The convergence problem in mean field games with a local coupling

Pierre Cardaliaguet

The convergence problem in mean field games with a local coupling

P. Cardaliaguet*

October 19, 2016

Abstract

The paper studies the convergence, as N tends to infinity, of a system of N coupled Hamilton-Jacobi equations (the Nash system) when the coupling between the players becomes increasingly singular. The limit equation is a mean field game system with local coupling.

In this paper we investigate the convergence of the Nash system associated with a differential game to the mean field game (MFG) system as the number of players tend to infinity. In differential game theory, the Nash system associated with a $N-$player differential game is a coupled system of N Hamilton-Jacobi equations. In our previous work [3], co-authored with Delarue, Lasry and Lions, we explained that the solution of the Nash system converges, as N tends to infinity, to the solution of the MFG system, which consists in a coupling between an Hamilton-Jacobi equation and a Fokker-Planck equation. We proved the result under the key assumption that the "coupling" between the equations is nonlocal and regularizing. In the present setting, we consider the case where this coupling is singular: in the Nash system, the payoffs of a player depends in a increasingly singular way on the players which are very close to him. We prove that, in this case, the solution of the Nash system converges to a solution of the Nash system with a local coupling.

To better explain what we have in mind, let us consider the Nash system

$$\begin{cases} -\partial_t v^{N,i}(t,\boldsymbol{x}) - \displaystyle\sum_{j=1}^N \Delta_{x_j} v^{N,i}(t,\boldsymbol{x}) + H(x_i, D_{x_i} v^{N,i}(t,\boldsymbol{x})) \\ \qquad\qquad + \displaystyle\sum_{j\neq i} D_p H(x_j, D_{x_j} v^{N,j}(t,\boldsymbol{x}))\cdot D_{x_j} v^{N,i}(t,\boldsymbol{x}) = F^{N,i}(\boldsymbol{x}) \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{in } [0,T]\times(\mathbb{R}^d)^N, \\ v^{N,i}(T,\boldsymbol{x}) = G(x_i) \qquad \text{in } (\mathbb{R}^d)^N. \end{cases} \tag{1}$$

In the above system, the N unknown maps $v^{N,i}$ depend on time and space in the form $(t,\boldsymbol{x})$ with $\boldsymbol{x}=(x_1,\dots,x_N)\in(\mathbb{R}^d)^N$. The data are the horizon T, the Hamiltonian $H:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$, the terminal condition $G:\mathbb{R}^d\to\mathbb{R}$ and the maps $F^{N,i}:(\mathbb{R}^d)^N\to\mathbb{R}$. The maps $F^{N,i}$ are called the coupling because they are responsible of all the interactions between the equations.

We are also interested in the associated system of N coupled stochastic differential equations (SDE):

$$dY_{i,t} = -D_p H\big(Y_{i,t}, Dv^{N,i}(t,\boldsymbol{Y}_t)\big)dt + \sqrt{2}dB^i_t, \qquad t\in[0,T],\ i\in\{1,\dots,N\}, \tag{2}$$

*Université Paris-Dauphine, PSL Research University, CNRS, Ceremade, 75016 Paris, France. cardaliaguet@ceremade.dauphine.fr

where $\boldsymbol{Y}_t = (Y_{1,t}, \cdots, Y_{N,t})$, $(v^{N,i})$ being the solution to (1) and the $((B^i_t)_{t\in[0,T]})_{i=1,\dots,N}$ being $d-$dimensional independent Brownian motions. In the language of differential games, the map $v^{N,i}$ is the value function associated with player $i \in \{1, \dots, N\}$ while $(Y_{i,t})$ is his optimal trajectory.

In order to expect a limit system, we suppose that the coupling map $F^{N,i}$ enjoys a symmetry property:

$$F^{N,i}(\boldsymbol{x}) = F^N(x_i, m^{N,i}_{\boldsymbol{x}})$$

where $F^N : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ is a given map and $m^{N,i}_{\boldsymbol{x}} = \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j}$ is the empirical measure of all players but i, $\mathcal{P}(\mathbb{R}^d)$ denoting the space of Borel probability measures on $\mathbb{R}^d$. Note that this assumption means that the players are indistinguishable: for a generic player i, players k and l (for $k \neq i$ and $l \neq i$) play the same role. Moreover, all the players have a cost function with the same structure. This key conditions ensures that the Nash system enjoys strong symmetry properties.

In contrast with [3], where F^N does not depend on N and is regularizing with respect to the measure, we assume here that the (F^N) become increasingly singular as $N \to +\infty$. Namely we suppose that there exists a smooth (local) maps $F : \mathbb{R}^d \times [0, +\infty) \to \mathbb{R}$ such that

$$\lim_{N\to+\infty} F^N(x_i, m\mathrm{dx}) = F(x_i, m(x_i)), \tag{3}$$

for any sufficiently smooth probability density $m\mathrm{dx} = m(x)\mathrm{dx}$. This assumption, which is the main difference with [3], is very natural in the context of mean field games. One expects (and we will actually prove) that the limit system is a MFG system with local interactions:

$$\begin{cases} -\partial_t u - \Delta u + H(x, Du) = F(x, m(t,x)) & \text{in } [0,T]\times\mathbb{R}^d, \\ \partial_t m - \Delta m - \mathrm{div}(mD_pH(x,Du)) = 0 & \text{in } [0,T]\times\mathbb{R}^d, \\ u(T,x) = G(x),\ m(0,\cdot) = m_0 & \text{in } \mathbb{R}^d \end{cases} \tag{4}$$

This system—which enjoys very nice properties—has been very much studied in the literature: see [9, 10, 20, 21, 22] and the references therein. The typical assumptions ensuring the MFG system to be well-posed are that F is monotone with respect to its second variable while H is convex in its second variable (plus growth conditions on F and H).

To explain in what extend the local couplings differs from the nonlocal ones, let us recall the ideas of proof in this later setting. The main difficulty is the complete lack of estimate independent of N for the solution of the Nash system (note that in our local setting, things are even worse since terms like $F^N(x_i, m^{N,i}_{\boldsymbol{x}})$ blow up unless $m^{N,i}_{\boldsymbol{x}}$ is very close to a regular density). To overcome this problem, the main ingredient in [3] is the existence and uniqueness of a classical solution to the so-called *master equation*. When $F^{N,i}(\boldsymbol{x}) = F(x_i, m^{N,i}_{\boldsymbol{x}})$, where $F : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ is sufficiently smooth, the master equation takes the form of a transport equation stated on the space of probability measures:

$$\begin{cases} -\partial_t U - \Delta_x U + H(x, D_x U) \\ \qquad - \displaystyle\int_{\mathbb{R}^d} \mathrm{div}_y\left[D_m U\right]\, dm(y) + \int_{\mathbb{R}^d} D_m U \cdot D_p H(y, D_x U)\, dm(y) = F(x,m) \\ \qquad\qquad \text{in } [0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d) \\ U(T,x,m) = G(x) \qquad \text{in } \mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d) \end{cases} \tag{5}$$

In the above equation, the unknown U is a scalar function depending on $(t,x,m) \in [0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)$, $D_m U$ denotes the derivative with respect to the measure m (see [3]). The interest of the

map U is that the $N-$tuple $(u^{N,i})$ defined by

$$u^{N,i}(t,\boldsymbol{x}) = U(t,x_i,m^{N,i}_{\boldsymbol{x}}), \qquad \boldsymbol{x}=(x_1,\dots,x_N)\in(\mathbb{R}^d)^N, \tag{6}$$

is an approximate solution to the Nash system (1) which enjoys very good regularity properties. Then the whole point in the proof in [3] consists in transferring the regularity of $u^{N,i}$ to $v^{N,i}$ by integrating both functions along the same (and carefully chosen) path.

When F is a local couplings, i.e., $F(x,m)=F(x,m(x))$ for any absolutely continuous measure m, the meaning of the master equation is not clear: obviously one cannot expect U to be a smooth solution to (5), if only because the coupling blows up at singular measures. As a consequence, the definition of the maps $u^{N,i}$ through (6) is dubious and, even if such a definition could make sense, there is no hope that the $u^{N,i}$ satisfy the regularity properties required in the computation of [3].

As the master equation for the limit problem does not seem to make much sense, we present here a completely different approach for the proof of the convergence (see Theorem 3.1 and its proof): it consists in comparing directly the solution of the Nash system to the solution of the MFG system *without using the master equation*. Although the argument are presented here for a local coupling, they easily adapt to the framework of [3], thus providing a simpler and more direct proof of some convergence results in [3]. Let us point out, however, that we do not recover all the convergence results in [3] and that the convergence rate is also sharper in [3]. In order to compare directly the solution of the Nash system $v^{N,i}$ and the u component of the MFG system (or, actually, a variant of this system), we build *different* and well chosen paths along which these functions behave in a same way. Then we overcome the difficulty that the paths are different (as well as the lack of estimate for $v^{N,i}$) by using the structure of the equation (convexity of the Hamiltonian and monotonicity of the map F), somehow reproducing the uniqueness argument for the MFG system [20] at the level of the difference $v^{N,i}-u$.

