MEAN FIELD GAMES MASTER EQUATIONS : FROM DISCRETE TO CONTINUOUS STATE SPACE

CHARLES BERTUCCI AND ALEKOS CECCHIN

Abstract. This paper studies the convergence of mean field games with finite state to mean field games with a continuous state space. We are mainly interested in the convergence of the solution of the associated master equations as the number of states tends to infinity. We present two approaches. The first one uses the system of characteristics of the master equation and the second one relies on the notion of monotone solutions. In both cases, we show that convergence holds under standard assumptions and that a rate of convergence can be obtained if a smooth solution to the limit equation exists.

1. Introduction

This paper is interested in the convergence of value functions of mean field games (MFG for short) in finite state space when the number of states tends to infinity. We show that if the MFG in finite state space is a suitable discretization of a continuous MFG, the value, i.e. the solution of the master equation, in finite state space converges toward the value of the MFG in continuous state space when the number of states tends to infinity.

1.1. **General introduction.** MFG are differential games involving non-atomic players which interact only through mean field terms. A general mathematical study of such games started with [24, 25] and independently in [22]. We refer to the books [15, 13] for a more complete presentation of the theory. By now several properties of those games are understood, we present the two properties which are the most helpful to understand the following. The first one is that Nash equilibria of the game can be characterized in terms of differential equations which may be stochastic or not, depending on the nature of the game, and which are ordinary differential equations (ODE for short) if the state space of the players is finite or partial differential equations (PDE for short) if the state space of the players is continuous. Such a characterization of the Nash equilibria is called the MFG system. The second main aspect of MFG is that an adversarial regime can be identified. In this so-called monotone regime, there is always a unique Nash equilibrium in the MFG. This property allows to define a concept of value in this situation. This value is the solution of a PDE called the master equation, which is set on a finite dimensional space if the state space of the players is finite and on an infinite dimensional space if the state space of the players is continuous. MFG master equations have attracted quite a lot of attention in the last years. In the finite state space, let us mention [10, 17, 18, 5, 16, 7] and the works [13, 26, 20, 19, 8, 14, 27] in the continuous state space case.

1.2. **Bibliographical comments.** The convergence of finite state MFG toward ones with continuous state space has been studied in [21] without common noise and at the level of the MFG system, where a compactness argument was established to prove their convergence result. The convergence of numerical schemes for MFG systems is also of related nature, especially for finite difference schemes, since the main difference lies in the fact that in numerical schemes the time is also discretized. Several results of convergence holds for finite difference schemes of MFG systems such as [1, 2, 3, 6, 4, 11, 9].

Date: February 22, 2022.

1.3. **Main results of the paper.** The approach of in this paper is twosome. First we study MFG without common noise. We establish that given a classical solution of the limit master equation, a rather direct approach yields a rate of convergence for both the value of the MFG and a useful quantity involving the optimal controls for a discretization only valid in the presence of idiosyncratic noise. We then present a less restrictive approach, which uses the system of characteristics and the monotonicity and which allows to prove a convergence rate for both the value of the MFG and the trajectories at the Nash equilibrium.

In a second time we study MFG with common noise in the monotone regime. We prove first a compactness result which allows to prove the convergence of the value of the MFG. We then establish that if a classical solution of the limit problem exists, then a rate of convergence for the value can also be proved. This part on MFG with common noise relies on an intrinsic study of the master equation through the concept of monotone solutions introduced in [7, 8] and used in [14]. This concept enables us to work with solution of the master equations which are merely continuous in the finite state space case and only continuous with respect to the measure variable in the continuous state space limit. The aforementioned convergence result can be seen as an illustration of the stability of these monotone solutions.

1.4. **Notation.**

- $\langle \cdot, \cdot \rangle$ stands for either the usual scalar product between two element of $\mathbb{R}^d$ or for the extension of the L^2 scalar product for functional spaces in duality, depending on the context.
- For a function $U : \mathcal{P}(\mathbb{T}^d) \to \mathbb{R}$, when it is defined we note
$$\frac{\delta U}{\delta m}(m, x) = \lim_{\theta \to 0} \frac{U((1-\theta)m + \theta\delta_x) - U(m)}{\theta}. \tag{1}$$
- For $\mu, \nu \in \mathcal{P}(E)$, with E a metric space, we denote by W_1 the Monge-Kantorovich distance between μ and ν.

2. The limit model

In this section we introduce the model at interest in the limit of an infinite number of states, without common noise. In this paper we shall focus on a one dimensional state space. This is mainly to simplify the already heavy notations. We leave to the interested reader the generalization to a higher dimensional state space and indicate along the paper the arguments who do not immediately extend to such a situation.

We first consider the master equation of unknown $U : [0,T] \times \mathbb{T} \times \mathcal{P}(\mathbb{T}) \to \mathbb{R}$, in dimension 1 in the torus:

$$\begin{aligned} & -\partial_t U - \sigma\partial_{xx}U + H(x, \partial_x U) - \left\langle \sigma\partial_{xx}m + \partial_x(\partial_p H(\cdot, \partial_x U)m), \frac{\delta U}{\delta m}(t, x, m, \cdot) \right\rangle = f(m)(x) \\ & U(T, x, m) = g(m)(x), \end{aligned} \tag{2}$$

where $\sigma > 0$, $H : \mathbb{T} \times \mathbb{R} \to \mathbb{R}, f, g : \mathcal{P}(\mathbb{T}) \to \mathcal{C}(\mathbb{T})$ and $T > 0$ are data of the problem. This equation corresponds to the following MFG. The dynamics of a player is

$$dX_t = \alpha(t, X_t)dt + \sqrt{2\sigma}dW_t, \tag{3}$$

where α is its closed-loop control and $(W_t)_{t\geq 0}$ is a Brownian motion on a standard filtered probability space. Given an anticipation $(\mu_t)_{t\in[0,T]}$ on the repartition of the players in the state space, the expected cost of this player is given by

$$J(\alpha, (\mu_t)_{t\geq 0}) = \mathbb{E}\left[\int_0^T L(x, \alpha(t, X_t)) + f(\mu_t)(X_t)dt + g(\mu_T)(X_T)\right], \tag{4}$$

where L is a cost function such that its Legendre transform L^* with respect to its second argument is equal to the Hamiltonian $H(x,p)$. Nash equilibria of the MFG can be characterized through the MFG system

$$\begin{cases} -\partial_t u - \sigma\partial_x^2 u + H(x,\partial_x u) = f(\mu_t)(x) \\ \partial_t \mu - \sigma\partial_x^2\mu - \partial_x(\partial_p H(x,\partial_x u)\mu) = 0 \\ u(T,x) = g(\mu_T)(x) \qquad \mu_0 = m_0. \end{cases} \tag{5}$$

Let us finish this section by giving the main assumptions that shall be in force for the rest of paper.

We assume that the couplings $f, g : \mathcal{M}(\mathbb{T}) \to \mathcal{C}^\gamma(\mathbb{T})$ are continuous mappings for some $\gamma \in (0,1)$. We also assume that they are monotone, i.e.

$$\begin{aligned} &\int_{\mathbb{T}} (f(x,m) - f(x,\widetilde{m}))(m-\widetilde{m})(dx) \geq 0 \qquad \forall m, \widetilde{m} \in \mathcal{P}(\mathbb{T}) \\ &\int_{\mathbb{T}} (g(x,m) - g(x,\widetilde{m}))(m-\widetilde{m})(dx) \geq 0 \qquad \forall m, \widetilde{m} \in \mathcal{P}(\mathbb{T}) \end{aligned} \tag{6}$$

The duration of the game $T > 0$ is arbitrary long but fixed.

The Hamiltonian $H(x,p)$ is smooth, uniformly convex in p.

3. Discretization

We construct approximations on the previous model with finite state and continuous time. The dynamics of the underlying Markov chain has the peculiarity that it jumps either right or left, with the convention that at the boundary it jumps on the other side. For any n, we consider then the n states $S^n = \{x_1^n, \dots, x_n^n\} = \{1/n, \dots, 1 = 0\}$ with mutual distance $\Delta x_n = 1/n$.

The discretization we study is the natural one, from a control or MFG perspective, see for instance the seminal papers [1, 2].

3.1. **The discrete model.**

We assume to control the jump rate on the right and on the left by means of functions denoted $\alpha_+^n, \alpha_-^n : [0,T] \times S^n \to [0,+\infty)$; the Markov chain X^n then satisfies

$$\begin{aligned} \mathbb{P}(X_{t+\Delta t}^n = x_{i+1}^n | X_t^n = x_i^n) &= \left(\frac{\alpha_+^n(t,x_i^n)}{\Delta x_n} + \frac{1}{\Delta x_n^2}\right)\Delta t + o(\Delta t), \\ \mathbb{P}(X_{t+\Delta t}^n = x_{i-1}^n | X_t^n = x_i^n) &= \left(\frac{\alpha_-^n(t,x_i^n)}{\Delta x_n} + \frac{1}{\Delta x_n^2}\right)\Delta t + o(\Delta t), \end{aligned} \tag{7}$$

with the convention that $x_{n+1}^n = x_1^n = 1/n$ and $x_0^n = x_n^n = 1$. Given anticipations $(\mu_t^n)_{t\geq 0}$ on the repartition of players in S^n the cost is given by

$$J^n(\alpha_\pm^n, \mu^n) = \mathbb{E}\left[\int_0^T L(X_t^n, \alpha_+^n(t,X_t^n)) + L(X_t^n, \alpha_-^n(t,X_t^n)) - L(X_t^n, 0) + f(\mu_t^n)(X_t^n)dt + g(\mu_T^n)(X_t^n)\right]. \tag{8}$$

The Nash equilibrium condition reads $\mu_t^n = \mathcal{L}(X_t^n)$.

For a function $u : S^n \to \mathbb{R}$ we denote the right and left first order n finite difference by

$$\Delta_+^n u(x) = \frac{u(x+\Delta x_n) - u(x)}{\Delta x_n} \qquad \Delta_-^n u(x) = \frac{u(x-\Delta x_n) - u(x)}{\Delta x_n}, \tag{9}$$

and the second order finite difference

$$\Delta_2^n u(x) = \frac{u(x+\Delta x_n) - 2u(x) + u(x-\Delta x_n)}{\Delta x_n^2}. \tag{10}$$

If $u : \mathbb{T} \to \mathbb{R}$ is smooth, then $\lim_{n\to\infty} \Delta_\pm^n u(x) = \pm\partial_x u(x)$ and $\lim_{n\to\infty} \Delta_2^n u(x) = \partial_x^2 u(x)$. The optimization provides the discrete HJB equation

$$-\frac{d}{dt}u^n + H_\uparrow(x, \Delta_+^n u^n(x)) + H_\downarrow(x, \Delta_-^n u^n(x)) - \sigma\Delta_2^n u^n(x) = f(\mu_t^n)(x), \qquad x \in S^n, \tag{11}$$

where

$$H_\uparrow(x,p) := -\inf_{\alpha\geq 0}\{L(x,\alpha) + \alpha p\}; \qquad H_\downarrow(x,p) := -\inf_{\alpha\geq 0}\{L(x,-\alpha) + \alpha p\} + L(x,0). \tag{12}$$

The optimal controls are given in feedback form by

$$\alpha_+^n(t,x) = -\partial_p H_\uparrow(x, \Delta_+^n u^n(x)), \qquad \alpha_-^n(t,x) = \partial_p H_\downarrow(x, \Delta_-^n u^n(x)), \qquad x \in S^n. \tag{13}$$

Let us remark that $H_\uparrow$ and $H_\downarrow$ are such that for $(x,p) \in \mathbb{T} \times \mathbb{R}$

$$\begin{aligned} H_\uparrow(x,p) + H_\downarrow(x,p) &= H(x,p) \\ -\partial_p H_\uparrow(x,p) - \partial_p H_\downarrow(x,p) &= -\partial_p H(x,p) \end{aligned} \tag{14}$$

The generator of the dynamics of X^n associated to this control is given by

$$\begin{aligned} \mathcal{L}^n\varphi(x) &= \left(\frac{-\partial_p H_\uparrow(x,\Delta_+^n u^n(x))}{\Delta x_n} + \frac{\sigma}{\Delta x_n^2}\right)[\varphi(x+\Delta x_n) - \varphi(x)] \\ &\quad + \left(\frac{\partial_p H_\downarrow(x,\Delta_-^n u^n(x))}{\Delta x_n} + \frac{\sigma}{\Delta x_n^2}\right)[\varphi(x-\Delta x_n) - \varphi(x)] \\ &= -\partial_p H_\uparrow(x,\Delta_+^n u^n(x))\Delta_+^n\varphi(x) + \partial_p H_\downarrow(x,\Delta_-^n u^n(x))\Delta_-^n\varphi(x) + \sigma\Delta_2^n\varphi(x). \end{aligned} \tag{15}$$

Hence the discrete Fokker-Planck equation associated to this generator is given by

$$\begin{aligned} &\frac{d}{dt}\mu^n(t,x) - \sigma\Delta_2^n\mu^n(t,x) - \Delta_+^n(\partial_p H_\uparrow(x,\Delta_+^n u^n(x))\mu^n(t,x)) \\ &\qquad + \Delta_-^n(\partial_p H_\downarrow(x,\Delta_-^n u^n(x))\mu^n(t,x)) = 0, \qquad x \in S^n. \end{aligned} \tag{16}$$

