic fronts in semilinear parabolic equations, Nonlinear Diff. Eq. Appl., **4**, 1997, 521–536.

[112] H. Oertel (Editor), *Prandtl's Essentials of Fluid Mechanics,* Springer-Verlag, New York, 2004

[113] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer, 2010

[114] K. Ohkitani and M. Yamada, *Inviscid and inviscid-limit behavior of a surface quasi-geostrophic flow,* Phys. Fluids, **9** (1997), 876–882

[115] G. Pavliotis and A. Stuart, *Multiscale Methods: Averaging and Homogenization*, Springer, 2010.

[116] J. Pedlosky, *Geophysical Fluid Dynamics,* Springer, New York, 1987

[117] Dynamics of curved fronts, P. Pelcé, Ed., Academic Press, 1988

[118] Y. Pinchover and J. Rubinstein, *An Introduction to Partial Differential Equations*, Cambridge University Press, 2005.

[119] M. Pierre, E. Rouy, *A tridimensional inverse shaping problem,* Communications in PDE, **21**, 1996, 1279–1305

[120] G. Polya, On the zeros of an integral function represented by Fourier's integral, Messenger of Math. **52**, 1923, 185-188

[121] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. I. Functional Analysis,* Academic Press, New York, 1972

[122] E. Risler, *Global convergence toward traveling fronts in nonlinear parabolic systems with a gradient structure* Ann. Inst. H. Poincaré Anal. Non Linaire, **25** (2008), 381–424

[123] R.T. Rockafellar, *Convex analysis,* Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997

[124] J.-M. Roquejoffre, Eventual monotonicity and convergence to travelling fronts for the solutions of parabolic equations in cylinders. Ann. Inst. H. Poincaré Anal. Non Linéaire, **14**, 1997, 499–552

[125] J.-M. Roquejoffre, *Convergence to steady states or periodic solutions in a class of Hamilton-Jacobi equations*, J. Math. Pures Appl., **80**, 2001, 85–104

[126] O. Savin, Regularity of flat level sets in phase transitions, Ann. of Math. **169**, 2009, 41–78

[127] E. Stein, *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals,* Princeton University Press, 1993

[128] E. Stein, *Singular Integrals and Differentiability Properties of Functions,* Princeton University Press, 1970

[129] S. Sternberg, *On differential equations on the torus,* Amer. J. Math., **79**, 1957, 397–402

[130] M. Taylor, *Partial Differential Equations. II. Qualitative Studies of Linear Equations,* Applied Mathematical Sciences **116**, Springer-Verlag, New York 1996

[131] R. Temam, *Navier-Stokes Equations: Theory and Numerical Analysis,* AMS Chelsea Pub., 2001

[132] C.E. Wayne, *Vortices and two-dimensional fluid motion,* Notices of the AMS, **58**, no. 1, 10–19

[133] H. Weinberger, On spreading speeds and traveling waves for growth and migration models in a periodic habitat. J. Math. Biol. **45**, 2002, 511–548. Erratum: J. Math. Biol. **46**, 2003, p. 190

[134] W. Wolibner, *Un theorème sur l'existence du mouvement plan d'un fluide parfait, homogène, incompressible, pendant un temps infiniment long. (French)* Mat. Z., **37** (1933), 698–726

[135] J. Xin, Analysis and modelling of front propagation in heterogeneous media, SIAM Rev., **42**, 2000, 161-230

[136] J. Xin, *An Introduction to Fronts in Random Media*, Springer, 2009

[137] V.I. Yudovich, *The loss of smoothness of the solutions of the Euler equation with time (Russian),* Dinamika Sploshn. Sredy **16** Nestacionarnye Probelmy Gidordinamiki (1974), 71–78

[138] V.I. Yudovich, *On the loss of smoothness of the soltions to the Euler equation and the inhereht instability of flows of an ideal fluid,* Chaos **10** (2000), 705–719

[139] V. I. Yudovich, *Non-stationary flows of an ideal incompressible fluid,* Zh Vych Mat, **3** (1963), 1032–1066

[140] A. Zlatos, *Exponential growth of the vorticity gradient for the Euler equation on the torus,* preprint arXiv:1310.6128

[141] Y. Achdou, I. Capuzzo Dolcetta, *Mean field games: numerical methods*, SIAM J. Numer. Anal. 48 (2010), 1136–1162.