We now give a flavor of our result in a particular case. Let us assume that

$$F^N(x,m) = F(\cdot,\xi^{\epsilon_N}\star m(\cdot))\star\xi^{\epsilon_N},$$

where $F=F(x,m)$ is globally Lipschitz continuous and increasing in m, $\epsilon_N := N^{-\beta}$ (for some $\beta>0$), $\xi^\epsilon(x)=\epsilon^{-d}\xi(x/\epsilon)$, ξ being a symmetric smooth nonnegative kernel with compact support. In order to describe the convergence of the solution $v^{N,i}=v^{N,i}(t,x_1,\dots,x_N)$ of the Nash system (15) to the solution $u=u(t,x)$ of the MFG system (4), we reduce the function $v^{N,i}$ to a function of a single variable by averaging it against the measure m_0: for $i\in\{1,\dots,N\}$, let

$$w^{N,i}(t_0,x_i,m_0) := \int\dots\int v^{N,i}(t_0,\boldsymbol{x})\prod_{j\neq i} m_0(dx_j) \qquad \text{where } \boldsymbol{x}=(x_1,\dots,x_N).$$

If β is not too large (namely $\beta\in(0,(3d(d+1))^{-1})$), then

$$\left\|w^{N,i}(t_0,\cdot,m_0)-u(t_0,\cdot)\right\|_{L^1(m_0)}\leqslant CN^{-\gamma},$$

where the constants C and $\gamma\in(0,1)$ depend on the regularity of the initial measure m_0 and β (see Corollary 3.6). Moreover, we show that the optimal trajectories $(Y_{i,t})$ converges to the optimal trajectory $(\tilde{X}_{i,t})$ associated with the limit MFG system and establish a propagation of

chaos property.

For a general sequence of couplings (F^N), converging to a local coupling F as in (3), our main result (Theorem 3.4) states that, if the Lipschitz regularity of F^N does not deteriorate too fast, namely $Lip(F^N) = o(N^{-1/(3d)})$ in space dimension $d \geqslant 3$, then $w^{N,i}$ converges to u and the optimal trajectories converge as well. The space dimension arises here through the convergence rate in the law of large number [1, 6]. Let us point out that, in order to avoid issue related to boundary conditions or problems at infinity, we will assume that the data are periodic in space, thus working in the torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$.

Mean field game theory started with the pioneering works by Lasry and Lions [18, 19, 20] and Caines, Huang and Malhamé [11, 12, 13, 14]. These authors introduced the mean field game system and discussed its properties: in particular, Lasry and Lions introduced the fundamental monotonicity condition on the coupling functions. They also discussed the various types of MFG systems (with local or nonlocal coupling, with or without diffusion).

The link between the MFG system (which can be seen as a differential game with infinitely many players) and the differential games with finitely many players has been the object of several contributions. Caines, Huang and Malhamé [14], and, later on, Delarue and Carmona [4] and Kolokoltsov, Troeva and Yang [15], explained how to use the solution of the MFG system to build $\epsilon-$Nash equilibria in the $N-$player game (in open loop form). The convergence of the Nash system (1) remained a puzzling issue for some time. The first results in that direction go back to [18, 20], in the "ergodic case", where the Nash system becomes a system of N coupled equation in dimension d (not Nd as in our setting): then one can obtain estimates which allow to pass to the limit. Another particular case is obtained when one is interested in Nash equilibria in open loop form: Fischer [7] and Lacker [16] discussed in what extend one can expect to obtain the MFG system at the limit. For the genuine Nash system (1), a first breakthrough was achieved by Lasry and Lions (see the presentation in [21]) who formally explained the mechanism towards convergence under suitable *a priori* estimates on the solution. For this they introduced the master equation and described (again mostly formally) its main properties.

The rigorous derivation of Lasry and Lions ideas took some time. The existence of a classical solution to the master equation has been obtained by several authors in different frameworks (Buckdahn, Li, Peng and Rainer [2] for the linear master equation without coupling, Gangbo and Swiech [8] for the master equation without diffusion and in short time horizon, Chassagneux, Crisan and Delarue [5] for the "first order" master equation, Lions [21] for an approach by monotone operators). The most general result so far is obtained in [3], where the second order master equation is proved to be well-posed even for more complex problems with *common noise*. The first convergence result for the Nash system is also obtained in [3] as a consequence of the well-posedness of the master equation. Its holds for Nash systems with common noise as well. The convergence is expressed in two ways: by comparing the $v^{N,i}$ to the solution of the master equation, or, as we do so here, by comparing the averaged function $w^{N,i}$ to the solution of the MFG system. The present paper is the first attempt to show the convergence for a coupling which becomes singular. It also provides an alternative approach of the convergence—without using the master equation.

The paper is organized in the following way: we first state our main notation and assumptions. In section 2, we prove uniform regularity estimates on the solution (u^N, m^N) of the perturbed MFG system (4) in which the right-hand side is replaced by F^N. We also compare (u^N, m^N) with (u, m). Section 3 is the core of the paper. The key step is the comparison

between the solution of the Nash system $v^{N,i}$ and u^N along well-chosen paths (Theorem 3.1). Then we collect all the estimate to obtain a rate of convergence between the averaged function $w^{N,i}$ and u (Theorem 3.4).

Acknowledgement: The author was partially supported by the ANR (Agence Nationale de la Recherche) project ANR-16-CE40-0015-01.

1 Notation and Assumptions

1.1 Notation

We sake of simplicity, the paper is written under the assumption that all maps are periodic in space. So the underlying state space is the torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. This simplifying assumption allows to discard possible problems at infinity (or at the boundary of a domain). We denote by $|\cdot|$ the euclidean norm in $\mathbb{R}^d$ and—by abuse of notation—the corresponding distance in $\mathbb{T}^d$. The ball centered at $x \in \mathbb{T}^d$ and of radius R is denoted by $B_R(x)$.

For $k \in \mathbb{N}$ and $\alpha \in (0,1)$, we denote by $C^{k+\alpha}$ the set of maps $u = u(x)$ which are of class C^k and $D^k u$ is $\alpha-$Holder continuous. When $u = u(t,x)$ is time dependent and $\alpha \in (0,1)$, we say that u is in $C^{0,\alpha}$ if

$$\|u\|_{C^{0,\alpha}} := \|u\|_\infty + \sup_{(t,x),(t',x')} \frac{|u(t,x) - u(t',x')}{|x-x'|^\alpha + |t-t'|^{\alpha/2}} < +\infty.$$

We say that u is in $C^{1,\alpha}$ if u and Du belong to $C^{0,\alpha}$. Finally $C^{2,\alpha}$ consists in the maps u such that D^2u and $\partial_t u$ belong to $C^{0,\alpha}$. It is known that, if u is in $C^{2,\alpha}$, then u is also in $C^{1,\alpha}$.

We denote by $\mathcal{P}(\mathbb{T}^d)$ the set of Borel probability measures on the torus $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$. It is endowed with the Monge-Kantorovitch distance:

$$\mathbf{d}_1(m,m') = \sup_\phi \int_{\mathbb{T}^d} \phi(y)\ d(m-m')(y),$$

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

1.2 Assumption

Throughout the paper, we suppose that the following conditions are in force.

- The Hamiltonian $H : \mathbb{T}^d \times \mathbb{R}^d \to \mathbb{R}$ is smooth, H and $D_p H$ are globally Lipschitz continuous in both variables and H is locally uniformly convex with respect to the second variable:

$$D^2_{pp} H(x,p) > 0 \qquad \forall (x,p) \in \mathbb{T}^d \times \mathbb{R}^d. \tag{7}$$

- $F : \mathbb{T}^d \times [0,+\infty) \to \mathbb{R}$ is smooth, globally Lipschitz continuous in both variables and increasing with respect to the second variable with $\partial_m F \geqslant \delta > 0$ for some $\delta > 0$.

- The terminal cost $G : \mathbb{T}^d \to \mathbb{R}$ is a smooth map.

- For any $N \in \mathbb{N}$, $F^N : \mathbb{T}^d \times \mathcal{P}(\mathbb{T}^d) \to \mathbb{R}$ is monotone:

$$\int_{\mathbb{T}^d} (F^N(x,m) - F^N(x,m') d(m-m')(x) \geqslant 0 \qquad \forall m, m' \in \mathcal{P}(\mathbb{T}^d).$$

- (difference between F^N and F) For any $R > 0$ and $\alpha \in (0,1)$, there exists $k_N^{R,\alpha} \to 0$ as $N \to +\infty$ such that
$$\|F^N(\cdot, m\mathrm{dx}) - F(\cdot, m(\cdot))\|_\infty \leqslant k_N^{R,\alpha} \tag{8}$$
for any density m such that $\|m\|_{C^\alpha} \leqslant R$.