For a function U defined on $\mathcal{P}(S^n)$ we denote by $\partial_{m_j}U$ its derivative along direction e_j; and denote equivalently $e_j = e_{x_j}$ and $\partial_{m_j}U = \partial_{m_{x_j}}U$, because we view $m \in \mathcal{P}(S^n)$ as $m = \sum_{j=1}^n m_j\delta_{x_j}$. More precisely, we will consider only derivatives along directions $(e_j - e_i)$, which are tangent vectors to the simplex. The discrete master equation for $U^n : [0,T] \times S^n \times \mathcal{P}(S^n)$ is then given by

$$\begin{aligned} &-\partial_t U^n(t,x,m) + H_\uparrow(x,\Delta_+^n U^n(t,x,m)) + H_\downarrow(x,\Delta_-^n U^n(t,x,m)) - \sigma\Delta_2^n U^n(t,x,m) - f(m)(x) \\ &\quad - \sum_{y\in S^n} m_y\left(\frac{-\partial_p H_\uparrow(y,\Delta_+^n U^n(y,m))}{\Delta x_n} + \frac{\sigma}{\Delta x_n^2}\right)\Big(\partial_{m_{y+\Delta x_n}}U^n(x,m) - \partial_{m_y}U^n(x,m)\Big) \\ &\quad - \sum_{y\in S^n} m_y\left(\frac{\partial_p H_\downarrow(y,\Delta_-^n U^n(y,m))}{\Delta x_n} + \frac{\sigma}{\Delta x_n^2}\right)\Big(\partial_{m_{y-\Delta x_n}}U^n(x,m) - \partial_{m_y}U^n(x,m)\Big) = 0 \end{aligned} \tag{17}$$

The last two terms are equal to

$$-\int_{\mathbb{T}} m(dy)\frac{-\partial_p H_\uparrow(y,\Delta_+^n U^n(y,m))}{\Delta x_n}\Big(\partial_{m_{y+\Delta x_n}}U^n(x,m) - \partial_{m_y}U^n(x,m)\Big) \tag{18}$$

$$-\int_{\mathbb{T}} m(dy)\frac{\partial_p H_\downarrow(y,\Delta_-^n U^n(y,m))}{\Delta x_n}\Big(\partial_{m_{y-\Delta x_n}}U^n(x,m) - \partial_{m_y}U^n(x,m)\Big) \tag{19}$$

$$-\int_{\mathbb{T}} m(dy)\frac{\sigma}{\Delta x_n^2}\Big(\partial_{m_{y+\Delta x_n}}U^n(x,m) - 2\partial_{m_y}U^n(x,m) + \partial_{m_{y-\Delta x_n}}U^n(x,m)\Big) \tag{20}$$

assuming that $m = \sum_{j=1}^n m_j\delta_{x_j}$.

Let us remark that in this formulation, the Hamiltonian is not $\mathcal{C}^2$ as a function of $(p_+,p_-) = (u(x+\Delta x_n) - u(x), u(x-\Delta x_n) - u(x))$, thus the existence of a classical solution to the master equation is not entirely clear.

3.2. **Heuristic derivation of the master equation.** In this section, we justify the previous discretization. Ultimately, we want to show that

$$\lim_{n\to\infty} U^n(t,x^n,m^n) = U(t,x,m), \tag{21}$$

when $|x^n - x| + W_1(m^n,m) \to 0$. Let us assume this and show that, formally, the master equation (17) converges indeed to (2). We first have

$$\lim_n \Delta_\pm^n U^n(t,x,m) = \pm\partial_x U(t,x,m), \qquad \lim_n \Delta_2^n U^n(t,x,m) = \partial_{xx}U(t,x,m) \tag{22}$$

and thus the first terms in (17) converge to the corresponding one in (2).

We recall the definition of the measure derivative: for a function $U : \mathcal{P}(\mathbb{T}) \to \mathbb{R}$ the first derivative $\frac{\delta U}{\delta m}(m; y)$ is defined by the limit

$$\frac{\delta U}{\delta m}(m, x) = \lim_{\theta \to 0} \frac{U((1-\theta)m + \theta\delta_x) - U(m)}{\theta}. \tag{23}$$

Recalling that $U^n(t, x, m) \approx U(t, x, \sum_j m_j \delta_{x_j})$ we then have

$$\begin{aligned}\partial_{m_{y\pm\Delta x_n}} U^n(x, m) - \partial_{m_y} U^n(x, m) &= \int_{\mathbb{T}} \frac{\delta U}{\delta m}(x, m; z)(\delta_{y\pm\Delta x_n} - \delta_y)(dz) \\ &= \frac{\delta U}{\delta m}(x, m; y \pm \Delta x_n) - \frac{\delta U}{\delta m}(x, m; y)\end{aligned} \tag{24}$$

and hence

$$\lim_n \frac{\partial_{m_{y\pm\Delta x_n}} U^n(x, m) - \partial_{m_y} U^n(x, m)}{\Delta x_n} = \pm\partial_y \frac{\delta U}{\delta m}(x, m; y). \tag{25}$$

Therefore the terms in (18)-(19) give

$$\begin{aligned}&\approx -\int_{\mathbb{T}} m(dy) \left[-\partial_p H_\uparrow(y, \partial_x U^n(y, m))\partial_y \frac{\delta U}{\delta m}(x, m; y) - \partial_p H_\downarrow(y, -\partial_x U^n(y, m))\partial_y \frac{\delta U}{\delta m}(x, m; y)\right] \\ &= \int_{\mathbb{T}} m(dy)\partial_p H(y, \partial_x U(y, m))\partial_y \frac{\delta U}{\delta m}(x, m; y) = -\left\langle \partial_x(\partial_p H(\cdot, \partial_x U(\cdot, m))m), \frac{\delta U}{\delta m}(x, m; \cdot)\right\rangle.\end{aligned} \tag{26}$$

while the term in (20) yields

$$\begin{aligned}&-\sigma \int_{\mathbb{T}} m(dy) \frac{\frac{\delta U}{\delta m}(x, m; y + \Delta x_n) - 2\frac{\delta U}{\delta m}(x, m; y) + \frac{\delta U}{\delta m}(x, m; y - \Delta x_n)}{\Delta x_n^n} \\ &\to -\sigma \int_{\mathbb{T}} m(dy)\partial_y^2 \frac{\delta U}{\delta m}(x, m; y) = -\sigma\left\langle \partial_{xx} m, \frac{\delta U}{\delta m}(x, m; \cdot)\right\rangle.\end{aligned} \tag{27}$$

This provides the remaining terms in (2).

As far as convergence of the trajectories is concerned, we study it by means of the generators. By (15) we have

$$\begin{aligned}\mathcal{L}^n\varphi(x) &\approx -\partial_p H_\uparrow(x, \partial_x U(x, \mu_t))\partial_x\varphi(x) - \partial_p H(x, -\partial_x U(x, \mu_t))\partial_x\varphi(x) + \sigma\partial_{xx}\varphi(x) \\ &= -\partial_p H(x, \partial_x U(x, \mu_t))\partial_x\varphi(x) + \sigma\partial_{xx}\varphi(x),\end{aligned}$$

which is the generator of the limiting dynamics, providing the convergence in distribution of X^n to X.

3.3. **Second.** We provide here a different discretization. Consider a control $\alpha^n : [0, T] \times S^n \to \mathbb{R}$, the transition rates

$$\mathbb{P}(X^n_{t+\Delta t} = x^n_{i\pm1} | X^n_t = x^n_i) = \left(\pm\frac{\alpha^n(t, x^n_i)}{2\Delta x_n} + \frac{\sigma}{\Delta x_n^2}\right)\Delta t + o(\Delta t) \tag{28}$$

and the cost

$$J^n(\alpha^n, \mu^n) = \mathbb{E}\left[\int_0^T L(X^n_t, \alpha^n(t, X^n_t)) + f(\mu^n_t)(X^n_t)dt + g(\mu^n_T)(X^n_T)\right]. \tag{29}$$

Note that the rates are non-negative if

$$|\alpha(t, x)| \le \frac{2\sigma}{\Delta x_n} = 2\sigma n, \tag{30}$$

which should hold true for n large enough.

This formulation provides a smooth Hamiltonian, and is allowed because of the additional viscosity. Instead, if there is no Brownian motion in the limit dynamics (3) then we should consider the discretization in (7).

Here, denoting

$$\Delta^n u(x) = \frac{u(x + \Delta x_n) - u(x - \Delta x_n)}{2\Delta x_n} \tag{31}$$

we derive the HJB equation

$$-\partial_t u^n - \sigma\Delta_2^n u^n(x) + H(x, \Delta^n u^n(x)) = f(x, \mu_t^n) \tag{32}$$

and the optimal control is

$$\alpha^n(t,x) = -\partial_p H(x, \Delta^n u^n(t,x)). \tag{33}$$

Therefore the master equation becomes

$$\begin{aligned} &-\partial_t U^n(x,m) + H(x, \Delta^n U^n(x,m)) - \sigma\Delta_2^n U^n(x,m) - f(x,m) \\ &\quad + \sum_{y\in S^n} m_y \frac{\partial_p H(y, \Delta^n U^n(y,m))}{2\Delta x_n}\Big(\partial_{m_{y+\Delta x_n}} U^n(x,m) - \partial_{m_{y-\Delta x_n}} U^n(x,m)\Big) \\ &\quad - \sum_{y\in S^n} m_y \frac{\sigma}{\Delta x_n^2}\Big(\partial_{m_{y+\Delta x_n}} U^n(x,m) - 2\partial_{m_y} U^n(x,m) + \partial_{m_{y-\Delta x_n}} U^n(x,m)\Big) = 0. \end{aligned} \tag{34}$$

and we can formally see the convergence of the above equation to (2) as in §3.2.

4. Convergence results in the absence of a common noise

We prove convergence, with a convergence rate, of the discrete mater equation (17) to (2) and then of the related optimal trajectories.

This monotonicity of f and g implies that both (2) and (17) admit at most one classical solution. It also implies the uniqueness of solutions of the systems of characteristics such as (5) or the discrete system

$$\begin{aligned} &-\frac{d}{dt}u^n + H_\uparrow(x, \Delta_+^n u^n(x)) + H_\downarrow(x, \Delta_-^n u^n(x)) - \sigma\Delta_2^n u^n(x) = f(\mu_t^n)(x), \qquad t\in(s,T), x\in S^n, \\ &\frac{d}{dt}\mu^n(t,x) - \sigma\Delta_2^n\mu^n(t,x) - \Delta_+^n(\partial_p H_\uparrow(x, \Delta_+^n u^n(x))\mu^n(t,x)) \\ &\qquad + \Delta_-^n(\partial_p H_\downarrow(x, \Delta_-^n u^n(x))\mu^n(t,x)) = 0, \qquad t\in(s,T), x\in S^n, \\ &\qquad u^n(T,x) = g(\mu_T^n)(x); \mu^n(s) = \widetilde{\mu^n}, \end{aligned} \tag{35}$$

Lemma 1. *Let (u^n,μ^n) be a solution of either (35) for given $s\in[0,T], \widetilde{\mu^n}\in\mathcal{P}(S^n)$, this solution satisfies*

$$\sup_{x\in S^n}|u^n(s,x)| \le (T-s)(\|f\|_\infty + \sup_x|\inf_\alpha L(x,\alpha)|) + \|g\|_\infty. \tag{36}$$

If f and g are Lipschitz in x, uniformly in m, then there exists $M>0$ such that for any $s\in[0,T], \widetilde{\mu^n}\in\mathcal{P}(S^n), n\ge 1$ and $x\in S^n$

$$|\Delta_\pm^n u^n(s,x)| \le M. \tag{37}$$

Proof. Recall that u^n is the value function of an optimal control problem for the dynamics (7) (given μ^n). It is convenient to use stochastic open-loop controls for the control problem and thus to consider a probabilistic representation of the dynamics of X^n. We now introduce such a representation.

Let then $\mathcal{N}$ be a Poisson random measure on $[0,T]\times[0,\infty)^2$ with intensity measure $\nu(d\theta)$ on $[0,\infty)^2$ given by

$$\nu(E) = \ell(E\cap([0,\infty)\times\{0\})) + \ell(E\cap(\{0\}\times[0,\infty))),$$

where ℓ is the Lebesgue measure on $\mathbb{R}$. The measure ν is in fact the sum of the intersection with the axes and has the property that

$$\int_{[0,\infty)^2}\varphi(\theta_+,\theta_-)\nu(d\theta) = \int_0^\infty \varphi(\theta_+,0)d\theta_+ + \int_0^\infty \varphi(0,\theta_-)d\theta_-.$$

Consider then the dynamics

$$dX_t^n = \int_{[0,\infty)^2}\Big(\Delta x_n \mathbb{1}_{\left(0,\frac{\sigma}{\Delta x_n^2}+\frac{\alpha_+^n(t,X_t^n)}{2\Delta x_n}\right]}(\theta_+) - \Delta x_n \mathbb{1}_{\left(0,\frac{\sigma}{\Delta x_n^2}+\frac{\alpha_-^n(t,X_t^n)}{2\Delta x_n}\right]}(\theta_-)\Big)\mathcal{N}(d\theta,dt) \tag{38}$$

for a control $(\alpha_+^n(t,x), \alpha_-^n(t,x)$. We can show that the generator is given by (calling $\lambda(\alpha_+, \alpha_-, \theta)$ the integrand above)

$$\int_{[0,\infty)^2} [\varphi(x + \lambda(\alpha_+(t,x), \alpha_-(t,x), \theta)) - \varphi(x)]\, \nu(d\theta) = \Delta_+^n \varphi(x) \alpha_+(t,x) + \Delta_-^n \varphi(x) \alpha_-(t,x) + \sigma \Delta_2^n \varphi(x),$$

which ensures that X^n has the transition rates as in (7). The advantage in using the representation (38) is that it permits to use stochastic open-loop controls, which are predictable stochastic processes.