[142] Y.Achdou, F.Camilli, I.Capuzzo Dolcetta, *Mean field games: numerical methods for the planning problem*, SIAM J. Control Opt. 50 (2012), 77–109.

[143] Y.Achdou, F.Camilli, I.Capuzzo Dolcetta, *Mean field games: convergence of a finite difference method*. SIAM J. ournal on Numerical Analysis, 2013, Vol. 51, No. 5, 2585—2612

[144] Achdou Y. *Mean field games: additional notes*. Unpublished notes. php.math.unifi.it/users/cime/.../201141-Notes.pdf

[145] Achdou Y., Perez, V. *Iterative strategies for solving linearized discrete mean field games systems.* Netw. Heterog. Media 7 (2012), no. 2, 197-217.

[146] Achdou Y., Lauriere M. *On the system of partial differential equations arising in mean field type control*. arXiv:1503.05044

[147] Achdou, Y., Buera, F. J., Lasry, J. M., Lions, P. L., and Moll, B. (2014). *Partial differential equation models in macroeconomics.* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 372(2028), 20130397.

[148] Achdou, Y., Han, J., Lasry, J. M., Lions, P. L., and Moll, B. (2014). *Heterogeneous agent models in continuous time.* Preprint.

[149] Ambrose, D. M. (2016). Small strong solutions for time-dependent mean field games with local coupling. Comptes Rendus Mathematique, 354(6), 589-594.

[150] Ambrosio, L., *Transport equation and Cauchy problem for non-smooth vector fields.* Calculus of variations and nonlinear partial differential equations, 1–41, Lecture Notes in Math., 1927, Springer, Berlin, 2008.

[151] Ambrosio, L., *Transport equation and Cauchy problem for BV vector fields.* Inv. Math. 158 (2004), 227-260.

[152] Ambrosio, L., Gigli, N., Savar, G. GRADIENT FLOWS IN METRIC SPACES AND IN THE SPACE OF PROBABILITY MEASURES. Second edition. Lectures in Mathematics ETH Zrich. Birkhuser Verlag, Basel, 2008.

[153] Andersson, D., Djehiche, B. (2011). *A maximum principle for SDEs of mean-field type.* Applied Mathematics & Optimization, 63(3), 341-356.

[154] Aumann R. *Markets with a continuum of traders.* Econometrica, 32(1/2), 1964.

[155] Bardi, M. (2012). *Explicit solutions of some linear-quadratic mean field games.* Networks and Heterogeneous Media, 7(2), 243-261.

[156] Bardi, M., Feleqi, E. (2012). *Nonlinear elliptic systems and mean field games.* Forthcoming.

[157] Bardi M. & Capuzzo Dolcetta I. (1996) OPTIMAL CONTROL AND VISCOSITY SOLUTIONS OF HAMILTON-JACOBI-BELLMAN EQUATIONS. Birkhäuser.

[158] Barles G. (1994) SOLUTIONS DE VISCOSITÉ DES ÉQUATIONS DE HAMILTON-JACOBI. Springer-Verlag, Berlin.

[159] A. Blanchet, G. Carlier, *Optimal transport and Cournot-Nash equilibria*, (2015) Mathematics of Operations Research.

[160] Brenier, Y. *Dcomposition polaire et rarrangement monotone des champs de vecteurs.* C. R. Acad. Sci. Paris Sr. I Math. 305, 19 (1987), 805–808.

[161] Bensoussan, A., Frehse, J. (2002) REGULARITY RESULTS FOR NONLINEAR ELLIPTIC SYSTEMS AND APPLICATIONS. Applied Mathematical Sciences, 151. Springer-Verlag, Berlin.

[162] Bensoussan, A., Frehse, J. (2002). *Smooth solutions of systems of quasilinear parabolic equations.* ESAIM: Control, Optimisation and Calculus of Variations, 8, 169-193.

[163] A. Bensoussan, K. C. J. Sung, S. C. P. Yam, and S. P. Yung *Linear-Quadratic Mean Field Games*, preprint.

[164] Bensoussan, A., Frehse, J. (2012), *Control and Nash Games with Mean Field effect*, Chinese Annals of Mathematics, Series B, 34(2), 161-192.

[165] Bensoussan, A., Frehse J., Yam, (2013). *Mean field games and mean field type control theory.* New York: Springer.

[166] Brenier, Y. *Polar factorization and monotone rearrangement of vector-valued functions.* Comm. Pure Appl. Math. 44, 4 (1991), 375–417.