- (uniform regularity of F^N for regular densities) For any $R > 0$ and $\alpha \in (0,1)$, there exists $\kappa^{R,\alpha} > 0$ such that, for any $N \in \mathbb{N}$,
$$\left|F^N(x, mdx) - F^N(y, m'dx)\right| \leqslant \kappa_{R,\alpha}\left(|x-y|^\alpha + \|m - m'\|_\infty\right) \tag{9}$$
for any density m, m' with $\|m\|_{C^\alpha}, \|m'\|_{C^\alpha} \leqslant R$.

- (regularity assumptions on F^N for general densities) For any $N \in \mathbb{N}$, there exists a constant $K_N \geqslant 1$ such that
$$\left|F^N(x,m) - F^N(x,m')\right| \leqslant K_N \qquad \forall x \in \mathbb{T}^d,\ \forall m, m' \in \mathcal{P}(\mathbb{T}^d). \tag{10}$$

A few comments are in order. Note that $K_N \to +\infty$ as $N \to +\infty$ because $F^N(x,m)$ blows up if m is a singular measure. So F^N becomes closer and closer to F while its regularity at general probability measures deteriorates. However, assumption (9) states that the F^N are uniformly Holder continuous when evaluated at probability densities which are Holder continuous.

The monotonicity assumptions on F and F^N and the convexity of H are known to ensure the uniqueness of the solution in the MFG systems: they are therefore natural in our study. The global Lipschitz regularity assumption on H is not completely natural in the context of MFG, but we do not know how to avoid it: it is required at every key step of the paper. Let us just note that it simplifies at lot the existence of solutions for the Nash system (see, for instance, [17]) as well as for the limit MFG system (see [20]): indeed, without the assumption that D_pH is bounded, existence of classical solution to (4) is related on an subtle interplay between the growth of H and of F.

We explain in Remark 3.2 that the strong monotonicity condition on F can be avoided (F non decreasing suffices), but the convergence rates in Theorems 3.1 and 3.4 then deteriorates a little.

1.3 Main example

Given F satisfying the above conditions, a typical example for the regularization F^N is the following:

Proposition 1.1. *Assume that $F^N = F^{\epsilon_N}$ with*
$$F^\epsilon(x,m) := F(\cdot, \xi^\epsilon \star m(\cdot)) \star \xi^\epsilon(x) \tag{11}$$
where $\epsilon_N \to 0$ as $N \to +\infty$ and $\xi^\epsilon(x) = \epsilon^{-d}\xi(x/\epsilon)$, ξ being a symmetric smooth nonnegative kernel with compact support. Then, for any N, F^N is monotone and satisfies (9).

Moreover the constants $k_N^{R,\alpha}$ and K_N can be estimated by
$$k_N^{R,\alpha} \leqslant C(1+R)\epsilon_N^\alpha, \qquad K_N \leqslant C\epsilon_N^{-d-1}, \tag{12}$$
where C depends on the regularity of F and of ξ.

Proof. We prove the result for F^ϵ with the transparent notations $k_\epsilon^{R,\alpha}$ and K_ϵ. Under the monotonicity assumption on F, one easily checks that the F^ϵ are monotone. Next we prove that the F^ϵ satisfy (9). Let $mdx, m'dx \in \mathcal{P}(\mathbb{T}^d)$ with $\|m\|_{C^\alpha}, \|m'\|_{C^\alpha} \leqslant R$. Then

$$\begin{aligned}
&\left|F^\epsilon(x, mdx) - F^\epsilon(y, m'dx)\right| \\
&\quad \leqslant \sup_y \left|F(x-y, m \star \xi^\epsilon(x-y)) - F(x'-y, m' \star \xi^\epsilon(x'-y))\right| \\
&\quad \leqslant C \sup_y \left[|x-x'| + |m \star \xi^\epsilon(x-y) - m' \star \xi^\epsilon(x'-y)|\right] \\
&\quad \leqslant C \sup_y \left[|x-x'| + R|x-x'|^\alpha + |m \star \xi^\epsilon(x'-y) - m' \star \xi^\epsilon(x'-y)|\right] \\
&\quad \leqslant C(R+1) \sup_y \left[|x-x'|^\alpha + \|m-m'\|_\infty\right].
\end{aligned}$$

We now estimate the constants $k_\epsilon^{R,\alpha}$ defined by

$$k_\epsilon^{R,\alpha} := \sup_m \|F^\epsilon(\cdot, m\mathrm{dx}) - F(\cdot, m(\cdot))\|_\infty,$$

where the supremum is taken over the densities m such that $\|m\|_{C^\alpha} \leqslant R$. Recall that, if ϕ is Holder continuous with $\|\phi\|_{C^\alpha} \leqslant R$, then

$$|\xi^\epsilon \star \phi(x) - \phi(x)| \leqslant R\epsilon^\alpha \int_{\mathbb{R}^d} \xi(y)|y|^\alpha.$$

One easily derive from this that

$$k_\epsilon^{R,\alpha} \leqslant C(1+R)\epsilon^\alpha$$

where C depends on the Lipschitz constant of F and on ξ.

Finally, we have to estimate K_ϵ, which is the Lipschitz constant of F^ϵ with respect to the measure m. Following [3], we know that

$$K_\epsilon = \sup_{m \in \mathcal{P}(\mathbb{T}^d)} \left\| D_y \frac{\delta F^\epsilon}{\delta m}(\cdot, m, \cdot) \right\|_\infty$$

(we refer to [3] for the definition of the derivative $\frac{\delta F^\epsilon}{\delta m}$). As

$$\frac{\delta F^\epsilon}{\delta m}(x, m, y) = \int_{\mathbb{R}^d} \sum_{k \in \mathbb{Z}^d} \xi^\epsilon(z-y-k) \partial_m F(z, \xi^\epsilon \star m(z)) \xi^\epsilon(x-z) dz$$

we have

$$\left| D_y \frac{\delta F^\epsilon}{\delta m}(x, m, y) \right| \leqslant \|\partial_m F\|_\infty \int_{\mathbb{R}^d} \sum_{k \in \mathbb{Z}^d} |D\xi^\epsilon(z-y-k)| \; |\xi^\epsilon(x-z)| dz.$$

Assuming that the support of ξ is contained in the ball B_ρ, we have therefore

$$\left| D_y \frac{\delta F^\epsilon}{\delta m}(x, m, y) \right| \leqslant \|\partial_m F\|_\infty \sum_{k \in \mathbb{Z}^d} \|D\xi^\epsilon(\cdot - z - k)\|_{L^\infty(B_{\rho\epsilon}(x))} \|\xi^\epsilon(x - \cdot)\|_1 \leqslant C\epsilon^{-d-1}.$$

□

2 Regularity estimates

Let $(t_0, m_0) \in [0,T] \times \mathcal{P}(\mathbb{T}^d)$ and let (u^N, m^N) and (u, m) be respectively the unique solution to the MFG systems

$$\begin{cases} -\partial_t u^N - \Delta u^N + H(x, Du^N) = F^N(x, m(t)) & \text{in } [t_0, T] \times \mathbb{T}^d, \\ \partial_t m^N - \Delta m^N - \operatorname{div}(m^N D_p H(x, Du^N)) = 0 & \text{in } [t_0, T] \times \mathbb{T}^d, \\ u^N(T, x) = G(x),\ m^N(t_0, \cdot) = m_0 & \text{in } \mathbb{T}^d \end{cases} \tag{13}$$

and

$$\begin{cases} -\partial_t u - \Delta u + H(x, Du) = F(x, m(t, x)) & \text{in } [0, T] \times \mathbb{T}^d, \\ \partial_t m - \Delta m - \operatorname{div}(m D_p H(x, Du)) = 0 & \text{in } [0, T] \times \mathbb{T}^d, \\ u(T, x) = G(x),\ m(t_0, \cdot) = m_0 & \text{in } \mathbb{T}^d \end{cases} \tag{14}$$

Following [20, 21], these systems are known to be well-posed. The aim of this section is to establish estimates for (u^N, m^N) independent of N and to compare (u^N, m^N) with (u, m).

Proposition 2.1. *Assume that m_0 has a positive density of class C^2. Then the (u^N, m^N) are bounded in $C^{2,\alpha} \times C^{0,\alpha}$ independently of N.*

Proof. As H is globally Lipschitz continuous, the existence and uniqueness of a solution to (13) and to (14) is well-known: see [20].

We now check the regularity of m^N. As $D_p H$ is bounded and m_0 is smooth, standard argument for parabolic equation in divergence form (see, e.g., Theorem 3.1 of chap. V in [17]) state that the m^N are bounded in $C^{0,\alpha}$ for some $\alpha \in (0,1)$. Note that α and the bound depend on $\|D_p H\|_\infty$ and $\|m_0\|_{C^2}$ only.