Therefore we have (recall that $(\mu_t^n)_{t_0 \le t \le T}$ is fixed)

$$u^n(t_0, x) = \inf_{\alpha^n} \mathbb{E}\left[\int_{t_0}^T L(X_t^n, (\alpha_+^n)_t) + L(X_t^n, (\alpha_-^n)_t) - L(X_t^n, 0) + f(X_t^n, \mu_t^n) dt + g(X_T^n, \mu_T^n) \right],$$

where X^n starts at $X_{t_0}^n = x$ and uses the stochastic control (α_+^n, α_-^n). Then we derive immediately the bound (36) by taking $(\alpha_+^n, \alpha_-^n)_t$ which minimizes the function $L(X_t^n, \alpha_+) + L(X_t^n, \alpha_-)$. As to (37), if $(\alpha_+^n, \alpha_-^n)_t$ is optimal for x (that is given by (13)) and we denote by X^n the process starting at x and by $\widetilde{X}^n$ the process starting at $x + \Delta x_n$, both with the control $(\alpha_+^n, \alpha_-^n)_t$, then

$$\begin{aligned}
& u(t_0, x + \Delta x_n) - u(t_0, x) \\
& \le \mathbb{E}\left[\int_{t_0}^T L(\widetilde{X}_t^n, (\alpha_+^n)_t) + L(\widetilde{X}_t^n, (\alpha_-^n)_t) - L(\widetilde{X}_t^n, 0) + f(\widetilde{X}_t^n, \mu_t^n) dt + g(\widetilde{X}_T^n, \mu_T^n) \right] \\
& \quad - \mathbb{E}\left[\int_{t_0}^T L(X_t^n, (\alpha_+^n)_t) + L(X_t^n, (\alpha_-^n)_t) - L(X_t^n, 0) + f(X_t^n, \mu_t^n) dt + g(X_T^n, \mu_T^n) \right] \\
& \le M \sup_{t_0 \le t \le T} \mathbb{E}|\widetilde{X}_t^n - X_t^n|,
\end{aligned}$$

where we have used the regularity of L in its first variable. We conclude by noticing that $\widetilde{X}_t^n - X_t^n = \Delta x_n$ for any t, because all the other terms cancel. By changing the roles of x and $x + \Delta x_n$, we obtain the opposite inequality and hence (37) follows. ☐

Clearly this estimate translates directly to the solution of the master equation thanks to its representation by the characteristics.

4.1. **Classical solutions.** We first prove convergence, with a convergence rate, in cases in which (2) and (17) admit a classical solution.

Proposition 2. *If U is the classical solution to* (2) *then $V^n(t,x,m) := U(t, x, \sum_{j=1}^n m_j \delta_{x_j})$ solves*

$$\begin{aligned}
& -\partial_t V^n(x,m) + H_\uparrow(x, \Delta_+^n V^n(x,m)) + H_\downarrow(x, \Delta_-^n V^n(x,m)) - \sigma \Delta_2^n V^n(x,m) - f(x,m) \\
& \quad + \sum_{y \in S^n} m_y \frac{\partial_p H_\uparrow(u, \Delta_+^n V^n(y,m))}{\Delta x_n} \Big(\partial_{m_{y+\Delta x_n}} V^n(x,m) - \partial_{m_y} V^n(x,m) \Big) \\
& \quad + \sum_{y \in S^n} m_y \frac{-\partial_p H_\downarrow(u, \Delta_-^n V^n(y,m))}{\Delta x_n} \Big(\partial_{m_{y-\Delta x_n}} V^n(x,m) - \partial_{m_y} V^n(x,m) \Big) \\
& \quad - \sum_{y \in S^n} m_y \frac{\sigma}{\Delta x_n^2} \Big(\partial_{m_{y+\Delta x_n}} V^n(x,m) - 2\partial_{m_y} V^n(x,m) + \partial_{m_{y-\Delta x_n}} V^n(x,m) \Big) = r^n(t,x,m),
\end{aligned} \tag{39}$$

with $|r^n(t,x,m)| \le C\omega(\frac{1}{n})$, where ω is a modulus of continuity of $\partial_x U$, $\partial_{xx} U$, $\frac{\delta U}{\delta m}$, $\partial_y \frac{\delta U}{\delta m}(\cdot, y)$ and $\partial_{yy} \frac{\delta U}{\delta m}(\cdot, y)$.

Proof. The proof of this statement follows from the fact that assuming this regularity on U, all the heuristics of section 3.2 are true with a modulus of convergence driven by ω. ☐

Theorem 3. *Let U be a classical solution to* (2) *and U^n be a classical solution to* (17)*. There exists a constant C (independent of n) such that, for V^n defined as in the previous result,*

$$|U^n(t,x,m) - V^n(t,x,m)| \leq C\omega\left(\frac{1}{n}\right), \qquad \forall t \in [0,T], x \in S^n, m \in \mathcal{P}(S_n) \tag{40}$$

$$\mathbb{E}\int_0^T |\Delta_+^n(U^n - V^n)(t, X_t^n, \mathrm{Law}(X_t^n))|^2 + |\Delta_-^n(U^n - V^n)(t, X_t^n, \mathrm{Law}(X_t^n))|^2 dt \leq C\omega^2\left(\frac{1}{n}\right) \tag{41}$$

where X_t^n is the optimal process of the MFG (28)-(29)*.*

Proof. Consider any initial time $t_0 \in [0,T)$ and distribution $\mu_0 = (\mu_{0,y})_{y\in S^n} \in \mathcal{P}(S^n)$ such that $\mu_{0,y} \neq 0$ for each $y \in S^n$, i.e. μ_0 belongs to the interior of the simplex, and let $(X_t^n)_{t_0\leq t\leq T}$ the optimal process of the MFG starting at (t_0,μ_0). Denote its law by $\left(\mu_t^n = \mathrm{Law}(X_t^n)\right)_{t_0\leq t\leq T}$. We denote $W^n = U^n - V^n$, $W_t^n = U_t^n - V_t^n = (U^n - V^n)(t, X_t^n, \mu_t^n)$ and expand $|U^n - V^n|^2(t, X_t^n, \mu_t^n)$. For any $t \in [t_0,T]$, Itô formula and then conditional expectation with respect to the initial condition (denoted $\mathbb{E}_0$) give

$$\begin{aligned}
\mathbb{E}_0|W_T^n|^2 - \mathbb{E}_0|W_t^n|^2 = \mathbb{E}_0\int_t^T \Bigg[&\left(|W^n(X_s^n + \Delta x_n, \mu_s^n)|^2 - |W_s^n|^2\right)\left(\frac{\sigma}{\Delta x_n^2} - \frac{\partial_p H_\uparrow(X_s^n, \Delta_+^n U_s^n)}{\Delta x_n}\right)\\
&+ \left(|W^n(X_s^n - \Delta x_n, \mu_s^n)|^2 - |W_s^n|^2\right)\left(\frac{\sigma}{\Delta x_n^2} + \frac{\partial_p H_\downarrow(X_s^n, \Delta_-^n U_s^n)}{\Delta x_n}\right)\\
&+ 2W_s^n(\partial_t U_s^n - \partial_t V_s^n)\\
&+ 2W_s^n(D^m U^n - D^m V^n)(X_s^n, \mu_s^n)\cdot\frac{d}{dt}\mu_s^n\Bigg] ds\\
= \mathbb{E}_0\int_t^T \Bigg[&\left(|W^n(X_s^n + \Delta x_n, \mu_s^n) - W_s^n|^2 + 2W_s^n(W^n(X_s^n + \Delta x_n, \mu_s^n) - W_s^n)\right)\left(\frac{\sigma}{\Delta x_n^2} - \frac{\partial_p H_\uparrow(X_s^n, \Delta_+^n U_s^n)}{\Delta x_n}\right)\\
&+ \left(|W^n(X_s^n - \Delta x_n, \mu_s^n) - W_s^n|^2 + 2W_s^n(W^n(X_s^n - \Delta x_n, \mu_s^n) - W_s^n)\right)\left(\frac{\sigma}{\Delta x_n^2} + \frac{\partial_p H_\downarrow(X_s^n, \Delta_-^n U_s^n)}{\Delta x_n}\right)\\
&+ 2W_s^n\Bigg(H_\uparrow(X_s^n, \Delta_+^n U_s^n) + H_\downarrow(X_s^n, \Delta_-^n U_s^n) - H_\uparrow(X_s^n, \Delta_+^n V_s^n) - H_\downarrow(X_s^n, \Delta_-^n V_s^n) - \sigma\Delta_2^n U_s^n + \sigma\Delta_2^n V_s^n\\
&\quad + \sum_{y\in S^n}\mu_{s,y}^n\frac{\partial_p H_\uparrow(y, \Delta_+^n U^n(y,\mu_s^n))}{\Delta x_n}\left(\partial_{m_{y+\Delta x_n}}U_s^n - \partial_{m_y}U_s^n\right)\\
&\qquad - \sum_{y\in S^n}\mu_{s,y}^n\frac{\partial_p H_\uparrow(y, \Delta_+^n V^n(y,\mu_s^n))}{\Delta x_n}\left(\partial_{m_{y+\Delta x_n}}V_s^n - \partial_{m_y}V_s^n\right)\\
&\quad - \sum_{y\in S^n}\mu_{s,y}^n\frac{\partial_p H_\downarrow(y, \Delta_-^n U^n(y,\mu_s^n))}{\Delta x_n}\left(\partial_{m_{y-\Delta x_n}}U_s^n - \partial_{m_y}U_s^n\right)\\
&\qquad + \sum_{y\in S^n}\mu_{s,y}^n\frac{\partial_p H_\downarrow(y, \Delta_-^n V^n(y,\mu_s^n))}{\Delta x_n}\left(\partial_{m_{y-\Delta x_n}}V_s^n - \partial_{m_y}V_s^n\right)\\
&\quad - \sum_{y\in S^n}\mu_{s,y}^n\frac{\sigma}{\Delta x_n^2}\left(\partial_{m_{y+\Delta x_n}}U_s^n - 2\partial_{m_y}U_s^n + \partial_{m_{y-\Delta x_n}}U_s^n\right)\\
&\qquad + \sum_{y\in S^n}\mu_{s,y}^n\frac{\sigma}{\Delta x_n^2}\left(\partial_{m_{y+\Delta x_n}}V_s^n - 2\partial_{m_y}V_s^n + \partial_{m_{y-\Delta x_n}}V_s^n\right)\Bigg) + r_s^n\\
&+ 2W_s^n\Bigg(- \sum_{y\in S^n}\mu_{s,y}^n\frac{\partial_p H_\uparrow(y, \Delta_+^n U^n(y,\mu_s^n))}{\Delta x_n}\left(\partial_{m_{y+\Delta x_n}}U_s^n - \partial_{m_y}U_s^n\right)\\
&\qquad + \sum_{y\in S^n}\mu_{s,y}^n\frac{\partial_p H_\downarrow(y, \Delta_-^n U^n(y,\mu_s^n))}{\Delta x_n}\left(\partial_{m_{y-\Delta x_n}}U_s^n - \partial_{m_y}U_s^n\right)
\end{aligned}$$