[167] Buckdahn, R., Li, J., Peng, S. *Mean-field backward stochastic differential equations and related partial differential equations.* Stochastic Process. Appl. 119 (2009), no. 10, 3133–3154.

[168] Buckdahn, R., Djehiche, B., Li, J., Peng, S. *Mean-field backward stochastic differential equations: a limit approach.* Ann. Probab. 37 (2009), no. 4, 1524–1565.

[169] Buckdahn, R., Li, J., Peng, S., & Rainer, C. (2014). Mean-field stochastic differential equations and associated PDEs. arXiv preprint arXiv:1407.1215.

[170] Burger, M., Di Francesco, M., Markowich, P., and Wolfram, M. T. (2013). *Mean field games with nonlinear mobilities in pedestrian dynamics.* arXiv preprint arXiv:1304.5201.

[171] L.A. Caffarelli, P.A. Markowich, M.T. Wolfram, *On a price formation free boundary value model by Lasry & Lions: The Neumann problem*, C. R. Acad. Sci. Paris, 349, 841-844, 2011.

[172] Camilli, F., & Silva, F. (2012). A semi-discrete in time approximation for a model 1st order-finite horizon mean field game problem. Networks and heterogeneous media, 7(2), 263-277.

[173] Cannarsa P. and Sinestrari C. SEMICONCAVE FUNCTIONS, HAMILTON-JACOBI EQUATIONS, AND OPTIMAL CONTROL. Progress in Nonlinear Differential Equations and their Applications, 58. Birkhäuser Boston Inc., Boston, MA, 2004.

[174] Cardaliaguet P., Lehalle C.-A. *Mean Field Game of Controls and an Application To Trade Crowding.* Pre-print.

[175] Cardaliaguet, P., Graber, J., Porretta, A., & Tonon, D. (2014). Second order mean field games with degenerate diffusion and local coupling. arXiv preprint arXiv:1407.7024.

[176] Cardaliaguet, P., & Graber, P. J. (2014). Mean field games systems of first order. arXiv preprint arXiv:1401.1789.

[177] P. Cardaliaguet, *Long time average of first order mean field games and weak KAM theory.* Preprint 2013.

[178] P. Cardaliaguet, *Weak solutions for first order mean field games with local coupling.* Preprint 2013.

[179] P. Cardaliaguet, G. Carlier, B. Nazaret, *Geodesics for a class of distances in the space of probability measures.* To appear in Calc. Var. and PDE.

[180] P. Cardaliaguet, J.-M. Lasry, P.-L. Lions, A. Porretta, *Long time average of mean field games*, Networks and Heterogeneous Media, 7 (2012), 2, 279–301.

[181] P. Cardaliaguet, J.-M. Lasry, P.-L. Lions, A. Porretta, *Long time average of mean field games with nonlocal coupling*, Preprint 2012.

[182] P. Cardaliaguet. *Notes on mean field games (from Lions' lectures at the Collge de France).* Unpublished notes. www.ceremade.dauphine.fr/ cardalia/MFG100629.pdf

[183] R. Carmona and F. Delarue. *Optimal control of McKean-Vlasov stochastic dynamics.* Technical report, 2012.

[184] R. Carmona and F. Delarue, (2013) *Probabilist analysis of Mean-Field Games.* SIAM Journal on Control and Optimization, 51(4), 2705-2734.

[185] R. Carmona, F. Delarue and A. Lachapelle *Control of McKean-Vlasov Dynamics versus Mean Field Games.* Mathematics and Financial Economics, 7(2), 131-166.

[186] Carmona, R., Delarue, F., and Lacker, D. (2014). Mean field games with common noise. arXiv preprint arXiv:1407.6181.

[187] M. Cirant. PhD Thesis, to appear.

[188] Cirant, M. (2014). *Multi-population Mean Field Games systems with Neumann boundary conditions.* To appear in Journal de Mathmatiques Pures et Appliques.

[189] P. Degond, J.-G. Liu, C. Ringhofer, *Large-scale dynamics of Mean-Field Games driven by local Nash equilibria*, Journal of Nonlinear Science (to appear).

[190] P. Degond, J.-G. Liu, C. Ringhofer, *Evolution of the distribution of wealth in an economic environment driven by local Nash equilibria*, preprint.