We now plug this estimate into the parabolic equation for u^N. For this we note that the map $(t,x) \to F^N(x, m^N(t))$ is uniformly Holder continuous. Indeed, in view of assumption (9) and the uniform regularity of m^N,

$$\begin{aligned} \left|F^N(x, m^N(t)) - F^N(x', m^N(t'))\right| &\leqslant \kappa^{R,\alpha}\left(|x-y|^\alpha + \|m^N(t,\cdot) - m^N(t',\cdot)\|_\infty\right) \\ &\leqslant C\left(|x-y|^\alpha + |t-t'|^{\alpha/2}\right) \end{aligned}$$

where $R := \|m^N\|_{C^{0,\alpha}}$. Since G is $C^{2+\alpha}$ and is independent of m^N and since H is Lipschitz continuous, standard estimates on Hamilton-Jacobi equations imply that the u^N are bounded in $C^{2,\alpha}$. □

Proposition 2.2. *Assume that m_0 has a positive density of class C^2. Then*

$$\sup_{t\in[0,T]} \|u^N(t,\cdot) - u(t,\cdot)\|_{H^1(\mathbb{T}^d)} + \|m^N - m\|_{L^2} \leqslant C k_N^{R,\alpha},$$

$$\sup_{t\in[t_0,T]} \|Du^N(t,\cdot) - Du(t,\cdot)\|_\infty \leqslant C\left(k_N^{R,\alpha}\right)^{\frac{2}{d+2}}.$$

where α, R and C depend on on the data and m_0, but not on N.

Proof. By standard computation (see [20]), we have

$$\begin{aligned} &\left[\int_{\mathbb{T}^d} (u^N - u)(m^N - m)\right]_0^T \\ &\quad = -\int_0^T \int_{\mathbb{T}^d} m\left(H(x, Du^N) - H(x, Du) - D_p H(x, Du)\cdot D(u^N - u)\right) \\ &\qquad -\int_0^T \int_{\mathbb{T}^d} m^N\left(H(x, Du) - H(x, Du^N) - D_p H(x, Du^N)\cdot D(u - u^N)\right) \\ &\qquad -\int_0^T \int_{\mathbb{T}^d} (F^N(x, m^N(t)) - F(x, m(t,x))(m^N(t,x) - m(t,x)). \end{aligned}$$

Note, on the one hand, that $m^N(0) = m(0) = m_0$ and $u^N(T) = u(T) = G$. So the left-hand side vanishes. One the other hand, m is bounded below by a positive constant (strong maximum principle) and the u^N and u are uniformly Lipschitz continuous. So, by assumption (7), we have

$$C^{-1}\int_0^T\int_{\mathbb{T}^d}|Du^N - Du|^2 \;\leqslant\; -\int_0^T\int_{\mathbb{T}^d}(F^N(x,m^N(t)) - F(x,m(t,x))(m^N(t,x) - m(t,x))$$

As $F = F(x,m)$ is increasing with $\partial_m F \geqslant \delta$ and assumption (8) holds,

$$\begin{aligned}
&\int_0^T\int_{\mathbb{T}^d}(F^N(x,m^N(t) - F(x,m(T,x)))(m^N - m)\\
&\quad= \int_0^T\int_{\mathbb{T}^d}\left[(F^N(x,m^N(t)) - F(x,m^N(t,x))) + (F(x,m^N(t,x)) - F(x,m(t,x))\right](m^N - m)\\
&\quad\geqslant -Ck_N^{R,\alpha}\|m^N - m\|_1 + \delta\int_0^T\int_{\mathbb{T}^d}(m^N(t,x) - m(t,x))^2.
\end{aligned}$$

We obtain therefore

$$C^{-1}\int_0^T\int_{\mathbb{T}^d}|Du^N - Du|^2 + \delta\int_0^T\int_{\mathbb{T}^d}(m^N(t,x) - m(t,x))^2 \leqslant Ck_N^{R,\alpha}\|m^N - m\|_1 \leqslant Ck_N^{R,\alpha}\|m^N - m\|_{L^2}.$$

Hence

$$\|Du^N - Du\|_{L^2} + \|m^N - m\|_{L^2} \leqslant Ck_N^{R,\alpha}.$$

In particular

$$\begin{aligned}
&\|F^N(\cdot,m^N(t)) - F(\cdot,m(t,\cdot))\|_{L^2}\\
&\quad\leqslant \|F^N(\cdot,m^N(t)) - F(\cdot,m^N(t,\cdot))\|_\infty + \|F(\cdot,m^N(t,\cdot)) - F(\cdot,m(t,\cdot))\|_{L^2}\\
&\quad\leqslant Ck_N^{R,\alpha} + C\|m^N - m\|_{L^2} \;\leqslant\; Ck_N^{R,\alpha}.
\end{aligned}$$

Therefore the difference $w := u^N - u$ satisfies

$$-\partial_t w - \Delta w = g(t,x)$$

with $g(t,x) = F^N(x,m^N(t)) - F(x,m(t,x)) - H(x,Du^N(t,x)) + H(x,Du(t,x))$. By our previous bounds, we have $\|g\|_{L^2} \leqslant Ck_N^{R,\alpha}$, so that classical estimates on the heat equation imply that

$$\sup_{t\in[t_0,T]}\|u^N(t,\cdot) - u(t,\cdot)\|_{H^1(\mathbb{T}^d)} \leqslant Ck_N^{R,\alpha}.$$

As, by interpolation, for any smooth map $\phi : \mathbb{T}^d \to \mathbb{R}$,

$$\|\phi\|_\infty \leqslant C\|\phi\|_{L^2}^{\frac{2}{d+2}}\sup\{\|D\phi\|_\infty^{\frac{d}{d+2}}, 1\}$$

and since D^2u^N is bounded independently of N, we infer that

$$\sup_{t\in[t_0,T]}\|Du^N(t,\cdot) - Du(t,\cdot)\|_\infty \leqslant C\left(k_N^{R,\alpha}\right)^{\frac{2}{d+2}}.$$

□

A straightforward consequence of Proposition 2.2 is the following estimate on optimal trajectories related with the MFG systems (13) and (14).

Corollary 2.3. *Let $m_0 \in \mathcal{P}(\mathbb{T}^d)$, (u^N, m^N) and (u, m) be the solution to the MFG system (13) and (14) respectively. Let $t_0 \in [0, T)$ and Z be a random variable with law m_0 which is independent of a Brownian motion (B_t). If $(\tilde{X}_t)$ and (X_t) are the solution to*

$$\begin{cases} d\tilde{X}_t = -D_p H(\tilde{X}_t, Du(t, \tilde{X}_t))dt + \sqrt{2}dB_t \\ \tilde{X}_{t_0} = Z \end{cases}$$

and

$$\begin{cases} dX_t = -D_p H(X_t, Du^N(t, X_t))dt + \sqrt{2}dB_t \\ X_{t_0} = Z \end{cases}$$

respectively, then

$$\mathbb{E}\left[\sup_{t\in[t_0,T]} \left|\tilde{X}_t - X_t\right|\right] \leqslant C\left(k_N^{R,\alpha}\right)^{\frac{2}{d+2}},$$

where C, R and α are as in Proposition 2.2.

Proof. By Proposition 2.2, we have

$$\sup_{t\in[t_0,T]} \|Du^N(t,\cdot) - Du(t,\cdot)\|_\infty \leqslant C\left(k_N^{R,\alpha}\right)^{\frac{2}{d+2}}.$$

The conclusion then follows by standard estimates on ordinary differential equations. □

3 Convergence

In this section, we consider, for an integer $N \geqslant 2$, a classical solution $(v^{N,i})_{i\in\{1,\dots,N\}}$ of the Nash system:

$$\begin{cases} -\partial_t v^{N,i}(t,\boldsymbol{x}) - \sum_j \Delta_{x_j} v^{N,i}(t,\boldsymbol{x}) + H\big(x_i, D_{x_i} v^{N,i}(t,\boldsymbol{x})\big) \\ \qquad\qquad + \sum_{j\neq i} D_p H\big(x_j, D_{x_j} v^{N,j}(t,\boldsymbol{x})\big)\cdot D_{x_j} v^{N,i}(t,\boldsymbol{x}) = F^N(x_i, m_{\boldsymbol{x}}^{N,i}) \qquad \text{in } [0,T]\times(\mathbb{T}^d)^N, \\ v^{N,i}(T,\boldsymbol{x}) = G(x_i) \qquad \text{in } (\mathbb{T}^d)^N, \end{cases} \tag{15}$$

where we set, for $\boldsymbol{x} = (x_1, \dots, x_N) \in (\mathbb{T}^d)^N$, $m_{\boldsymbol{x}}^{N,i} = \dfrac{1}{N-1}\sum_{j\neq i} \delta_{x_j}$.

Our aim is to prove that the $v^{N,i}$ are close to u, where (u, m) is the solution of the MFG system (14). For this we first compare $v^{N,i}$ and u^N, where (u^N, m^N) is the solution of the perturbed MFG system (13).