$$
\begin{aligned}
&\qquad + \sum_{y\in S^n} \mu^n_{s,y} \frac{\partial_p H_\uparrow(y, \Delta^n_+ U^n(y,\mu^n_s))}{\Delta x_n} \Big(\partial_{m_{y+\Delta x_n}} V^n_s - \partial_{m_y} V^n_s\Big)\\
&\qquad - \sum_{y\in S^n} \mu^n_{s,y} \frac{\partial_p H_\downarrow(y, \Delta^n_- U^n(y,\mu^n_s))}{\Delta x_n} \Big(\partial_{m_{y-\Delta x_n}} V^n_s - \partial_{m_y} V^n_s\Big)\\
&\quad + \sum_{y\in S^n} \mu^n_{s,y} \frac{\sigma}{\Delta x_n^2} \Big(\partial_{m_{y+\Delta x_n}} U^n_s - 2\partial_{m_y} U^n_s + \partial_{m_{y-\Delta x_n}} U^n_s\Big)\\
&\qquad - \sum_{y\in S^n} \mu^n_{s,y} \frac{\sigma}{\Delta x_n^2} \Big(\partial_{m_{y+\Delta x_n}} V^n_s - 2\partial_{m_y} V^n_s + \partial_{m_{y-\Delta x_n}} V^n_s\Big)\Big)\Big] ds\\
&= \mathbb{E}_0 \int_t^T \Big[|\Delta^n_+ W^n_s|^2 (\sigma - \Delta x_n \partial_p H_\uparrow(X^n_s, \Delta^n_+ U^n_s)) + |\Delta^n_- W^n_s|^2 (\sigma + \Delta x_n \partial_p H_\downarrow(X^n_s, \Delta^n_- U^n_s))\\
&\quad + 2W^n_s(\sigma \Delta^n_2 W^n_s - \partial_p H_\uparrow(X^n_s, \Delta^n_+ U^n_s)\Delta^n_+ W^n_s + \partial_p H_\downarrow(X^n_s, \Delta^n_- U^n_s)\Delta^n_- W^n_s)\\
&\quad + 2W^n_s\Big(H_\uparrow(X^n_s, \Delta^n_+ U^n_s) + H_\downarrow(X^n_s, \Delta^n_- U^n_s) - H_\uparrow(X^n_s, \Delta^n_+ V^n_s) - H_\downarrow(X^n_s, \Delta^n_- V^n_s) - \sigma \Delta^n_2 W^n_s + r^n_s\\
&\quad + \sum_{y\in S^n} \mu^n_{s,y} (\partial_p H_\uparrow(y, \Delta^n_+ U^n(y,\mu^n_s)) - \partial_p H_\uparrow(y, \Delta^n_+ V^n(y,\mu^n_s))) \frac{\partial_{m_{y+\Delta x_n}} V^n_s - \partial_{m_y} V^n_s}{\Delta x_n}\\
&\quad - \sum_{y\in S^n} \mu^n_{s,y} (\partial_p H_\downarrow(y, \Delta^n_- U^n(y,\mu^n_s)) - \partial_p H_\downarrow(y, \Delta^n_- V^n(y,\mu^n_s))) \frac{\partial_{m_{y-\Delta x_n}} V^n_s - \partial_{m_y} V^n_s}{\Delta x_n}\Big)\Big] ds\\
&= \mathbb{E}_0 \int_t^T \Big[|\Delta^n_+ W^n_s|^2 (\sigma - \Delta x_n \partial_p H_\uparrow(X^n_s, \Delta^n_+ U^n_s)) + |\Delta^n_- W^n_s|^2 (\sigma + \Delta x_n \partial_p H_\downarrow(X^n_s, \Delta^n_- U^n_s))\\
&\quad + 2W^n_s\Big(- \partial_p H_\uparrow(X^n_s, \Delta^n_+ U^n_s)\Delta^n_+ W^n_s + H_\uparrow(X^n_s, \Delta^n_+ U^n_s) - H_\uparrow(X^n_s, \Delta^n_+ V^n_s)\\
&\qquad + \partial_p H_\downarrow(X^n_s, \Delta^n_- U^n_s)\Delta^n_- W^n_s + H_\downarrow(X^n_s, \Delta^n_- U^n_s) - H_\downarrow(X^n_s, \Delta^n_- V^n_s) + r^n_s\\
&\quad + \sum_{y\in S^n} \mu^n_{s,y} (\partial_p H(y, \Delta^n U^n(y,\mu^n_s)) - \partial_p H(y, \Delta^n V^n(y,\mu^n_s))) \frac{\partial_{m_{y+\Delta x_n}} V^n_s - \partial_{m_{y-\Delta x_n}} V^n_s}{2\Delta x_n}\Big)\Big] ds.
\end{aligned}
$$

Since $|\Delta^n_\pm U^n_s| \le M$, we have

$$
\sigma \pm \Delta x_n \partial_p H_{\downarrow\uparrow}(X^n_s, \Delta^n_\mp U^n_s) \ge \frac{\sigma}{2} \qquad \text{if } \Delta x_n \le \frac{1}{C},
$$

for C depending on M and H. As $W_T = 0$, using the convexity inequality $AB \le \varepsilon A^2 + \frac{1}{4\varepsilon} B^2$ and the bounds

$$
\left| \frac{\partial_{m_{y\pm\Delta x_n}} V^n_s - \partial_{m_y} V^n_s}{\Delta x_n} \mp D^m U(X^n_s, \mu^n_s) \right| \le \omega\Big(\frac{1}{n}\Big)
$$

and $|D^m U| \le C$, as well as the fact that $H_\uparrow$ and $H_\downarrow$ are locally Lipschitz we obtain

$$
\begin{aligned}
&\mathbb{E}_0 |W^n_t|^2 + \frac{\sigma}{2} \mathbb{E}_0 \int_t^T \Big(|\Delta^n_+ W^n_s|^2 + |\Delta^n_- W^n_s|^2\Big) ds\\
&\quad \le C \mathbb{E}_0 \int_t^T |W^n_s| \Big(|\Delta^n_+ W^n_s| + |\Delta^n_- W^n_s| + \sum_{y\in S^n} \mu^n_{s,y} |\partial_p H_\uparrow(y, \Delta^n_+ U^n(y,\mu^n_s)) - \partial_p H_\uparrow(y, \Delta^n V^n_+(y,\mu^n_s))|\\
&\qquad\qquad \sum_{y\in S^n} \mu^n_{s,y} |\partial_p H_\downarrow(y, \Delta^n_- U^n(y,\mu^n_s)) - \partial_p H_\downarrow(y, \Delta^n_- V^n(y,\mu^n_s))| + |r^n_s| + \omega\Big(\frac{1}{n}\Big)\Big) ds\\
&\quad \le C \mathbb{E}_0 \int_t^T |W^n_s| \Big(|\Delta^n_+ W^n_s| + |\Delta^n_- W^n_s| + \omega\Big(\frac{1}{n}\Big)\Big) ds\\
&\quad \le C \mathbb{E}_0 \int_t^T |W^n_s|^2 ds + \frac{\sigma}{4C} \mathbb{E}_0 \int_t^T \Big(|\Delta^n_+ W^n_s|^2 + |\Delta^n_- W^n_s|^2\Big) ds + C\omega^2\Big(\frac{1}{n}\Big).
\end{aligned}
$$

This gives

$$\mathbb{E}_0|W_t^n|^2 + \frac{\sigma}{4}\mathbb{E}_0\int_t^T \Big(|\Delta_+^n W_s^n|^2 + |\Delta_-^n W_s^n|^2\Big)\,ds \le C\mathbb{E}_0\int_t^T |W_s^n|^2 ds + C\omega^2\Big(\frac{1}{n}\Big) \tag{42}$$

and thus Gronwall's inequality yields

$$\sup_{t\in[t_0,T]} \mathbb{E}_0|W_t^n|^2 \le C\omega^2\Big(\frac{1}{n}\Big) \qquad \mathbb{P}-a.s. \tag{43}$$

At $t = t_0$, the above inequality gives

$$|U^n(t_0, X_{t_0}, \mu_0) - V^n(t_0, X_{t_0}, \mu_0)| \le C\omega\Big(\frac{1}{n}\Big) \qquad \mathbb{P}-a.s., \tag{44}$$

which, since $\mathrm{Law}(X_{t_0}) = \mu_0$ is supported on the entire S^n, provides

$$|U^n(t_0, x, \mu_0) - V^n(t_0, x, \mu_0)| \le C\omega\Big(\frac{1}{n}\Big) \tag{45}$$

for any $t_0 \in [0, T]$, $x \in S^n$ and μ_0 in the interior of $\mathcal{P}(S^n)$. Since U^n and V^n are continuous in the measure argument, the above inequality holds for any $\mu \in \mathcal{P}(S^n)$, which provides (40), but only for $n \ge 4M$; changing the value of the constant, (40) holds for any n.

Finally, letting $t_0 = 0$, applying (43) into (42) and taking the expectation, we obtain (41). □

We now turn to the convergence of the trajectories at equilibrium. Consider an initial distribution (at time 0) m_0 of the limit MFG, and a random variable ξ (with values in $\mathbb{T}$) with $\mathrm{Law}(\xi) = m_0$. For the discretization, let $E_i^n = [x_i^n - \frac{1}{2n}, x_i^n - \frac{1}{2n})$ and

$$m_0^n = \sum_{i=1}^n m_0(E_i^n)\delta_{x_i^n}, \qquad \xi^n = \sum_{i=1}^n x_i^n \mathbb{1}_{\{\xi\in E_i^n\}}. \tag{46}$$

We have $\mathrm{Law}(\xi^n) = m_0^n$ and $\sqrt[k]{\mathbb{E}|\xi_n - \xi|^k} \le \frac{1}{2n}$ for any integer $k \ge 1$.

Let X^n be the trajectory at equilibrium for the discrete MFG with initial condition ξ^n. Hence the control of the players is given by $\alpha_+^n(t,x) = -\partial_p H_\uparrow(x, \Delta_+^n U^n(t, x, \mathrm{Law}(X_t^n)))$ (and similarly for α_-^n), where U^n is the classical solution to (17). Let also X be the optimal process for the limit MFG (5) with initial condition ξ. The associated control is thus given by $\alpha(t,x) = -\partial_p H(x, \partial_x U(t, x, \mathrm{Law}(X_t)))$, where U is the classical solution of (2). Let $\mathcal{D}([0,T],\mathbb{T})$ be the space of càdlàg functions endowed with the Skorokhod J_1 topology, and recall that W_1 is the 1-Wasserstein distance on $\mathcal{P}(\mathbb{T})$.

Theorem 4 (Convergence of trajectories)**.** *We have*

$$\lim_n X^n = X \qquad \text{in law in } \mathcal{D}([0,T],\mathbb{T}). \tag{47}$$

As a consequence

$$\lim_n \sup_{0\le t\le T} W_1(\mathrm{Law}(X_t^n), \mathrm{Law}(X_t)) = 0 \tag{48}$$

Proof. Let $\overline{X}^n$ and $\widetilde{X}^n$ be the processes starting at ξ^n, with dynamics given by (7), with controls therein given by $(\overline{\alpha}_+^n(t,x), \overline{\alpha}_-^n(t,x)) = (-\partial_p H_\uparrow(x, \Delta_+^n U(t, x, \mathrm{Law}(\overline{X}_t^n))), \partial_p H_\downarrow(x, \Delta_-^n U(t, x, \mathrm{Law}(\overline{X}_t^n))))$ and $(\widetilde{\alpha}_+^n(t,x), \widetilde{\alpha}_-^n(t,x)) = (-\partial_p H_\uparrow(x, \partial_x U(t, x, \mathrm{Law}(X_t))), \partial_p H_\downarrow(x, \partial_x U(t, x, \mathrm{Law}(X_t))))$ respectively. The SDE representation (38), applying then (41), Jensen's inequality and the Lipschitz continuity of $\partial_x U$ in x and m (in W_1) and recalling that $W_1(\mathrm{Law}(X), \mathrm{Law}(Y)) \le \mathbb{E}|X - Y|$, give

$$\begin{aligned}
\mathbb{E}\sup_{0\le s\le t}|X_s^n - \overline{X}_s^n| &\le \mathbb{E}\int_0^t |\partial_p H_\uparrow(x, \Delta_+^n U^n(s, X_s^n, \mathrm{Law}(X_s^n))) - \partial_p H_\uparrow(x, \Delta_+^n U(s, \overline{X}_s^n, \mathrm{Law}(\overline{X}_s^n)))| \\
&\quad + |\partial_p H_\downarrow(x, \Delta_-^n U^n(s, X_s^n, \mathrm{Law}(X_s^n))) - \partial_p H_\downarrow(x, \Delta_-^n U(s, \overline{X}_s^n, \mathrm{Law}(\overline{X}_s^n)))|ds \\
&\le \mathbb{E}\int_0^t |\partial_p H_\uparrow(x, \Delta_+^n U^n(s, X_s^n, \mathrm{Law}(X_s^n))) - \partial_p H_\uparrow(x, \Delta_+^n U(s, X_s^n, \mathrm{Law}(X_s^n)))| \\
&\quad + |\partial_p H_\downarrow(x, \Delta_-^n U^n(s, X_s^n, \mathrm{Law}(X_s^n))) - \partial_p H_\downarrow(x, \Delta_-^n U(s, X_s^n, \mathrm{Law}(X_s^n)))|ds
\end{aligned}$$

$$
\begin{aligned}
&+\mathbb{E}\int_0^t |\partial_p H_\uparrow(x,\Delta_+^n U(s,X_s^n,\mathrm{Law}(X_s^n))) - \partial_p H_\uparrow(x,\Delta_+^n U(s,\overline{X}_s^n,\mathrm{Law}(\overline{X}_s^n)))| \\
&+ |\partial_p H_\downarrow(x,\Delta_-^n U(s,X_s^n,\mathrm{Law}(X_s^n))) - \partial_p H_\downarrow(x,\Delta_-^n U(s,\overline{X}_s^n,\mathrm{Law}(\overline{X}_s^n)))| ds \\
\le{}& C\omega\left(\frac{1}{n}\right) + C\mathbb{E}\int_0^t \sup_{0\le r\le s}|X_r^n - \overline{X}_r^n| ds,
\end{aligned}
$$

and therefore Gronwall's lemma yields

$$
\mathbb{E}\sup_{0\le t\le T}|X_t^n - \overline{X}_t^n| \le C\omega\left(\frac{1}{n}\right). \tag{49}
$$

Similarly, recalling that $||\Delta_\pm^n U \mp \partial_x U||_\infty \le \omega(\frac{1}{n})$, we have

$$
\begin{aligned}
\mathbb{E}\sup_{0\le s\le t}|\widetilde{X}_s^n - \overline{X}_s^n| \le{}& \mathbb{E}\int_0^t |\partial_p H_\uparrow(x,\partial_x U(s,\widetilde{X}_s^n,\mathrm{Law}(X_s))) - \partial_p H_\uparrow(x,\Delta_+^n U(s,\overline{X}_s^n,\mathrm{Law}(\overline{X}_s^n)))| \\
& |\partial_p H_\downarrow(x,\partial_x U(s,\widetilde{X}_s^n,\mathrm{Law}(X_s))) - \partial_p H_\downarrow(x,\Delta_-^n U(s,\overline{X}_s^n,\mathrm{Law}(\overline{X}_s^n)))| ds \\
\le{}& C\omega\left(\frac{1}{n}\right) + C\mathbb{E}\int_0^t |\widetilde{X}_s^n - \overline{X}_s^n| + W_1(\mathrm{Law}(\widetilde{X}_s^n),\mathrm{Law}(\overline{X}_s^n)) + W_1(\mathrm{Law}(\widetilde{X}_s^n),\mathrm{Law}(X_s))ds \\
\le{}& C\omega\left(\frac{1}{n}\right) + C\sup_{0\le t\le T} W_1(\mathrm{Law}(\widetilde{X}_t^n),\mathrm{Law}(X_t)) + C\mathbb{E}\int_0^t \sup_{0\le r\le s}|\widetilde{X}_r^n - \overline{X}_r^n| ds
\end{aligned}
$$

and thus, applying Gronwall's inequality, we get

$$
\mathbb{E}\sup_{0\le t\le T}|\widetilde{X}_t^n - \overline{X}_t^n| \le C\omega\left(\frac{1}{n}\right) + C\sup_{0\le t\le T} W_1(\mathrm{Law}(\widetilde{X}_t^n),\mathrm{Law}(X_t)),
$$

which, together with (49), implies

$$
\mathbb{E}\sup_{0\le t\le T}|\widetilde{X}_t^n - X_t^n| \le C\omega\left(\frac{1}{n}\right) + C\sup_{0\le t\le T} W_1(\mathrm{Law}(\widetilde{X}_t^n),\mathrm{Law}(X_t)). \tag{50}
$$

Now, if we show that $\widetilde{X}^n \to X$ in law in $\mathcal{D}([0,T],\mathbb{T})$ then, as a consequence,

$$
\lim_n \sup_{0\le t\le T} W_1(\mathrm{Law}(\widetilde{X}_t^n),\mathrm{Law}(X_t)) = 0
$$

and thus the claim (47) follows from (50).