[191] Chassagneux, J. F., Crisan, D., & Delarue, F. (2014). Classical solutions to the master equation for large population equilibria. arXiv preprint arXiv:1411.3009.

[192] Ekeland, I., and Tmam, R. Convex analysis and variational problems, english ed., vol. 28 of Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999. Translated from the French.

[193] Clarke F. H. Optimization and nonsmooth analysis. Second edition. Classics in Applied Mathematics, 5. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.

[194] Di Perna R. J., Lions P.-L. *Ordinary differential equations, transport theory and Sobolev spaces.* Inv. Math. 98 (1989), 511-547.

[195] Fan, K. (1952). *Fixed-point and Minimax Theorems in Locally Convex Topological Linear Spaces.* Proc. Nat. Acad. Sci. U.S.A. 38(2) (1952), 121-126.

[196] Feleqi H. *The derivation of ergodic mean field game equations for several populations of players.* To appear in Dynamic Games and Applications.

[197] Fischer M. (2014) On the connection between symmetric N-player games and mean field games. Preprint.

[198] Fleming W.H., Soner H.M. (1993) Controlled Markov processes and viscosity solution. Springer-Verlag, New-York.

[199] Friedman A. (1971) Differential Games. Wiley, New York.

[200] Gangbo, W., Nguyen, T., Tudorascu, A. *Hamilton-Jacobi equations in the Wasserstein space.* Methods Appl. Anal. 15 (2008), no. 2, 155–183.

[201] Gangbo, W., Swiech A. (2014) *Metric viscosity solutions of Hamilton-Jacobi equations depending on local slopes.* Preprint.

[202] Gangbo, W., Swiech A. (2014) *Existence of a solution to an equation arising from the theory of mean field games.* Preprint.

[203] Gomes D.A., Mohr J. and Souza R.R. *Discrete time, finite state space mean field games.* Journal de Mathmatiques Pures et Appliques, 93(3) (2010), 308-328.

[204] Gomes, D. A., Mohr, J., Souza, R. R. (2010). *Mean field limit of a continuous time finite state game.* arXiv preprint arXiv:1011.2918.

[205] Gomes, D. A., Mohr, J., Souza, R. R. (2013). *Continuous time finite state mean field games.* Applied Mathematics & Optimization, 68(1), 99-143.

[206] D. Gomes, G. Pires and H. Sanchez-Morgado. *A-priori estimates for stationary mean-field games.* To appear in Networked and Heterogeneous Media.

[207] Gomes, D. A., Patrizi, S., Voskanyan, V. (2013). *On the existence of classical solutions for stationary extended mean field games.* Nonlinear Analysis: Theory, Methods & Applications, 99, 49-79.

[208] Gomes, D. A., Voskanyan, V. K. (2013). *Extended mean field games-formulation, existence, uniqueness and examples.* arXiv preprint arXiv:1305.2600.

[209] Gomes, D. A., Pimentel, E., Sánchez-Morgado, H. (2015) *Time dependent mean-field games in the subquadratic case.* Communications in Partial Differential Equations, 40(1), 40-76.

[210]

[211] Gomes, D. A., Pimentel, E., Snchez-Morgado, H. (2016). Time-dependent mean-field games in the superquadratic case. ESAIM: Control, Optimisation and Calculus of Variations, 22(2), 562-580.

[212] Gomes and Saude. *Mean field games models - a brief survey.* Dynamic Games and Applications 4(2), 2014, 110–154.

[213] Gomes, D. A., Pimentel, E. *Time dependent mean-field games with logarithmic nonlinearities.* arXiv preprint arXiv:1407.1261, 2014.

[214] Gomes, D. A., Mitake H. *Existence for stationary mean field games with quadratic hamiltonians with congestion.*

[215] Graber, P. J. (2014). *Optimal Control of First-Order Hamilton-Jacobi Equations with Linearly Bounded Hamiltonian.* Applied Mathematics & Optimization, 1-40.

[216] Guant O. *Mean field games and applications to economics.* PhD thesis, Universit Paris-Dauphine, 2009.

[217] Guant, O. *A reference case for mean field games models.* J. Math. Pures Appl. (9) 92 (2009), no. 3, 276–294.