3.1 Estimates between $v^{N,i}$ and u^N

Let us fix $t_0 \in [0, T)$, $m_0 \in \mathcal{P}(\mathbb{T}^d)$ with a C^2 density. We consider the solution (u^N, m^N) of the MFG system (13). Following Proposition 2.1, we know that u^N is bounded in $C^{2,\alpha}$ (for some $\alpha \in (0, 1)$): we will use this uniform regularity all along the section.

Let $(Z_i)_{i\in\{1,\dots,N\}}$ be an i.i.d family of N random variables of law m_0. We set $\boldsymbol{Z} = (Z_i)_{i\in\{1,\dots,N\}}$. Let also $((B_t^i)_{t\in[0,T]})_{i\in\{1,\dots,N\}}$ be a family of N independent d-dimensional Brownian Motions which is also independent of $(Z_i)_{i\in\{1,\dots,N\}}$. We consider the systems of SDEs with variables $(\boldsymbol{X}_t = (X_{i,t})_{i\in\{1,\dots,N\}})_{t\in[0,T]}$ and $(\boldsymbol{Y}_t = (Y_{i,t})_{i\in\{1,\dots,N\}})_{t\in[0,T]}$:

$$\begin{cases} dX_{i,t} = -D_p H\big(X_{i,t}, D_{x_i} u^N(t, X_{i,t})\big)dt + \sqrt{2}dB_t^i \qquad t \in [t_0, T] \\ X_{i,t_0} = Z_i, \end{cases} \tag{16}$$

and

$$\left\{ \begin{array}{l} dY_{i,t} = -D_p H\big(Y_{i,t}, D_{x_i} v^{N,i}(t, \boldsymbol{Y}_t)\big)dt + \sqrt{2} dB_t^i \qquad t \in [t_0, T] \\ Y_{i,t_0} = Z_i. \end{array} \right. \tag{17}$$

Note that the $(X_{i,t})$ are i.i.d. with law $m^N(t)$. Moreover, because of the symmetry properties of the $(v^{N,i})_{i\in\{1,\dots,N\}}$, the processes $(X_{i,t}, Y_{i,t})_{t\in[t_0,T]})_{i\in\{1,\dots,N\}}$ are exchangeable. Note finally that the X_i and the Y_i depends on N, but we do not write this dependence explicitly for the sake of simplicity.

Theorem 3.1. *Assume that N is so large that*

$$\left\{ \begin{array}{ll} K_N N^{-\frac{1}{d}} \leqslant \bar{C}^{-1} & \text{if } d \geqslant 3 \\ K_N N^{-\frac{1}{2}} \log(N) \leqslant \bar{C}^{-1} & \text{if } d = 2 \end{array} \right. \tag{18}$$

for some constant $\bar{C}$ depending on m_0 but independent of N. Then, for any $i \in \{1, \dots, N\}$,

$$\mathbb{E}\left[\sup_{t\in[t_0,T]} |X_{i,t} - Y_{i,t}| \right] \leqslant \left\{ \begin{array}{ll} C K_N^{\frac{1}{2}} N^{-\frac{1}{2d}} & \text{if } d \geqslant 3 \\ C K_N^{\frac{1}{2}} N^{-\frac{1}{4}} \log^{\frac{1}{2}}(N) & \text{if } d = 2 \end{array} \right.$$

and

$$\mathbb{E}\left[|u^N(t_0, Z_i) - v^{N,i}(t_0, \boldsymbol{Z})|\right] \leqslant \left\{ \begin{array}{ll} C\left(K_N^{\frac{3}{2}} N^{-\frac{1}{2d}} + K_N N^{-\frac{1}{d}}\right) & \text{if } d \geqslant 3 \\ C\left(K_N^{\frac{3}{2}} N^{-\frac{1}{4}} \log^{\frac{1}{2}}(N) + K_N N^{-\frac{1}{2}} \log(N)\right) & \text{if } d = 2 \end{array} \right.$$

where the constant C depends on m_0 but not on N.

Proof. For simplicity, we work with $t_0 = 0$. Let us introduce a few notations: for $\boldsymbol{x} = (x_j)_{j\in\{1,\dots,N\}} \in \mathbb{T}^{Nd}$ and $z \in \mathbb{T}^d$, let us denote by $v^{N,i}(t, z, \boldsymbol{x}^i)$ the value of $v^{N,i}(t, \cdot)$ evaluated at the point $\tilde{\boldsymbol{x}} = (\tilde{x}_j)_{j\in\{1,\dots,N}$ obtained from $\boldsymbol{x}$ by replacing x_i by z (i.e, $\tilde{x}_j = x_j$ if $j \neq i$ and $\tilde{x}_i = z$). We also denote by $\mathbb{E}^{\boldsymbol{Z}}$ the conditional expectation with respect to $\boldsymbol{Z}$.

By Ito's formula, we have

$$\mathbb{E}^Z\left[u^N(T, X_{i,T})\right] = \mathbb{E}^Z\left[u^N(0, Z_i) + \int_0^T \partial_t u^N + \Delta u^N - Du^N \cdot D_p H(X_{i,t}, Du^N(t, X_{i,t}))\ dt\right],$$

where u^N and its derivatives are evaluated at $(t, X_{i,t})$. As u^N solves (13), we get therefore

$$\begin{array}{rcl} u^N(0, \boldsymbol{Z}) & = & \mathbb{E}^{\boldsymbol{Z}}\left[\int_0^T (-H(X_{i,t}, Du^N) \right. \\ & & \left. + D_p H(X_{i,t}, Du^N) \cdot Du^N + F^N(X_{i,t}, m^N(t)))\ dt + G(X_{i,T})\right]. \end{array}$$

We now compute the variation of $v^{N,i}(t, X_{i,t}, \boldsymbol{Y}_t^i)$. Since the B.M. driving $X_{i,t}$ and those driving the $(Y_{j,t})_{j\neq i}$ are independent, we have, using the equation satisfied by $v^{N,i}$,

$$\begin{array}{rcl} dv^{N,i}(t, X_{i,t}, \boldsymbol{Y}_t^i) & = & (\partial_t v^{N,i} + \sum_j \Delta_{x_j} v^{N,i} - \sum_{j\neq i} D_{x_j} v^{N,i} \cdot D_p H(y_{j,t}, D_{x_i} v^{N,i}(t, \boldsymbol{Y}_t)))dt \\ & & -D_{x_i} v^{N,i} \cdot D_p H(X_{i,t}, Du^N(t, X_{i,t}))dt + \sqrt{2}\sum_j D_{x_j} v^{N,i} \cdot dB_t^j \\ & = & (H(X_{i,t}, D_{x_i} v^{N,i}) - D_{x_i} v^{N,i} \cdot D_p H(X_{i,t}, Du^N(t, X_{i,t})))dt \\ & & -F^N(X_{i,t}, m^{N,i}_{\boldsymbol{Y}_t})dt + \sqrt{2}\sum_j D_{x_j} v^{N,i} \cdot dB_t^j \end{array}$$

where $v^{N,i}$ and its derivatives are evaluated at $(t, X_{i,t}, \boldsymbol{Y}_t^i)$. Hence

$$\begin{aligned} v^{N,i}(0,\boldsymbol{Z}) \;=\; & \mathbb{E}^{\boldsymbol{Z}}\left[\int_0^T (-H(X_{i,t}, D_{x_i}v^{N,i}) + D_{x_i}v^{N,i}\cdot D_pH(X_{i,t}, Du^N(t,X_{i,t})) \right. \\ & \left. +F^N(X_{i,t}, m^{N,i}_{\boldsymbol{Y}_t}))\; dt + G(X_{i,T})\right]. \end{aligned}$$

So

$$\begin{aligned} & u^N(0,Z_i) - v^{N,i}(0,\boldsymbol{Z}) \\ & \quad = \mathbb{E}^{\boldsymbol{Z}}\left[\int_0^T \big[H(X_{i,t}, D_{x_i}v^{N,i}) - H(X_{i,t}, D_{x_i}u^N) - D_pH(X_{i,t}, Du^N)\cdot(D_{x_i}v^{N,i} - Du^N) \right. \\ & \qquad \left. +(F^N(X_{i,t}, m^N(t)) - F^N(X_{i,t}, m^{N,i}_{\boldsymbol{Y}_t}))\big]dt\right] \end{aligned} \tag{19}$$

Recall that Du^N is bounded by some constant R independently of N. Let us set, for $z \geqslant 0$,

$$\Psi(z) = \begin{cases} z^2 & \text{if } z \in [0,1] \\ 2z-1 & \text{if } z \geqslant 1 \end{cases} \tag{20}$$

From Lemma 3.3, there exists $C_0 > 0$ (which depends on R) such that

$$H(x,q) - H(x,p) - D_pH(x,p)\cdot(q-p) \geqslant C_0^{-1}\Psi(|q-p|) \qquad \forall p,q \text{ with } |p| \leqslant R.$$