The convergence $\widetilde{X}^n \to X$ is given by the convergence of the generators (plus convergence in law of the initial conditions), noticing that there is no dependence on the law in $\widetilde{X}^n$ and hence the depencence on $\mathrm{Law}(X_t)$ in the two processes can be treated as a dependence on time. Indeed, for any function $\varphi \in \mathcal{C}^2(\mathbb{T})$, the generator of $\widetilde{X}^n$ is (for $x \in S^n$)

$$
\begin{aligned}
\mathcal{L}_t^n\varphi(x) ={}& \left(-\frac{\partial_p H_\uparrow(x,\partial_x U(t,x,\mathrm{Law}(X_t)))}{\Delta x_n} + \frac{\sigma}{\Delta x_n^2}\right)[\varphi(x+\Delta x_n) - \varphi(x)] \\
&+ \left(\frac{\partial_p H_\downarrow(x,\partial_x U(t,x,\mathrm{Law}(X_t)))}{\Delta x_n} + \frac{\sigma}{\Delta x_n^2}\right)[\varphi(x-\Delta x_n) - \varphi(x)] \\
={}& -\partial_p H_\uparrow(x,\partial_x U(t,x,\mathrm{Law}(X_t)))\Delta_+^n\varphi(x) + \partial_p H_\downarrow(x,\partial_x U(t,x,\mathrm{Law}(X_t)))\Delta_-^n\varphi(x) + \sigma\Delta_2^n\varphi(x) \\
\to{}& -\partial_p H(x,\partial_x U(t,x,\mathrm{Law}(X_t)))\partial_x\varphi(x) + \sigma\partial_{xx}\varphi(x),
\end{aligned}
$$

which is the generator of the process X. □

4.2. **Convergence through MFG system.** Without assuming regularity on f and g, convergence results can still be established, namely by using the MFG system, which represents the characteristic curves of the master equation. This analysis is concerned with the first discretization (17).

Theorem 5. *Assume that f and g are monotone and W_1-Lipschitz continuous, and f is bounded in $\mathcal{C}^\gamma(\mathbb{T})$ and g in $\mathcal{C}^{2+\gamma}(\mathbb{T})$, uniformly in m, for a $\gamma \in (0,1)$. Then there exists $\lambda \in (0,1)$ such that, for any n*

$$|U^n(t,x,m) - U(t,x,m)| \leq \frac{C}{n^\lambda}, \qquad \forall t \in [0,T], x \in S^n, m \in \mathcal{P}(S_n). \tag{51}$$

Moreover, let X^n be the state process of players which plays optimally at the equilibrium in the MFG (7)-(8), with initial distribution m_0^n at $t=0$, let X be the state process of a player which plays optimally at equilibrium in the limit MFG (3)-(4), with initial condition m_0 at $t=0$. Then $W_1(m_0^n, m_0) \leq \frac{C}{n}$ implies

$$\sup_{0 \leq t \leq T} W_1(\mathrm{Law}(X_t^n), \mathrm{Law}(X_t)) \leq \frac{C}{n^\lambda} \tag{52}$$

and further

$$\lim_n X^n = X \qquad \textit{in law in } \mathcal{D}([0,T],\mathbb{T}). \tag{53}$$

Proof. Fix m_0 and m_0^n such that $W_1(m_0^n, m_0) \leq \frac{C}{n}$ and an initial time t_0. We consider a solution (u,m) of the MFG system (5) starting at (t_0, m_0), and denote by X the corresponding optimal process given by (3) with $\alpha(t,x) = -\partial_p H(x, \partial_x u(t,x))$ therein.

Step 1. The assumption of the space regularity of f and g gives that $u \in \mathcal{C}^{2+\gamma'}(\mathbb{T})$ in space, uniformly in time, for a $\gamma' \in (0,1)$. Let $\widehat{X}^n$ be the process given by (7) with $\alpha_+(t,x) = -\partial_p H_\uparrow(x, \partial_x u(t,x)); \alpha_-(t,x) = \partial_p H_\downarrow(x, -\partial_x u(t,x))$, and $\widetilde{X}^n$ be given by (7) with $\alpha_+(t,x) = -\partial_p H_\uparrow(x, \Delta_+^n u(t,x)); \alpha_-(t,x) = \partial_p H_\downarrow(x, \Delta_-^n u(t,x))$. Since $\partial_x u$ is Lipschitz in $\mathbb{T}$, we derive

$$\mathbb{E}\Big[\sup_{t_0 \leq t \leq T} |\widetilde{X}^n - \widehat{X}_t^n|\Big] \leq \frac{C}{n}. \tag{54}$$

Convergence in law of $\widehat{X}^n$ to X, in $\mathcal{D}([0,T],\mathbb{T})$ holds by convergence of the generators. From Lemma 6 below, we have

$$\lim_n \widetilde{X}^n = X \qquad \text{in law in } \mathcal{D}([0,T],\mathbb{T}) \tag{55}$$

and

$$\sup_{0 \leq t \leq T} W_1(\mathrm{Law}(\widetilde{X}_t^n), \mathrm{Law}(X_t)) \leq \frac{C}{n^\lambda}. \tag{56}$$

Since $u(t,\cdot) \in \mathcal{C}^{2+\gamma'}(\mathbb{T})$, using (56), the Lipschitz-continuity of f and g, and (5), we obtain

$$\begin{aligned} &-\partial_t u + H_\uparrow(x, \Delta_+^n u(x)) + H_\downarrow(x, \Delta_-^n u(x)) - \sigma \Delta_2^n u(x) = f(x, \widetilde{\mu}_t^n) + r^n(t,x), \qquad x \in S^n, \\ &u(T,x) = g(x, \widetilde{\mu}_t^n) + r^n(T,x), \end{aligned} \tag{57}$$

with $|r^n(t,x)| \leq \frac{C}{n^\lambda}$ (for a possibly new value of λ), whereas $\widetilde{\mu}_t^n = \mathrm{Law}(\widetilde{X}_t^n)$. Thus (u,m) almost solves the finite MFG system (35).

Step 2. Let (u^n, μ^n) be the solutions to the MFG (35) starting at (t_0, m_0^n) and X^n the associated state process. Denote the initial random distribution on the state by ξ^n. We stress that this is the same initial condition as $\widetilde{X}^n$. Thanks to (57), u (restricted to S^n) can be seen as the value function of a control problem with dynamics (7) and cost

$$\begin{aligned} \widetilde{J}^n(\alpha, \widetilde{\mu^n}) = \mathbb{E}\Big[\int_0^T & L(X_t^n, \alpha_+^n(t, X_t^n)) + L(X_t^n, \alpha_-^n(t, X_t^n)) + f(X_t^n, \widetilde{\mu}_t^n) + r^n(t, X_t^n) dt \\ & + g(X_T^n, \mu_T^n) + r^n(T, X_T^n)\Big]. \end{aligned} \tag{58}$$

We first compute u^n on $\widetilde{X}^n$: denoting $\widetilde{\alpha}_+^n(t,x) = -\partial_p H_\uparrow(x, \Delta_+^n u(t,x)), \widetilde{\alpha}_-^n(t,x) = \partial_p H_\downarrow(x, \Delta_-^n u(t,x))$ and $\alpha_+^n(t,x) = -\partial_p H_\uparrow(x, \Delta_+^n u^n(t,x)), \alpha_-^n(t,x) = \partial_p H_\downarrow(x, \Delta_-^n u^n(t,x))$,

$$\mathbb{E}[u^n(t_0, \xi^n)] = \mathbb{E}[u^n(T, \widetilde{X}_T^n)] + \mathbb{E}\int_{t_0}^T \left(-\partial_t u^n - \sigma \Delta_2^n u^n - \widetilde{\alpha}_+^n \Delta_+^n u^n - \widetilde{\alpha}_-^n \Delta_-^n u^n\right)(s, \widetilde{X}_s^n) ds$$

$$= \mathbb{E}\Big[g(\widetilde{X}^n_T,\mu^n_T) + \int_{t_0}^T \left(-H_\uparrow(x,\Delta^n_+ u^n) - H_\downarrow(x,\Delta^n_- u^n) - \widetilde{\alpha}^n_+\Delta^n_+ u^n - \widetilde{\alpha}^n_-\Delta^n_- u^n\right)(s,\widetilde{X}^n_s) + f(\widetilde{X}^n_s,\mu^n_s)ds\Big]$$

$$\le \mathbb{E}\Big[g(\widetilde{X}^n_T,\mu^n_T) + \int_{t_0}^T \big(-H_\uparrow(x,\Delta^n_+ u^n) - L(x,\widetilde{\alpha}^n_+) - H_\downarrow(x,\Delta^n_- u^n) - L(x,\widetilde{\alpha}^n_-) - \widetilde{\alpha}^n_+\Delta^n_+ u^n - \widetilde{\alpha}^n_-\Delta^n_- u^n$$
$$+ L(x,\widetilde{\alpha}^n_+) + L(x,\widetilde{\alpha}^n_-)\big)(s,\widetilde{X}^n_s) + f(\widetilde{X}^n_s,\mu^n_s)ds\Big],$$

which gives using the strong convexity of L

$$\mathbb{E}\int_{t_0}^T \frac{1}{C}|\alpha^n - \widetilde{\alpha}^n|^2(s,\widetilde{X}_s)ds \le J^n(\widetilde{\alpha}^n,\mu^n) - \mathbb{E}[u^n(t_0,\xi^n)]. \tag{59}$$

We recall that

$$\mathbb{E}[u^n(t_0,\xi^n)] = J^n(\alpha^n,\mu^n) = \mathbb{E}\Big[g(X^n_T,\mu^n_T) + \int_{t_0}^T \left(L(x,\alpha^n_+) + L(x,\alpha^n_-)\right)(s,X^n_s) + f(X^n_s,\mu^n_s)ds\Big].$$

On the other hand, evaluating u on X^n, a similar argument yields

$$\mathbb{E}\int_{t}^T \frac{1}{C}|\alpha^n - \widetilde{\alpha}^n|^2(s,X_s)ds \le \widetilde{J}^n(\alpha^n,\widetilde{\mu}^n) - \mathbb{E}[u(t_0,\xi^n)]. \tag{60}$$

Summing (59) and (60), and then using the monotonicity assumption, we obtain (recalling that $\mathrm{Law}(X^n_s) = \mu_s$ and $\mathrm{Law}(\widetilde{X}^n_s) = \widetilde{\mu}^n_s$)

$$\begin{aligned}
&\mathbb{E}\int_{t_0}^T \frac{1}{C}|\alpha^n - \widetilde{\alpha}^n|^2(s,X^n_s) + \frac{1}{C}|\alpha^n - \widetilde{\alpha}^n|^2(s,\widetilde{X}^n_s)ds\\
&\quad\le J^n(\widetilde{\alpha}^n,\mu^n) - J^n(\alpha^n,\mu^n) + \widetilde{J}^n(\alpha^n,\widetilde{\mu}^n) - \widetilde{J}^n(\widetilde{\alpha}^n,\widetilde{\mu}^n)\\
&\quad= \mathbb{E}\Big[\int_{t_0}^T \Big(f(\widetilde{X}^n_s,\mu^n_s) - f(X^n_s,\mu^n_s) + f(X^n_s,\widetilde{\mu}^n_s) - f(\widetilde{X}^n_s,\widetilde{\mu}^n_s) + r^n(s,X^n_s) - r^n(s,\widetilde{X}^n_s)\Big)ds\\
&\qquad\quad + g(\widetilde{X}^n_T,\mu^n_T) - g(X^n_T,\mu^n_T) + g(X_T,\widetilde{\mu}^n_T) - g(\widetilde{X}^n_T,\widetilde{\mu}^n_T) + r^n(T,X^n_T) - r^n(T,\widetilde{X}^T_s)\Big]\\
&\quad\le \int_{t_0}^T ds\int_{\mathbb{T}}(f(x,\mu^n_s) - f(x,\widetilde{\mu}^n_s)(\widetilde{\mu}^n_s - \mu^n_s)(dx) + \int_{\mathbb{T}}(g(x,\mu^n_T) - g(x,\widetilde{\mu}^n_T)(\widetilde{\mu}^n_T - \mu^n_T)(dx) + \frac{C}{n^\lambda}\\
&\quad\le \frac{C}{n^\lambda},
\end{aligned}$$

which provides

$$\mathbb{E}\int_{t_0}^T |\alpha^n - \widetilde{\alpha}^n|^2(s,X^n_s) + |\alpha^n - \widetilde{\alpha}^n|^2(s,\widetilde{X}^n_s)ds \le \frac{C}{n^\lambda} \tag{61}$$