[218] Guant, O., Lions, P.-L., Lasry, J.-M. *Mean Field Games and Applications.* Paris-Princeton Lectures on Mathematical Finance 2010. Tankov, Peter; Lions, Pierre-Louis; Laurent, Jean-Paul; Lasry, Jean-Michel; Jeanblanc, Monique; Hobson, David; Guant, Olivier; Crpey, Stphane; Cousin, Areski. Springer. Berlin. 2011. pages 205-266

[219] Guant, O. (2011). *Existence and uniqueness result for mean field games with congestion effect on graphs.* Applied Mathematics & Optimization, 1-13.

[220] Guant, O. *From infinity to one: The reduction of some mean field games to a global control problem.* Unpublished notes.

[221] Guant, O. (2013). *Mean Field Games with a Quadratic Hamiltonian: A Constructive Scheme.* In Advances in Dynamic Games (pp. 229-241). Birkhuser Boston.

[222] Guant, O. *A note on the uniqueness result for mean field games with congestion (after Pierre-Louis Lions).* Unpublished notes.

[223] Hewitt, E. and Savage, L. J. (1955). *Symmetric measures on Cartesian products.* Transactions of the American Mathematical Society, 80:470-501.

[224] J. Horowitz and R.L. Karandikar. Mean rates of convergence of empirical measures in the Wasserstein metric. Journal of Computational and Applied Mathematics, 55(3):261–273, 1994.

[225] M. Huang, *Large-population LQG games involving a major player: The Nash certainty equivalence principle*, SIAM J. Control Optim., 48 (2010), pp. 3318-3353.

[226] M. Huang, P. E. Caines, and R. P. Malham, *Individual and mass behaviour in large population stochastic wireless power control problems: centralized and Nash equilibrium solutions.* Proc. 42nd IEEE Conf. Decision Contr., Maui, Hawaii, pp. 98-103, Dec. 2003.

[227] M. Huang, P. E. Caines, and R. P. Malham, *Uplink power adjustment in wireless communication systems: a stochastic control analysis.* IEEE Trans. Automat. Contr., 49(2004), pp. 1693-1708.

[228] Huang, M., Malhamé, R.P. Caines, P.E. (2006). *Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle.* Communication in information and systems. Vol. 6, No. 3, pp. 221-252.

[229] Huang, M., Caines, P.E., Malhamé, R.P. (2007). *Large-Population Cost-Coupled LQG Problems With Nonuniform Agents: Individual-Mass Behavior and Decentralized ε-Nash Equilibria.* IEEE Transactions on Automatic Control, 52(9), p. 1560-1571.

[230] Huang, M., Caines, P.E., Malhamé, R.P. (2007). *The Nash Certainty Equivalence Principle and McKean-Vlasov Systems: an Invariance Principle and Entry Adaptation.* 46th IEEE Conference on Decision and Control, p. 121-123.

[231] Huang, M., Caines, P.E., Malhamé, R.P. (2007). *An Invariance Principle in Large Population Stochastic Dynamic Games.* Journal of Systems Science & Complexity, 20(2), p. 162-172.

[232] Huang, M., Caines, P. E., Malham, R. P. (2010). *The NCE (mean field) principle with locality dependent cost interactions.* Automatic Control, IEEE Transactions on, 55(12), 2799-2805.

[233] Huang, M., Caines, P. E., Malham, R. P. (2012). *Social optima in mean field LQG control: centralized and decentralized strategies.* Automatic Control, IEEE Transactions on, 57(7), 1736-1751.

[234] Kingman, J. F. C. *Uses of exchangeability.* Ann. Probability 6 (1978), no. 2, 183–197.

[235] Kolokoltsov, V. N., Li, J., Yang, W. (2011). *Mean field games and nonlinear Markov processes.* Preprint arXiv:1112.3744.

[236] Lachapelle A. *Human crowds and groups interactions: a mean field games approach.* Pre-print.

[237] Lachapelle A., Salomon J. and Turinici G. *Computation of mean field equilibria in economics.* Mathematical Models and Methods in Applied Sciences, Issue 0, Vol. 1, pp1-22, 2010.

[238] A. Lachapelle, M.T. Wolfram, *On a mean field game approach modeling congestion and aversion in pedestrian crowds*, Trans. Res.: Part B: Methodological, 45(10), 1572-1589, 2011

[239] Lacker, D. (2014). A general characterization of the mean field limit for stochastic differential games. arXiv preprint arXiv:1408.2708.