Therefore

$$\begin{aligned} & u^N(0,Z_i) - v^{N,i}(0,\boldsymbol{Z}) \\ & \quad \geqslant \mathbb{E}^{\boldsymbol{Z}}\left[\int_0^T C_0^{-1}\Psi\left(|D_{x_i}v^{N,i}(t,X_{i,t},\boldsymbol{Y}_t^i) - D_{x_i}u^N(t,X_{i,t})|\right) \right. \\ & \qquad \left. +(F^N(X_{i,t}, m^N(t)) - F^N(X_{i,t}, m^{N,i}_{\boldsymbol{Y}_t}))\big]dt\right] \end{aligned} \tag{21}$$

Computing in the same way the variation of the terms $-u^N(t,Y_{i,t}) + v^{N,i}(t,\boldsymbol{Y}_t)$, we find

$$\begin{aligned} & -u^N(0,Z_i) + v^{N,i}(0,\boldsymbol{Z}) \\ & \quad = \mathbb{E}^{\boldsymbol{Z}}\left[\int_0^T \big[H(Y_{i,t}, Du^N) - H(Y_{i,t}, D_{x_i}v^{N,i}) - D_pH(Y_{i,t}, D_{x_i}v^{N,i})\cdot(Du^N - D_{x_i}v^{N,i}) \right. \\ & \qquad \left. +(F^N(Y_{i,t}, m^{N,i}_{\boldsymbol{Y}_t}) - F^N(Y_{i,t}, m^N(t))\big]dt\right] \end{aligned}$$

where Du^N and $D_{x_i}v^{N,i}$ are computed at $(t,Y_{i,t})$ and $(t,\boldsymbol{Y}_t)$ respectively.

In order to estimate the first term in the right-hand side, we use Lemma 3.3 to infer the existence of a constant $c_0 > 0$ (which depends on the uniform bound R on $\|Du^N\|_\infty$) such that

$$H(x,q) - H(x,p) - D_pH(x,p)\cdot(q-p) \geqslant c_0\min\{|p-q|^2, c_0\} \qquad \forall p,q \text{ with } |q| \leqslant R.$$

Therefore

$$\begin{aligned} & \mathbb{E}[-u^N(0,Z_i) + v^{N,i}(0,\boldsymbol{Z})] \\ & \quad \geqslant \mathbb{E}\left[\int_0^T c_0\min\{|Du^N(t,Y_{i,t}) - D_{x_i}v^{N,i}(t,\boldsymbol{Y}_t)|^2, c_0\} \right. \\ & \qquad \left. +(F^N(Y_{i,t}, m^{N,i}_{\boldsymbol{Y}_t}) - F^N(Y_{i,t}, m^N(t))\big]dt\right]. \end{aligned}$$

Combining this inequality with (21), we obtain

$$\begin{aligned} 0 \;\geqslant\; & \mathbb{E}\left[\int_0^T C_0^{-1}\Psi\left(|D_{x_i}v^{N,i}(t,X_{i,t},\boldsymbol{Y}_t^i) - D_{x_i}u^N(t,X_{i,t})|\right)\right] \\ & +\mathbb{E}\left[\int_0^T c_0\min\{|Du^N(t,Y_{i,t}) - D_{x_i}v^{N,i}(t,\boldsymbol{Y}_t)|^2, c_0\}\right] \\ & +\mathbb{E}\left[\int_0^T F^N(X_{i,t}, m^N(t)) - F^N(X_{i,t}, m^{N,i}_{\boldsymbol{Y}_t}) - F^N(Y_{i,t}, m^N(t)) + F^N(Y_{i,t}, m^{N,i}_{\boldsymbol{Y}_t}))dt\right] \end{aligned}$$

Let us set $m^N_{\boldsymbol{X}_t} = \frac{1}{N}\sum_j \delta_{X_{j,t}}$ and $m^N_{\boldsymbol{Y}_t} = \frac{1}{N}\sum_j \delta_{Y_{j,t}}$. We note that $\mathbf{d}_1(m^{N,i}_{\boldsymbol{Y}_t}, m^N_{\boldsymbol{Y}_t}) \leqslant CN^{-1}$. Moreover, as the $(X_{i,t})$ are i.i.d. with law $m^N(t)$, a result by Dereich, Scheutzow and Schottstedt [6] implies that, for $d \geqslant 3$,

$$\mathbb{E}\left[\mathbf{d}_1\left(m^N_{\boldsymbol{X}_t}, m^N(t)\right)\right] \leqslant CN^{-\frac{1}{d}}.$$

For $d = 2$, the estimate becomes (see Ajtai, Komlos and Tusnády [1]),

$$\mathbb{E}\left[\mathbf{d}_1\left(m^N_{\boldsymbol{X}_t}, m^N(t)\right)\right] \leqslant CN^{-\frac{1}{2}}\log(N).$$

As F^N is K_N-Lipschitz continuous (recall (10)) we obtain (for $d \geqslant 3$)

$$\begin{aligned} CK_N N^{-\frac{1}{d}} \;\geqslant\; & \mathbb{E}\left[\int_0^T C_0^{-1}\Psi\left(|D_{x_i}v^{N,i}(t, X_{i,t}, \boldsymbol{Y}^i_t) - D_{x_i}u^N(t, X_{i,t})|\right)\right] \\ & \mathbb{E}\left[\int_0^T c_0 \min\{|Du^N(t, Y_{i,t}) - D_{x_i}v^{N,i}(t, \boldsymbol{Y}_t)|^2, c_0\}\right] \\ & + \mathbb{E}\left[\int_0^T F^N(X_{i,t}, m^N_{\boldsymbol{X}_t}) - F^N(X_{i,t}, m^N_{\boldsymbol{Y}_t}) - F^N(Y_{i,t}, m^N_{\boldsymbol{X}_t}) + F^N(Y_{i,t}, m^N_{\boldsymbol{Y}_t}))dt\right] \end{aligned}$$

We now sum these expressions over i. Since

$$\begin{aligned} & \sum_i F^N(X_{i,t}, m^N_{\boldsymbol{X}_t}) - F^N(X_{i,t}, m^N_{\boldsymbol{Y}_t}) - F^N(Y_{i,t}, m^N_{\boldsymbol{X}_t}) + F^N(Y_{i,t}, m^N_{\boldsymbol{Y}_t}) \\ & \quad = \int_{\mathbb{T}^d} (F^N(x, m^N_{\boldsymbol{X}_t}) - F^N(x, m^N_{\boldsymbol{Y}_t}))d(m^N_{\boldsymbol{X}_t} - m^N_{\boldsymbol{Y}_t})(x) \;\geqslant\; 0, \end{aligned}$$

we obtain:

$$\begin{aligned} CK_N N^{1-\frac{1}{d}} \;\geqslant\; & \sum_i \mathbb{E}\left[\int_0^T C_0^{-1}\Psi\left(|D_{x_i}v^{N,i}(t, X_{i,t}, \boldsymbol{Y}^i_t) - D_{x_i}u^N(t, X_{i,t})|\right)\right] \\ & + \sum_i \mathbb{E}\left[\int_0^T c_0 \min\{|Du^N(t, Y_{i,t}) - D_{x_i}v^{N,i}(t, \boldsymbol{Y}_t)|^2, c_0\}\right]. \end{aligned}$$

By symmetry of the $(v^{N,i})$, the random variables

$$D_{x_i}v^{N,i}(t, X_{i,t}, \boldsymbol{Y}^i_t) - D_{x_i}u^N(t, X_{i,t})$$

have the same law for any i. In the same way, the random variables

$$Du^N(t, Y_{i,t}) - D_{x_i}v^{N,i}(t, \boldsymbol{Y}_t)$$

have the same law for any i. We have therefore, for any $i \in \{1, \dots, N\}$ and $d \geqslant 3$,

$$\begin{aligned} CK_N N^{-\frac{1}{d}} \;\geqslant\; & \mathbb{E}\left[\int_0^T C_0^{-1}\Psi\left(|D_{x_i}v^{N,i}(t, X_{i,t}, \boldsymbol{Y}^i_t) - D_{x_i}u^N(t, X_{i,t})|\right)\right] \\ & + \mathbb{E}\left[\int_0^T c_0 \min\{|Du^N(t, Y_{i,t}) - D_{x_i}v^{N,i}(t, \boldsymbol{Y}_t)|^2, c_0\}\right]. \end{aligned} \tag{22}$$

In view of the SDEs satisfied by the $(X_{i,t})$ and by the $(Y_{i,t})$, we have

$$\begin{aligned} |X_{i,t} - Y_{i,t}| \;\leqslant\; & \int_0^t |-D_pH(X_{i,s}, Du^N(s, X_{i,s})) + D_pH(Y_{i,s}, D_{x_i}v^{N,i}(s, \boldsymbol{Y}_s))|\, ds \\ \leqslant\; & \int_0^t |-D_pH(X_{i,s}, Du^N(s, X_{i,s})) + D_pH(Y_{i,s}, Du^N(s, Y_{i,s}))|\, ds \\ & + \int_0^t |-D_pH(Y_{i,s}, Du^N(s, Y_{i,s})) + D_pH(Y_{i,s}, D_{x_i}v^{N,i}(s, \boldsymbol{Y}_s))\, ds \\ \leqslant\; & C\int_0^t |X_{i,s} - Y_{i,s}|\, ds + C\int_0^T \min\{|Du^N(s, Y_{i,s}) - D_{x_i}v^{N,i}(s, \boldsymbol{Y}_s)|, \|D_pH\|_\infty\}\, ds \end{aligned}$$

where we have used the bound and Lipschitz regularity of D_pH as well as the uniform Lipschitz bound of Du^N in the space variable x. So, by Gronwall's inequality and (22), we obtain, for any $i \in \{1, \dots, N\}$ and $d \geqslant 3$,

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X_{i,t} - Y_{i,t}|\right] \leqslant C K_N^{\frac{1}{2}} N^{-\frac{1}{2d}}. \tag{23}$$

When $d = 2$, the right-hand side has to be replaced by $C K_N^{\frac{1}{2}} N^{-\frac{1}{4}} \log^{\frac{1}{2}}(N)$.