Step 3. We can now estimate the distance between X^n and $\widetilde{X}^n$. Since $\partial_x u$ is Lipschitz continuous in x, so are $\Delta^n u$ and $\widetilde{\alpha}^n_\pm$, and thus, applying (61) and Jensen's inequality, we obtain

$$\begin{aligned}
\mathbb{E}\Big[\sup_{t_0\le s\le t}|X^n_t - \widetilde{X}^n_t|\Big] &\le \mathbb{E}\int_{t_0}^t |\alpha^n_+(s,X^n_s) - \widetilde{\alpha}^n_+(s,\widetilde{X}^n_s)| + |\alpha^n_-(s,X^n_s) - \widetilde{\alpha}^n_-(s,\widetilde{X}^n_s)|ds\\
&\le +C\mathbb{E}\int_{t_0}^t |X^n_s - \widetilde{X}^n_s|ds + C\sqrt{\mathbb{E}\int_{t_0}^T |\alpha^n - \widetilde{\alpha}^n|^2(s,X^n_s)ds}\\
&\le \frac{C}{n^{\lambda/2}} + C\int_{t_0}^t \mathbb{E}\Big[\sup_{t_0\le r\le s}|X^n_s - \widetilde{X}^n_s|\Big]ds
\end{aligned}$$

and hence Gronwall's lemma yields

$$\mathbb{E}\Big[\sup_{t_0\le s\le T}|X^n_s - \widetilde{X}^n_s|\Big] \le \frac{C}{n^{\lambda/2}}. \tag{62}$$

This estimate (if the processes starts at 0), together with (56), provides (52), while with (55) it proves (53).

Step 4. Finally, in order to estimate $|u^n(t,x) - u(t,x)|$, let $J^n(t_0,x,\beta,\mu^n)$ be the cost (8) where μ^n is fixed and the dynamics starts at (t_0,x) with control β, and similarly $\widetilde{J}^n(t_0,x,\beta,\widetilde{\mu}^n)$

for the cost in (58). Clearly $u^n(t_0,x) = \inf_\beta J^n(t_0,x,\beta,\mu^n)$ and $u(t_0,x) = \inf_\beta \widetilde{J}^n(t_0,x,\beta,\widetilde{\mu}^n)$, the infimum being over open-loop controls. Let $\widehat{\beta}$ be an optimal control for $J^n(t_0,x,\beta,\mu^n)$ with corresponding process $X^{n,x}$, with $X^{n,x}_{t_0} = x$. We get, applying the W_1-Lipschitz-continuity of f and g and (62),

$$
\begin{aligned}
u(t_0,x) - u^n(t_0,x) &\le \widetilde{J}^n(t_0,x,\widehat{\beta},\widetilde{\mu}^n) - J^n(t_0,x,\widehat{\beta},\mu^n)\\
&\le C \sup_{t_0\le s\le T} W_1(\mu^n_s,\widetilde{\mu}^n_s) + \frac{C}{n^\lambda}\\
&\le \frac{C}{n^{\lambda/2}},
\end{aligned}
$$

where we have also used the uniform bound on r^n which appears in $\widetilde{J}$. By changing the roles of μ^n and $\widetilde{\mu}^n$ we derive also the opposite inequality, thus

$$
|u^n(t_0,x) - u(t_0,x)| \le \frac{C}{n^{\lambda/2}}, \tag{63}
$$

where C is independent of t_0, x, m_0.

Finally, to obtain (51), recall that $u^n(t_0,x) = U^n(t_0,x,m^n_0)$ and $u(t_0,x) = U(t_0,x,m_0)$. Therefore (63) and the Lipschitz continuity of U in m provide (51). ☐

Lemma 6. *Let $\alpha \in \mathcal{C}^{\gamma/2,\gamma}([0,T]\times\mathbb{T})$ and let Y satisfy (3) with α and $Y_0 \sim m_0$, Y_n satisfy (7) with $\alpha_\pm$ the positive and negative part of α and $Y^n_0 \sim m^n_0$, with $W_1(m^n_0,m_0) \le \frac{1}{n}$. Then there exists λ such that*

$$
\sup_{0\le t\le T} W_1(\mathrm{Law}(Y^n_t),\mathrm{Law}(Y_t)) \le \frac{C}{n^\lambda} \tag{64}
$$

and

$$
\lim_n Y^n = Y \qquad \text{in law in } \mathcal{D}([0,T],\mathbb{T}). \tag{65}
$$

Proof. We employ the convergence of the generators: denote by $\mathcal{L}^n$ and $\mathcal{L}$ the generators of Y_n and Y, respectively. For a function $\varphi \in \mathcal{C}^{2+\gamma}(\mathbb{T})$, we have

$$
\begin{aligned}
&|\mathcal{L}^n_t\varphi(x) - \mathcal{L}\varphi(x)|\\
&\quad= \bigg| \left(\frac{\alpha_+(t,x)}{\Delta x_n} + \frac{\sigma}{\Delta x_n^2}\right)[\varphi(x+\Delta x_n) - \varphi(x)] + \left(\frac{\alpha_-(t,x)}{\Delta x_n} + \frac{\sigma}{\Delta x_n^2}\right)[\varphi(x-\Delta x_n) - \varphi(x)]\\
&\qquad - \alpha(t,x)\partial_x\varphi(x) - \sigma\partial_{xx}\varphi(x)\bigg|\\
&\quad= \Big|\alpha_+(t,x)\Delta^n_+\varphi(x) + \alpha_-(t,x)\Delta^n_-\varphi(x) + \sigma\Delta^n_2\varphi(x) - \alpha(t,x)\partial_x\varphi(x) - \sigma\partial_{xx}\varphi(x)\Big|\\
&\quad\le ||\alpha(t,\cdot)||_\infty \frac{||\partial_{xx}\varphi||_\infty}{n} + \frac{||\partial_{xx}\varphi||_\gamma}{n^\gamma},
\end{aligned}
$$

that is

$$
\sup_{0\le t\le T} ||\mathcal{L}^n_t\varphi - \mathcal{L}\varphi||_\infty \le \frac{1}{n^\gamma}(1 + ||\alpha(\cdot,\cdot)||_\infty)||\partial_{xx}\varphi||_\gamma \tag{66}
$$

Convergence of the generators provides (65). To prove (64), let $\mathcal{S}^n$ and $\mathcal{S}$ be the semigroups corresponding to $\mathcal{L}^n$ and $\mathcal{L}$:

$$
\mathcal{S}^n_{t,s}\varphi(x) = \mathbb{E}[\varphi(X^n_s)|X^n_t = x], \qquad \mathcal{S}_{t,s}\varphi(x) = \mathbb{E}[\varphi(X_s)|X_t = x]
$$

and we have

$$
\begin{aligned}
&\mathcal{S}^n_{t,s}\mathcal{S}^n_{s,r} = \mathcal{S}^n_{t,r}\\
&\frac{d}{dt}\mathcal{S}^n_{t,s} = \mathcal{S}^n_{t,s}\mathcal{L}^n_s, \qquad \frac{d}{ds}\mathcal{S}^n_{t,s} = -\mathcal{L}^n_t\mathcal{S}^n_{t,s}
\end{aligned}
$$

and similarly for $\mathcal{S}$ and $\mathcal{L}$. Thus we can write

$$
\mathcal{S}^n_{t,s}\varphi - \mathcal{S}_{t,s}\varphi = \mathcal{S}^n_{t,r}\mathcal{S}_{r,s}|^{r=s}_{r=t}\varphi = \int_t^s \frac{d}{dr}\mathcal{S}^n_{t,r}\mathcal{S}_{r,s}\varphi dr = \int_t^s \mathcal{S}^n_{t,r}(\mathcal{L}^n_r - \mathcal{L}_r)\mathcal{S}_{r,s}\varphi dr. \tag{67}
$$

Thanks to Feynman-Kac formula, the function $u(t,x) = \mathcal{S}_{t,s}\varphi(x)$ solves the parabolic PDE

$$\begin{cases} \partial_t u + \sigma \partial_{xx} u + \alpha(t,x)\partial_x u = 0, \\ u(s,x) = \varphi(x). \end{cases}$$

Hence Schauder's estimates give

$$\sup_{0\le t\le s\le T} ||\mathcal{S}_{t,s}\varphi||_{2+\gamma} \le C||\varphi||_{2+\gamma}$$

for a constant C depending on $||\alpha||_{\gamma/2,\gamma}$ and T. From this and (66), and using the fact that the transition operator $\mathcal{S}^n$ is a contraction, we obtain

$$||\mathcal{S}^n_{t,s}\varphi - \mathcal{S}_{t,s}\varphi||_\infty \le \frac{C}{n^\gamma}||\varphi||_{2+\gamma}. \tag{68}$$

We now show that the above estimate, together with the estimate on the initial condition, imply

$$|\mathbb{E}[\varphi(Y^n_t)] - \mathbb{E}[\varphi(Y_t)]| \le \frac{C}{n^\gamma}||\varphi||_{2+\gamma}, \tag{69}$$

uniformly in t, that is

$$\left| \int_{\mathbb{T}} \varphi d(\mathrm{Law}(Y^n_t) - \mathrm{Law}(Y_t)) \right| \le \frac{C}{n^\gamma}||\varphi||_{2+\gamma}. \tag{70}$$

Indeed,

$$\begin{aligned} \mathbb{E}[\varphi(Y^n_t)] - \mathbb{E}[\varphi(Y_t)] &= \int_{\mathbb{T}} S^n_{0,t}\varphi(x)m^n_0(dx) - \int_{\mathbb{T}} S_{0,t}\varphi(x)m_0(dx) \\ &= \int_{\mathbb{T}} (S^n_{0,t} - S_{0,t})\varphi(x)m^n_0(dx) + \int_{\mathbb{T}} S_{0,t}\varphi(x)(m^n_0 - m_0)(dx) \end{aligned}$$

and, estimating the first term by (68) and the second term by $W_1(m^n_0, m_0) \le \frac{1}{N}$ and again by parabolic estimates, we get

$$\begin{aligned} |\mathbb{E}[\varphi(Y^n_t)] - \mathbb{E}[\varphi(Y_t)]| &\le \frac{C}{n^\gamma}||\varphi||_{2+\gamma} + ||\partial_x(S_{0,t}\varphi(x))||_\infty W_1(m^n_0, m_0) \\ &\le \frac{C}{n^\gamma}||\varphi||_{2+\gamma} + \frac{C}{n}||\varphi||_{2+\gamma}. \end{aligned}$$

Finally, the estimate (64) is provided by an estimate of the distance in (70) (for functions in $\mathcal{C}^{2+\gamma}$) in terms of powers of the Wasserstein distance. Denote by ζ_k the Zolotarev metric of order k, defined by

$$\zeta_k(\mu,\nu) = \sup\left\{ \int_{\mathbb{T}} \varphi d(\mu-\nu) : \varphi \in \mathcal{F}_k \right\},$$

where $\mathcal{F}_k$ is the set of $\varphi \in \mathcal{C}^{k-1}(\mathbb{T})$ with $\varphi(0) = \varphi'(0) = \cdots = \varphi^{k-1}(0) = 0$ and $\varphi^{(k-1)}$ is 1-Lipschitz. We have $\zeta_1 = W_1$ and, by [12],

$$\zeta_1 \le c_k \zeta_k^{\frac{1}{k}},$$

where c_k is a constant depending only on k. Since we are on the torus, $\mathcal{F}_3$ coincides with the set of $\varphi \in \mathcal{C}^{2+1}$ with φ, φ' and φ'' bounded by 1 and 1-Lipschitz. By (70) we have, for any $\varphi \in \mathcal{F}_3$,

$$\int_{\mathbb{T}} \varphi d(\mathrm{Law}(Y^n_t) - \mathrm{Law}(Y_t)) \le \frac{C}{n^\gamma}||\varphi||_{2+1} \le \frac{C}{n^\gamma}$$

and thus

$$\zeta_3(\mathrm{Law}(Y^n_t), \mathrm{Law}(Y_t)) \le \frac{C}{n^\gamma},$$

which yields

$$W_1(\mathrm{Law}(Y^n_t), \mathrm{Law}(Y_t)) \le \frac{C}{n^{\gamma/3}}.$$

□

5. The case of common noise

We now turn to a case of MFG involving a common noise. As already mentioned the approach here is quite different. Namely we make an extensive use the stability of monotone solutions introduced in [7, 8]. We postpone recalling the definitions of monotone solutions and we now present the master equations in which we are interested.