[240] Ladyženskaja O.A., Solonnikov V.A and Ural'ceva N.N LINEAR AND QUASILINEAR EQUATIONS OF PARABOLIC TYPE. Translations of Mathematical Monographs, Vol. 23 American Mathematical Society, Providence, R.I. 1967

[241] Laguzet, L. and Turinici, G. (2015). *Individual vaccination as Nash equilibrium in a SIR model: the interplay between individual optimization and societal policies.* Preprint HAL hal-01100579.

[242] Lasry, J.-M., Lions, P.-L. *Large investor trading impacts on volatility.* Ann. Inst. H. Poincar Anal. Non Linaire 24 (2007), no. 2, 311–323.

[243] Lasry, J.-M., Lions, P.-L. *Mean field games.* Jpn. J. Math. 2 (2007), no. 1, 229–260.

[244] Lasry, J.-M., Lions, P.-L. *Jeux champ moyen. II. Horizon fini et contrle optimal.* C. R. Math. Acad. Sci. Paris 343 (2006), no. 10, 679–684.

[245] Lasry, J.-M., Lions, P.-L. *Jeux champ moyen. I. Le cas stationnaire.* C. R. Math. Acad. Sci. Paris 343 (2006), no. 9, 619–625.

[246] Lasry, J.-M., Lions, P.-L., Guant O. *Application of Mean Field Games to Growth Theory.* Preprint, 2009.

[247] Lieberman, G. M. SECOND ORDER PARABOLIC DIFFERENTIAL EQUATIONS. World Scientific, 1996.

[248] Lions, P.-L. *Cours au Collège de France*. www.college-de-france.fr.

[249] Lucas Jr, R. E., Moll, B. (2011). *Knowledge Growth and the Allocation of Time* (No. w17495). National Bureau of Economic Research.

[250] P.A. Markowich, N. Matevosyan, J.F. Pietschmann, M.T. Wolfram, *On a parabolic free boundary value problem modeling price formation*, Math. Models Methods Appl. Sci., 11(19), 1929-1957, 2009.

[251] A. Mas-Colell. On a theorem of Schmeidler. J. Math. Econ., 3:201-206, 1984.

[252] A. Meszaros and F. Silva. (2015) *A variational approach to second order mean field games with density constraints: the stationary case.* ArXiv:1502.06026.

[253] D. Monderer and L.S. Shapley. *Potential games.* Games and Economic Behavior, 14:124-143, 1996.

[254] S. L. Nguyen and M. Huang. *Linear-quadratic-gaussian mixed games with continuum parametrized minor players.* SIAM J. Control Optim., 50(5):2907-2937, 2012.

[255] Nourian, M., Caines, P., Malhame, R., Huang, M. (2011). *Mean field LQG control in leader-follower stochastic multi-agent systems: Likelihood ratio based adaptation.* IEEE Trans. Automat. Control 57 (2012), no. 11, 2801-2816.

[256] Porretta, A. (2013). *On the Planning Problem for the Mean Field Games System.* Dynamic Games and Applications, 1-26.

[257] Porretta, A. (2013). *Weak Solutions to Fokker-Planck Equations and Mean Field Games.* Archive for Rational Mechanics and Analysis, 1-62.

[258] S.T. Rachev and L. Rschendorf. MASS TRANSPORTATION PROBLEMS. Vol. I: Theory; Vol. II : Applications. Springer-Verlag,, 1998.

[259] Santambrogio, F. (2011). *A Modest Proposal for MFG with Density Constraints.* Net. Het. Media, Volume 7, Issue 2, 2012 Pages 337-347.

[260] D. Schmeidler. Equilibrium points of nonatomic games. J. Stat. Phys., 7:295-300, 1973.

[261] Tembine, H., Zhu, Q., Basar, T. (2011). *Risk-sensitive mean-field stochastic differential games.* In Proc. 18th IFAC World Congress.

[262] Sznitman, A.-S. TOPICS IN PROPAGATION OF CHAOS. Cours de l'Ecole d't de Saint-Flour, Lecture notes in mathematics vol. 1464, Springer (1989).

[263] Villani C. TOPICS IN OPTIMAL TRANSPORTATION. vol. 58 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2003.

[264] Villani, C. OPTIMAL TRANSPORT : OLD AND NEW, Springer, Berlin, (2009)

[265] Yong, J., and Zhou, X. Y. (1999). STOCHASTIC CONTROLS: HAMILTONIAN SYSTEMS AND HJB EQUATIONS (Vol. 43). Springer Science & Business Media.