In order to estimate $u^N(0, Z_i) - v^{N,i}(0, \boldsymbol{Z})$, we come back to (19). By the Lipschitz continuity of H and F^N we have:

$$\begin{aligned}
&\mathbb{E}\left[|u^N(0, Z_i) - v^{N,i}(0, \boldsymbol{Z})|\right] \\
&\quad \leqslant \mathbb{E}\left[\int_0^T C|D_{x_i} v^{N,i}(t, X_{i,t}, \boldsymbol{Y}_t^i) - D_{x_i} u^N(t, X_{i,t})| + K_N \mathbf{d}_1(m^N(t), m^{N,i}_{\boldsymbol{Y}_t^i}) dt\right].
\end{aligned} \tag{24}$$

Let us first estimate the first term in the right-hand side of (24) (for $d \geqslant 3$): we use inequality (22) and the fact that Ψ is convex and increasing:

$$\begin{aligned}
&T^{-1}\mathbb{E}\left[\int_0^T |D_{x_i} v^{N,i}(t, X_{i,t}, \boldsymbol{Y}_t^i) - D_{x_i} u^N(t, X_{i,t})| dt\right] \\
&\quad \leqslant \Psi^{-1}\left(T^{-1}\mathbb{E}\left[\int_0^T \Psi\left(|D_{x_i} v^{N,i}(t, X_{i,t}, \boldsymbol{Y}_t^i) - D_{x_i} u^N(t, X_{i,t})|\right) dt\right]\right) \\
&\quad \leqslant \Psi^{-1}\left(C K_N N^{-\frac{1}{d}}\right)
\end{aligned}$$

So, if we suppose as in assumption (18) that $K_N N^{-\frac{1}{d}} \leqslant C^{-1}$, we obtain

$$\mathbb{E}\left[\int_0^T |D_{x_i} v^{N,i}(t, X_{i,t}, \boldsymbol{Y}_t^i) - D_{x_i} u^N(t, X_{i,t})| dt\right] \leqslant C K_N^{\frac{1}{2}} N^{-\frac{1}{2d}}.$$

To estimate the second term in the right-hand side of (24), we note that

$$\begin{aligned}
\mathbf{d}_1(m^N(t), m^{N,i}_{\boldsymbol{Y}_t^i}) &\leqslant \mathbf{d}_1(m^N(t), m^{N,i}_{\boldsymbol{X}_t^i}) + \mathbf{d}_1(m^{N,i}_{\boldsymbol{X}_t^i}, m^{N,i}_{\boldsymbol{Y}_t^i}) \\
&\leqslant \mathbf{d}_1(m^N(t), m^{N,i}_{\boldsymbol{X}_t^i}) + \frac{1}{N}\sum_{j\neq i} |X_{j,t} - Y_{j,t}|.
\end{aligned}$$

So, using, on the one hand, the fact that the $(X_{i,t})$ are i.i.d. with law $m^N(t)$ and the result by Dereich, Scheutzow and Schottstedt [6] and, on the other hand, inequality (23), we have

$$\mathbb{E}\left[\mathbf{d}_1(m^N(t), m^{N,i}_{\boldsymbol{Y}_t^i})\right] \leqslant C\left(N^{-\frac{1}{d}} + K_N^{\frac{1}{2}} N^{-\frac{1}{2d}}\right).$$

This proves that, if $d \geqslant 3$,

$$\mathbb{E}\left[|u^N(0, Z_i) - v^{N,i}(0, \boldsymbol{Z})|\right] \leqslant C\left(K_N^{\frac{3}{2}} N^{-\frac{1}{2d}} + K_N N^{-\frac{1}{d}}\right).$$

When $d = 2$, the right-hand side becomes $C\left(K_N^{\frac{3}{2}} N^{-\frac{1}{4}} \log^{\frac{1}{2}}(N) + K_N N^{-\frac{1}{2}} \log(N)\right)$. □

Remark 3.2. A variant of Theorem 3.1 can be obtained by replacing u^N by u in the definition of the X_i and in the whole proof, thus avoiding the approximation argument of Propositions 2.1 and 2.2. As the assumption $\partial_m F \geqslant \delta$ is only used in Proposition 2.2, this condition can then be removed. The price to pay is a deterioration of the convergence rate because the left-hand side of (22) has to involve a term of the form $\sup_t \|F(\cdot, m(t)) - F^N(\cdot, m(t))\|_\infty$.

In the proof we used the

Lemma 3.3. *Assume that $D^2_{pp}H > 0$ and let Ψ be defined by (20). Then, for any $R > 0$, there exists $C_0, c_0 > 0$ such that*

$$H(x,q) - H(x,p) - D_pH(x,p)\cdot(q-p) \geqslant C_0^{-1}\Psi(|q-p|) \qquad \forall p, q \text{ with } |p| \leqslant R.$$

and

$$H(x,q) - H(x,p) - D_pH(x,p)\cdot(q-p) \geqslant c_0 \min\{|p-q|^2, c_0\} \qquad \forall p, q \text{ with } |q| \leqslant R.$$

Proof. As $D^2_{pp}H > 0$, there exists $\theta > 0$, depending on R, such that $D^2_{pp}H \geqslant \theta$ in $\mathbb{T}^d \times B_{2R}(0)$. Let $x \in \mathbb{T}^d$, $p, q \in \mathbb{R}^d$ with $|p| \leqslant R$. If $|q-p| \leqslant R$, then by the lower bound $D^2_{pp}H \geqslant \theta$ we have

$$H(x,q) - H(x,p) - D_pH(x,p)\cdot(q-p) \geqslant \frac{\theta}{2}|p-q|^2.$$

Now assume that $|q-p| > R$. Let $\hat{q}$ be the projection of q onto the ball $B_R(p)$. Then (omitting the x dependence which plays no role)

$$\begin{aligned}
&H(q) - H(p) - D_pH(p)\cdot(q-p)\\
&\quad = H(q) - H(\hat{q}) - D_pH(\hat{q})\cdot(q-\hat{q}) + H(\hat{q}) - H(p) - D_pH(p)\cdot(\hat{q}-p)\\
&\qquad +(D_pH(\hat{q}) - D_pH(p))\cdot(q-\hat{q})\\
&\quad \geqslant \frac{\theta}{2}|p-\hat{q}|^2 + R^{-1}(|q-p| - R)(D_pH(\hat{q}) - D_pH(p))\cdot(\hat{q}-p)
\end{aligned}$$

since $q-\hat{q}$ and $\hat{q}-p$ are collinear and $|\hat{q}-p| = R$. Using once more the lower bound on $D^2_{pp}H$ in $B_R(p)$, we get

$$H(q) - H(p) - D_pH(p)\cdot(q-p) \geqslant \frac{\theta}{2}R^2 + (|q-p| - R)R\theta.$$

This gives the first result. The second one is obtained in the same way: the inequality holds if $|q-p| \leqslant R$. Otherwise, let $\hat{p}$ be the projection of p onto $B_R(q)$. Then

$$\begin{aligned}
&H(q) - H(p) - D_pH(p)\cdot(q-p)\\
&\quad = H(q) - H(\hat{p}) - D_pH(\hat{p})\cdot(q-\hat{p}) + H(\hat{p}) - H(p) - D_pH(p)\cdot(\hat{p}-p)\\
&\qquad +(D_pH(\hat{p}) - D_pH(p))\cdot(q-\hat{p})\\
&\quad \geqslant \frac{\theta}{2}|p-\hat{p}|^2 = \frac{\theta}{2}R^2.
\end{aligned}$$

□

3.2 Putting the estimates together

Here we fix a initial condition $(t_0, m_0) \in [0,T] \times \mathcal{P}(\mathbb{T}^d)$, where m_0 has a positive density of class C^2. Let $v^{N,i}$ be the solution of the Nash system (15). We reduce $v^{N,i}$ to the variables (t, x_i) by setting:

$$w^{N,i}(t_0, x_i, m_0) := \int_{\mathbb{T}^d} \dots \int_{\mathbb{T}^d} v^{N,i}(t_0, \boldsymbol{x}) \prod_{j \neq i} m_0(dx_j) \qquad \text{where } \boldsymbol{x} = (x_1, \dots, x_N).$$