The master equation in the continuous state space is

$$\begin{aligned}
&-\partial_t U - \sigma\partial_{xx}U + H(x,\partial_x U) + \int_{\mathbb{T}} \partial_p H(x,\partial_x U(t,y,m))D^m U(t,x,m;y)m(dy)\\
&\qquad -\sigma\int_{\mathbb{T}} \partial_y D^m U(t,x,m;y)m(dy) + \lambda(U - \mathcal{A}^* U(t,x,\mathcal{A}m)) = f(x,m)\\
&U(T,x,m) = g(x,m).
\end{aligned}\tag{71}$$

It is very similar to the one at interest in the previous section, expect for the presence of terms modeling common noise or common shocks. The common noise is here similar to the one introduced in [10, 8]. At random times, of intensity $\lambda > 0$, all the players are affected by the map $\mathcal{A} : \mathcal{P}(\mathbb{T}) \to \mathcal{P}(\mathbb{T})$, and $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$. To fix ideas, mainly for the discretization, we take $\mathcal{A}$ of the form

$$\mathcal{A}m := \int_{\mathbb{T}} K(\cdot,y)m(dy), \tag{72}$$

where $K : \mathbb{T}^2 \to \mathbb{R}$ is a smooth function which satisfies $K \geq 0, \int K(x,y)dx = 1$.

In this section we shall only be concerned with the first discretization introduced earlier. That is, at the discrete state level, we are interested in the following master equation

$$\begin{aligned}
&-\partial_t U^n(x,m) + H_\uparrow(x,\Delta_+^n U^n(x,m)) + H_\downarrow(\Delta_-^n U^n(x,m)) - \sigma\Delta_2^n U^n(x,m) - f(m)(x)\\
&\quad + \sum_{y\in S^n} m_y\left(\frac{\partial_p H_\uparrow(y,(\Delta_+^n U^n(y,m))}{\Delta x_n} + \frac{\sigma}{\Delta x_n^2}\right)\Big(\partial_{m_{y+\Delta x_n}} U^n(x,m) - \partial_{m_y} U^n(x,m)\Big)\\
&\quad - \sum_{y\in S^n} m_y\left(\frac{\partial_p H_\downarrow(y,\Delta_-^n U^n(y,m))}{\Delta x_n} + \frac{\sigma}{\Delta x_n^2}\right)\Big(\partial_{m_{y-\Delta x_n}} U^n(x,m) - \partial_{m_y} U^n(x,m)\Big)\\
&\quad + \lambda(U - A_n^* U(t,A_n m)) = 0\\
&U(T,x,m) = g(m)(x)
\end{aligned}\tag{73}$$

where A_n is the discretization of $\mathcal{A}$ given by

$$(A_n m)_i = \frac{1}{n}\sum_j K_{i,j} m_j, \tag{74}$$

where $(K_{i,j})_{1\leq i,j\leq n}$ is a suitable discretization of K such that for all $1 \leq j \leq n$, $\sum_i K_{i,j} = n$. To lighten the notation of the following, we also introduce the operators $F^n, G^n : \mathcal{P}(S^n) \times \mathbb{R}^n$ defined by

$$\begin{aligned}
(G^n(m,p))_i &= f(x_i,m) - H_\uparrow(x,\Delta_+^n p(x_i)) - H_\downarrow(x,\Delta_-^n p(x_i)) + \sigma\Delta_2^n p(x_i)\\
(F^n(m,p))_i &= -\sum_{j=1}^n D_{p_i}(G^n(m,p))_j m_j,
\end{aligned}\tag{75}$$

where $p \in \mathbb{R}^n$ is interpreted as a real function $S^n \to \mathbb{R}$. Using these notations, (73) can be written

$$-\partial_t U(t,\cdot,m) + (F^n(m,U)\cdot\nabla_m)U + \lambda(U - A_n^* U(t,A_n m)) = G^n(m,U). \tag{76}$$

5.1. Monotone solutions of master equations.

The following is a brief reminder on monotone solutions. We refer to [7, 8] for details on this concept. We start with the definitions.

Definition 7. *A continuous function U, C^2 in space, is a monotone solution of* (71) *if for any measure $\nu \in \mathcal{M}(\mathbb{T})$ C^2 function φ of the space variable and C^1 function ψ of the time variable,* *for any* $(t_*, m_*) \in [0,T) \times \mathcal{P}(\mathbb{T})$ *point of strict minimum of* $(t,m) \to \langle U(t,m) - \varphi, m - \nu\rangle - \psi(t)$, *the following holds*

$$
\begin{aligned}
&-\frac{d\psi(t_*)}{dt} + \langle -\sigma\partial_{xx}U + H(x,\partial_x U) + \lambda(U - \mathcal{A}^*U(\mathcal{A}m_*)), m_* - \nu\rangle \geq \langle f(m_*), m_* - \nu\rangle \\
&- \langle U - \varphi, \partial_x(\partial_p H(\cdot,\partial_x U)m_*)\rangle - \sigma\langle \partial_{xx}(U-\varphi), m_*\rangle.
\end{aligned} \tag{77}
$$

The same type of definition holds at the discrete level and it is given in this situation by

Definition 8. *For* $n > 0$*, a continuous function* U *is a monotone solution of* (73) *if for any* $\nu \in \mathcal{M}(S^n)$*,* $a \in \mathbb{R}^n$*,* C^1 *function of the time* ψ*, and* $(t_*, m_*) \in [0,T) \times \mathcal{P}(S^n)$ *point of strict minimum of* $(t,m) \to \langle U(t,m) - \varphi, m - \nu\rangle - \psi(t)$*, the following holds*

$$
\begin{aligned}
-\frac{d\psi(t_*)}{dt} + \lambda\langle U - A_n^*U(A_n m_*), m_* - \nu\rangle &\geq \langle G^n(m_*, U(t_*, m_*)), m_* - \nu\rangle \\
&+ \langle F^n(m_*, U(t_*, m_*)), U(t_*, m_*) - a\rangle.
\end{aligned} \tag{78}
$$

The previous concept of solution possesses several properties. First we can mention that under the standing assumptions on the monotonicity of f and g, there is at most one monotone solution of either (71) or (73). Those solutions also enjoy several stability properties, in some sense, this part is an illustration of this fact. We continue this section with results of existence of such solutions.

Proposition 9. *Assume that* g *is monotone,* f *is strictly monotone and that they satisfy for* $\alpha, \beta \in (0,1]$*,*

$$
\begin{aligned}
\sup_m \|f(\cdot,m)\|_{1+\alpha} + \sup_{m,\nu} \frac{\|f(\cdot,m) - f(\cdot,\nu)\|_{1+\alpha}}{W_1(m,\nu)^\beta} < \infty, \\
\sup_m \|g(\cdot,m)\|_{2+\alpha} + \sup_{m,\nu} \frac{\|g(\cdot,m) - g(\cdot,\nu)\|_{2+\alpha}}{W_1(m,\nu)^\beta} < \infty.
\end{aligned} \tag{79}
$$

Assume that K *is a smooth function. Then there exists a unique monotone solution to both* (71) *and* (73)*.*

Proof. For the continuous state space, a similar result can be found in [8] and for the discrete case, a similar result is in [7]. In both cases, the only difference lies in the fact that the Hamiltonian can have a quadratic growth. We leave to the reader these immediate generalizations. □

Remark that the definition of monotone solution requires some regularity with respect to the space variable in the continuous case whereas, obviously, no such assumption is needed in the finite state case. An important consequence of this fact is that, stated as it is, some uniform continuity of the spatial derivatives with respect to the measure variable are needed. If such results on the first order derivatives of the solutions are fairly easy to obtain, they require slight additional assumptions for second order derivatives. Even if the setting of Proposition 9 is sufficient to obtain such result, we mention here this difficulty for two reasons. The first one is to explain why this questions shall pop out in the study of the convergence of the master equations, namely because we are going to use this property for the limit equation. Secondly because we believe that this point is of some importance and we shall explain how it can be dealt with in another manner later on in this part. We end this section with the following.

Proposition 10. *Under the assumption of Proposition 9, the unique monotone solution* U *of* (71) *is such that* $\nabla_x U$ *and* $\Delta_x U$ *are continuous on* $[0,T] \times \mathcal{P}(\mathbb{T})$*.*

The proof of this statement is in [8].

5.2. **A discrete parabolic estimate.** In this section we present estimates on the semi discrete heat equation, that is discretized in space but not in time. These estimates, in the flavor of parabolic regularity, is at the same time, fundamental to obtain compactness on the sequence $(U_n)_{n\geq 0}$ of solutions of (73), quite technical to establish, and not particularly interesting in itself since much more involved results are already well-known in the continuous setting. However, because we could not find sufficiently similar results in the literature, we take some time to explain the proof of such a result.

Our aim is to establish regularity results on the ODE

$$\begin{aligned} \dot{u}(t) &= \Lambda u + f(t), \\ u(0) &= g, \end{aligned} \tag{80}$$

where Λ is defined by

$$\Lambda = n^2 \begin{pmatrix} -2 & 1 & \dots & \dots & 1 \\ 1 & -2 & 1 & \dots & \dots \\ 0 & 1 & -2 & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \dots & \dots & 1 & -2 \end{pmatrix}$$

Clearly Λ is a space discretization of the Laplacian operator. We prove the following.

Theorem 11. *Assume that g is the evaluation of a smooth function on S^n. If f is uniformly bounded by a constant C, then, the solution u of* (80) *satisfies*

$$|n^{\alpha-1}\Lambda u_i(t)| \leq C, \tag{81}$$

for a constant C independent of n and any $\alpha \in (0, \frac{1}{2})$.
If f satisfies for constants $C \geq 0$ and $\alpha \in (\frac{1}{2}, 1]$,

$$n^{\alpha}\left|f_i(t) - f_{i+1}(t)\right| \leq C, \tag{82}$$

then, the solution u of (80) *satisfies that Λu is bounded by a constant independent of n.*

Remark 12. *The inequality* (81) *is a sort of α-Hölder estimate on the discrete spatial gradient of u and the inequality* (82) *is a sort of α-Hölder estimate on f.*

Proof. Let $\widetilde{f} : \mathbb{R} \to \mathbb{R}$ be the function which is 1-periodic, and the linear interpolation of f on $[0, 1]$ with $\widetilde{f}(\frac{i}{n}) = f_i$. Let us note $(c_k)_{k\in\mathbb{Z}}$ the Fourier exponents of $\widetilde{f}$. Because $\widetilde{f}$ is continuous, we deduce that it is the sum of its Fourier series, hence

$$f_j = \sum_{k\in\mathbb{Z}} c_k e^{\frac{2ik\pi j}{n}}, \tag{83}$$

Let us define $Q_{kj} = n^{-\frac{1}{2}} e^{\frac{2i\pi kj}{n}}$ and $\lambda_k = 2(1 - \cos(\frac{2k\pi}{n}))$. The vector Q_k is an eigenvector of Λ associated to the eigenvalue λ_k. Then, u is given by

$$u(t) = e^{t\Lambda} g + \int_0^t e^{(t-s)\Lambda} f(s) ds. \tag{84}$$

From which we deduce that

$$\begin{aligned} (\Lambda u(t))_l &= (\Lambda e^{t\Lambda} g)_l - \int_0^t \sum_{k,j} Q_{kl} Q_{kj} \lambda_k e^{-\lambda_k (t-s)} f_j(s) ds \\ &= (\Lambda e^{t\Lambda} g)_l - \int_0^t \sum_{k,j=1}^n Q_{kl} Q_{kj} \lambda_k e^{-\lambda_k (t-s)} \sum_{p\in\mathbb{Z}} c_p(s) e^{\frac{2i\pi jp}{n}} ds \\ &= (\Lambda e^{t\Lambda} g)_l - \int_0^t \sum_{k=1}^n e^{\frac{2ikl\pi}{n}} \lambda_k e^{-\lambda_k (t-s)} \sum_{j\in\mathbb{Z}} c_{nj-k}(s) ds \end{aligned} \tag{85}$$

We can now observe that estimates on Λu can be obtained by using decay assumptions on the Fourier coefficients of $\widetilde{f}$. Unlike for classical regularity (C^k for $k \in \mathbb{N}$), Hölder regularity does not immediately translates into a decay of the Fourier coefficients but rather on their cumulative sum. Namely, if $\widetilde{f} \in C^\alpha$ for $\alpha \in (0, 1)$, the following holds

$$\sum_{|k|\in[2^m, 2^{m+1}]} |c_k(t)| \leq C 2^{m(\frac{1}{2}-\alpha)}, \tag{86}$$

see for instance the first chapter of [23]. In particular, the series is summable if $\alpha > \frac{1}{2}$, which gives the second part of the result. The first part is obtained by remarking that since $\widetilde{f}$ is, uniformly in n, in L^∞, the sequence $(c_k(t))_{k\in\mathbb{Z}}$ is, uniformly in t and n, bounded in ℓ^2. Thus multiplying both sides of the last equation by $n^{\alpha-1}$, we deduce the required result. $\square$

5.3. **An estimate on the solution of the master equation without common noise.** We use the previous estimate to derive an estimate on the solution of (17), that is (73) in the case $\mathcal{A} = 0$. We use the 1-Wasserstein distance W_1 on $\mathcal{P}(\mathbb{T})$ and recall that U^n is evaluated on measures of the form $m = \sum_{j=1}^n m_j \delta_{\frac{j}{n}}$.