Let u be the solution to the MFG system (14). Combining, Proposition 2.2 and Theorem 3.1 we have:

Theorem 3.4. *If condition* (18) *holds, then*

$$\left\| w^{N,i}(t_0, \cdot, m_0) - u(t_0, \cdot) \right\|_{L^1(m_0)} \leqslant \begin{cases} C\left(K_N^{\frac{3}{2}} N^{-\frac{1}{2d}} + K_N N^{-\frac{1}{d}} + k_N^{R,\alpha} \right) & \text{if } d \geqslant 3 \\ C\left(K_N^{\frac{3}{2}} N^{-\frac{1}{4}} \log^{\frac{1}{2}}(N) + K_N N^{-\frac{1}{2}} \log(N) + k_N^{R,\alpha} \right) & \text{if } d = 2 \end{cases}, \tag{25}$$

where R and α do not depend on N (but depend on m_0). In particular, $w^{N,i}(t_0, \cdot)$ converges to $u(t_0, \cdot)$ in $L^1(\mathbb{T}^d)$ as soon as $K_N = o(N^{\frac{1}{3d}})$ if $d \geqslant 3$ and $K_N = o(N^{\frac{1}{6}} / \log^{\frac{1}{3}}(N))$ if $d = 2$.

Next we discuss the convergence of the optimal solutions and the propagation of chaos. Let (Z_i) be an i.i.d family of N random variables of law m_0. We set $\boldsymbol{Z} = (Z_1, \dots, Z_N)$. Let also $((B_t^i)_{t \in [0,T]})_{i \in \{1,\dots,N\}}$ be a family of N independent Brownian motions which is also independent of (Z_i). We consider the optimal trajectories $(\boldsymbol{Y}_t = (Y_{1,t}, \dots, Y_{N,t}))_{t \in [t_0,T]}$ for the $N-$player game:

$$\begin{cases} dY_{i,t} = -D_p H(Y_{i,t}, D_{x_i} v^{N,i}(t, \boldsymbol{Y}_t))dt + \sqrt{2} dB_t^i, & t \in [t_0, T] \\ Y_{i,t_0} = Z_i \end{cases}$$

and the optimal solution $(\tilde{\boldsymbol{X}}_t = (\tilde{X}_{1,t}, \dots, \tilde{X}_{N,t}))_{t \in [t_0,T]}$ to the limit MFG system:

$$\begin{cases} d\tilde{X}_{i,t} = -D_p H\left(\tilde{X}_{i,t}, Du\big(t, \tilde{X}_{i,t}\big)\right) dt + \sqrt{2} dB_t^i, & t \in [t_0, T] \\ \tilde{X}_{i,t_0} = Z_i. \end{cases}$$

The next result provides an estimate of the distance between the solutions:

Theorem 3.5. *Under the assumption of Theorem 3.4, we have*

$$\mathbb{E}\left[\sup_{t \in [t_0,T]} \left| Y_{i,t} - \tilde{X}_{i,t} \right| \right] \leqslant \begin{cases} C\left[K_N^{\frac{1}{2}} N^{-\frac{1}{2d}} + \left(k_N^{R,\alpha} \right)^{\frac{2}{d+2}} \right] & \text{if } d \geqslant 3 \\ C\left[K_N^{\frac{1}{2}} N^{-\frac{1}{4}} \log^{\frac{1}{2}}(N) + \left(k_N^{R,\alpha} \right)^{\frac{1}{2}} \right] & \text{if } d = 2 \end{cases} \tag{26}$$

where the constant $C > 0$ is independent of N and N. In particular, Y_i converges to $\tilde{X}_i$ if $K_N = o(N^{\frac{1}{d}})$ if $d \geqslant 3$ and $K_N = o(N^{\frac{1}{2}} / \log(N))$ if $d = 2$.

The proof is an immediate application of Corollary 2.3 and Theorem 3.1. We finally apply the above estimates to our main example:

Corollary 3.6. *Assume that* $F^N = F^{\epsilon_N}$ *where*

$$F^\epsilon(x,m) = F(\cdot, \xi^\epsilon \star m(\cdot)) \star \xi^\epsilon(x)$$

and where ξ^ϵ *is as in the example in Proposition 1.1. If one chooses* $\epsilon_N = N^{-\beta}$*, with* $\beta \in (0, (3d(d+1))^{-1})$*, then there exists* $\gamma \in (0,1)$ *such that*

$$\left\| w^{N,i}(t_0,\cdot,m_0) - u(t_0,\cdot) \right\|_{L^1(m_0)} \leqslant CN^{-\gamma}$$

and

$$\mathbb{E}\left[\sup_{t\in[t_0,T]} \left| Y_{i,t} - \tilde{X}_{i,t} \right| \right] \leqslant CN^{-\gamma}.$$

Proof. From Proposition 1.1, we can choose

$$k_N^{R,\alpha} \leqslant C(1+R)\epsilon_N^\alpha = CN^{-\alpha\beta}, \qquad K_N \leqslant C\epsilon_N^{-d-1} = CN^{\beta(N+1)}.$$

Inserting these inequality into (25) gives (for $d \geqslant 3$),

$$\left\| w^{N,i}(t_0,\cdot,m_0) - u(t_0,\cdot) \right\|_{L^1(m_0)} \leqslant C\left(N^{3\beta(d+1)/2 - \frac{1}{2d}} + N^{\beta(d+1)-\frac{1}{d}} + N^{\alpha\beta} \right),$$

where the right-hand side is of order $N^{-\gamma}$ for some $\gamma \in (0,1)$ thanks to our choice of β. In the same way, by (26),

$$\mathbb{E}\left[\sup_{t\in[t_0,T]} \left| Y_{i,t} - \tilde{X}_{i,t} \right| \right] \leqslant C\left[N^{\beta(d+1)/2 - \frac{1}{2d}} + N^{-\alpha\beta\frac{2}{d+2}} \right],$$

which also yield to an algebraic rate of convergence. Computation for the case $d = 2$ is similar. □

References

[1] Ajtai, M., Komlos, J., and Tusnády, G. (1984). *On optimal matchings.* Combinatorica, 4(4), 259-264.

[2] Buckdahn, R., Li, J., Peng, S., and Rainer, C. (2014). *Mean-field stochastic differential equations and associated PDEs.* arXiv preprint arXiv:1407.1215.

[3] Cardaliaguet P., Delarue F., Lasry J.-M., and Lions P.-L. (2015) *The master equation and the convergence problem in mean field games.* Preprint.

[4] Carmona, R., and Delarue, F. (2013) *Probabilist analysis of Mean-Field Games.* SIAM Journal on Control and Optimization, 51(4), 2705-2734.

[5] Chassagneux, J. F., Crisan, D., and Delarue, F. (2014). *Classical solutions to the master equation for large population equilibria.* arXiv preprint arXiv:1411.3009.

[6] Dereich, S., Scheutzow, M., and Schottstedt, R. (2013). *Constructive quantization: approximation by empirical measures.* In Annales de l'IHP, Probabilités et Statistiques, 49 (4), 1183-1203.

[7] Fischer M. (2014) *On the connection between symmetric N-player games and mean field games.* Preprint.

[8] Gangbo, W., Swiech A. (2014) *Existence of a solution to an equation arising from the theory of mean field games.* Preprint.

[9] 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.

[10] Gomes, D. A., Pimentel, E., Sánchez-Morgado, H. (2016). Time-dependent mean-field games in the superquadratic case. ESAIM: Control, Optimisation and Calculus of Variations, 22(2), 562-580.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] Kolokoltsov, V. N., Troeva, M., Yang, W. (2014). *On the Rate of Convergence for the Mean-Field Approximation of Controlled Diffusions with Large Number of Players.* Dyn Games Appl, 4, 208–230.

[16] Lacker, D. (2014). *A general characterization of the mean field limit for stochastic differential games.* arXiv preprint arXiv:1408.2708.

[17] 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

[18] Lasry, J.-M., Lions, P.-L. *Jeux à champ moyen. I. Le cas stationnaire.* C. R. Math. Acad. Sci. Paris 343 (2006), no. 9, 619–625.

[19] Lasry, J.-M., Lions, P.-L. *Jeux à champ moyen. II. Horizon fini et contrôle optimal.* C. R. Math. Acad. Sci. Paris 343 (2006), no. 10, 679–684.

[20] Lasry, J.-M., Lions, P.-L. *Mean field games.* Jpn. J. Math. 2 (2007), no. 1, 229–260.

[21] Lions, P.-L. *Cours au Collège de France.* www.college-de-france.fr.

[22] Porretta, A. (2013). *Weak Solutions to Fokker-Planck Equations and Mean Field Games.* Archive for Rational Mechanics and Analysis, 1-62.