Theorem 13. *If f and g satisfy the assumption of Proposition 9 with $\beta = 1$, then there exists a constant C independent of n such that*

$$|U^n(t,x,m) - U^n(\widetilde{t},\widetilde{x},\widetilde{m})| \le C\left(\sqrt[4]{|t-\widetilde{t}|} + |x-\widetilde{x}| + \sqrt{W_1(m,\widetilde{m})}\right) \tag{87}$$

for any $t,\widetilde{t} \in [0,T]$, $x,\widetilde{x} \in S^n$, $m = \sum_{j=1}^n m_j\delta_{x_j}, \widetilde{m} = \sum_{j=1}^n \widetilde{m}_j \delta_{x_j} \in \mathcal{P}(S^n)$. Moreover, the discrete gradient of Δ_+^n also satisfies the same estimate.

Proof. **Step 1.** The uniform Lipschitz continuity in space x can be proven exactly as in Lemma 1.

Step 2. To prove the estimate in m, fix the initial time t and consider the two solutions of the associated MFG system (35) (u,μ) and $(\widetilde{u},\widetilde{\mu})$ with $\mu_t = m$, $\widetilde{\mu}_t = \widetilde{m}$. (Let us omit n in the notation.) Recall that $U^n(t,x,m) = u(t,x)$ and $U^n(t,x,\widetilde{m}) = \widetilde{u}(t,x)$. Let ξ and $\widetilde{\xi}$ be two random variables (the initial conditions) which attain the minimum in the 1-Wasserstein distance, i. e. $\mathrm{Law}(\xi) = m$, $\mathrm{Law}(\widetilde{\xi}) = \widetilde{m}$ and

$$\mathbb{E}|\xi - \widetilde{\xi}| = W_1(m,\widetilde{m}). \tag{88}$$

Consider the optimal feedback control for (u,m): $\alpha(s,x) = (\alpha^+,\alpha^-) = \big(-\partial_p H_\uparrow(x,\Delta_+^n u^n(s,x))$ $,\partial_p H_\downarrow(x,(\Delta_-^n u^n(s,x))\big)$, and similarly let $\widetilde{\alpha}$ be the optimal feedback for $\widetilde{u},\widetilde{m}$. Let X^ξ be the state process driven by the control α, with $X_t^\xi = \xi$, and $\widetilde{X}^{\widetilde{\xi}}$ be the process driven by the control $\widetilde{\alpha}$ with $\widetilde{X}_t^{\widetilde{\xi}} = \widetilde{\xi}$. For μ fixed and a control β (open-loop or feedback), denote by $J(t,\xi,\beta,\mu)$ the cost in (8) starting at t, ξ, and similarly $J(t,\widetilde{\xi},\beta,\widetilde{\mu})$.

We compute u on $\widetilde{X}$: we have

$$\begin{aligned}\mathbb{E}[u(t,\widetilde{\xi})] &= \mathbb{E}[u(T,\widetilde{X}_T)] + \mathbb{E}\int_t^T \left(-\partial_t u - \sigma\Delta_2^n u - \widetilde{\alpha}_+\Delta_+^n u - \widetilde{\alpha}_-\Delta_-^n u\right)(s,\widetilde{X}_s)ds \\ &= \mathbb{E}\Big[g(\widetilde{X}_T,\mu_T) + \int_t^T \left(-H_\uparrow(x,\Delta_+^n u) - H_\downarrow(x,\Delta_-^n u) - \widetilde{\alpha}_+\Delta_+^n u - \widetilde{\alpha}_-\Delta_-^n u\right)(s,\widetilde{X}_s) + f(\widetilde{X}_s,\mu_s)ds.\end{aligned}$$

A similar computation as in the proof of Theorem 5 yields

$$\mathbb{E}\int_t^T \frac{1}{C}|\alpha - \widetilde{\alpha}|^2(s,\widetilde{X}_s)ds \le J(t,\widetilde{\xi},\widetilde{\alpha},\mu) - \mathbb{E}[u(t,\widetilde{\xi})]. \tag{89}$$

Similarly, we get

$$\mathbb{E}\int_t^T \frac{1}{C}|\alpha - \widetilde{\alpha}|^2(s,X_s)ds \le J(t,\xi,\alpha,\widetilde{\mu}) - \mathbb{E}[\widetilde{u}(t,\xi)], \tag{90}$$

and we have

$$\begin{aligned}\mathbb{E}[u(t,\xi)] &= J(t,\xi,\alpha,\mu)\\ \mathbb{E}[\widetilde{u}(t,\widetilde{\xi})] &= J(t,\widetilde{\xi},\widetilde{\alpha},\widetilde{\mu}).\end{aligned}$$

Summing (89) and (90), adding and subtracting $\mathbb{E}[u(t,\xi)]$ and $\mathbb{E}[\widetilde{u}(t,\widetilde{\xi})]$ and then using the monotonicity assumption, we obtain (recalling that $\mathrm{Law}(X_s) = \mu_s$ and $\mathrm{Law}(\widetilde{X}_s) = \widetilde{\mu}_s$)

$$\begin{aligned}&\mathbb{E}\int_t^T \frac{1}{C}|\alpha - \widetilde{\alpha}|^2(s,X_s) + \frac{1}{C}|\alpha - \widetilde{\alpha}|^2(s,\widetilde{X}_s)ds\\ &\qquad \le \mathbb{E}[u(t,\xi) - u(t,\widetilde{\xi}) + \widetilde{u}(t,\widetilde{\xi}) - \widetilde{u}(t,\xi)]\end{aligned}$$

$$
\begin{aligned}
&+ J(t,\widetilde{\xi},\widetilde{\alpha},\mu) - J(t,\xi,\alpha,\mu) + J(t,\widetilde{\xi},\widetilde{\alpha},\mu) + J(t,\xi,\alpha,\widetilde{\mu}) - J(t,\widetilde{\xi},\widetilde{\alpha},\widetilde{\mu})\\
&= \int_{\mathbb{T}} (u-\widetilde{u})(0,x)d(m-\widetilde{m})(x) + \mathbb{E}\Big[\int_t^T \Big(f(\widetilde{X}_s,\mu_s) - f(X_s,\mu_s) + f(X_s,\widetilde{\mu}_s) - f(\widetilde{X}_s,\widetilde{\mu}_s)\Big)ds\\
&\quad + g(\widetilde{X}_T,\mu_T) - g(X_T,\mu_T) + g(X_T,\widetilde{\mu}_T) - g(\widetilde{X}_T,\widetilde{\mu}_T)\Big]\\
&= \int_{\mathbb{T}} (u-\widetilde{u})(0,x)d(m-\widetilde{m})(x)\\
&\quad + \int_t^T ds \int_{\mathbb{T}} (f(x,\mu_s) - f(x,\widetilde{\mu}_s)(\widetilde{\mu}_s-\mu_s)(dx) + \int_{\mathbb{T}} (g(x,\mu_T) - g(x,\widetilde{\mu}_T)(\widetilde{\mu}_T-\mu_T)(dx)\\
&\leq \int_{\mathbb{T}} (u-\widetilde{u})(0,x)d(m-\widetilde{m})(x)
\end{aligned}
$$

We now bound the r.h.s using the Lipschitz continuity of u and $\widetilde{u}$ to obtain :

$$
\mathbb{E}\int_t^T |\alpha-\widetilde{\alpha}|^2(s,X_s) + |\alpha-\widetilde{\alpha}|^2(s,\widetilde{X}_s)ds \leq CW_1(m,\widetilde{m}) \tag{91}
$$

Step 3. We now use a Lipschitz property on the discrete gradient of u, or $\widetilde{u}$ (see Step 6 below):

$$
|\Delta^n_{\pm}u(x) - \Delta^n_{\pm}u(\widetilde{x})| \leq C|x-\widetilde{x}|. \tag{92}
$$

If this is true, then applying (91) and Jensen's inequality, we obtain

$$
\begin{aligned}
\mathbb{E}|X^{\xi}_s - \widetilde{X}^{\widetilde{\xi}}_s| &\leq \mathbb{E}|\xi-\widetilde{\xi}| + \mathbb{E}\int_t^s |\alpha_+(r,X_r) - \widetilde{\alpha}_+(r,\widetilde{X}_r)| + |\alpha_-(r,X_r) - \widetilde{\alpha}_-(r,\widetilde{X}_r)|dr\\
&\leq \mathbb{E}|\xi-\widetilde{\xi}| + C\mathbb{E}\int_t^s |X_r-\widetilde{X}_r|dr + C\sqrt{\mathbb{E}\int_t^T |\alpha-\widetilde{\alpha}|^2(r,X_r)ds}\\
&\leq C(\mathbb{E}|\xi-\widetilde{\xi}| + \sqrt{W_1(m,\widetilde{m})} + C\int_t^s \mathbb{E}|X_r-\widetilde{X}_r|dr
\end{aligned}
$$

and thus Gronwall's lemma yields

$$
\sup_{t\leq s\leq T} \mathbb{E}|\widetilde{X}^{\widetilde{\xi}}_s - X^{\xi}_s| \leq C\sqrt{W_1(m,\widetilde{m})}. \tag{93}
$$

Step 4. We bound the value functions, classicaly using the characteristics, by

$$
\begin{aligned}
|U^n(t,x,m) - U^n(t,x,\widetilde{m})| &= |u(t,x) - \widetilde{u}(t,x)| \leq C\sup_{t\leq s\leq T} W_1(\mu_s,\widetilde{\mu_s})\\
&\leq C\sup_{t\leq s\leq T}\mathbb{E}|\widetilde{X}^{\widetilde{\xi}}_s - X^{\xi}_s| \leq C(\mathbb{E}|\xi-\widetilde{\xi}|)^{\frac{1}{2}} = C\sqrt{W_1(m,\widetilde{m})}.
\end{aligned}
$$

Step 5. To prove the estimate in time, let $\widetilde{t} > t$ and consider the HJB equation (11) starting at t. For μ^n fixed, u^n represents the value function corresponding to the cost (8) and we have

$$
u^n(t,x) = U^n(t,x,m), \qquad u^n(\widetilde{t},x) = U^n(\widetilde{t},x,\mu^n_{\widetilde{t}}),
$$

where μ^n_s is the Law of the process starting at (t,m) with control given by $\Delta^n u$. Hence the dynamic programmin principle gives

$$
u^n(t,x) = \mathbb{E}\left[\int_t^{\widetilde{t}} L(X^n_s,\Delta^n_+u^n(s,X^n_s)) + L(X^n_s,\Delta^n_-u^n(s,X^n_s)) + f(X^n_s,\mu^n_s)ds + u^n(\widetilde{t},X^n_{\widetilde{t}})\right],
$$

where X is now the same process as before but conditioned with $X_t = x$. Since u^n is uniformly Lipschitz in space, $\Delta^n_+u^n$ and $\Delta^n_-u^n$ are uniformly bounded and we have

$$
\begin{aligned}
|u^n(t,x) - u^n(\widetilde{t},x)| &\leq |u^n(t,x) - \mathbb{E}u^n(\widetilde{t},X^n_{\widetilde{t}})| + |\mathbb{E}u^n(\widetilde{t},X^n_{\widetilde{t}}) - u^n(\widetilde{t},x)|\\
&\leq C(\widetilde{t}-t) + C\mathbb{E}|X^n_{\widetilde{t}} - x|.
\end{aligned}
$$

We bound the latter term by using the SDE representation (38):

$$
\begin{aligned}
\mathbb{E}|X^n_{\tilde{t}} - x|^2 &\leq C\mathbb{E}\left|\int_t^{\tilde{t}} \lambda(\Delta^n_\pm u^n(s, X^n_s), \theta)\nu(d\theta)ds\right|^2 + C\mathbb{E}\left|\int_t^{\tilde{t}} \lambda(\Delta^n_\pm u^n(s, X^n_s), \theta)(\mathcal{N}(d\theta, ds - \nu(d\theta)ds)\right|^2 \\
&\leq C\mathbb{E}\left|\int_t^{\tilde{t}} \Delta x_n \frac{\partial_p H_\uparrow(X^n_s, \Delta^n_+ u^n(s, X^n_s)) + \partial_p H_\downarrow(X^n_s, \Delta^n_- u^n(s, X^n_s))}{\Delta x_n} ds\right|^2 \\
&\quad + C\mathbb{E}\int_t^{\tilde{t}} \left|\lambda(\Delta^n_\pm u^n(s, X^n_s), \theta)\right|^2 \nu(d\theta)ds \\
&\leq C(\tilde{t} - t)^2 + C\mathbb{E}\int_t^{\tilde{t}} \Delta x_n^2 \left(\frac{1}{\Delta x_n^2} + (\Delta^n_+ u^n(s, X^n_s))_- + \frac{1}{\Delta x_n^2} + (\Delta^n_- u^n(s, X^n_s))_-\right) ds \\
&\leq C(\tilde{t} - t)^2 + C(\tilde{t} - t)
\end{aligned}
$$

and therefore $\mathbb{E}|X^n_{\tilde{t}} - x| \leq C\sqrt{\tilde{t} - t}$, which yields

$$|u^n(t, x) - u^n(\tilde{t}, x)| \leq C\sqrt{\tilde{t} - t}.$$

Similarly we get

$$W_1(\mu^n_{\tilde{t}}, m) \leq C\sqrt{\tilde{t} - t}$$

and hence, applying the Hölder continuity in m,

$$
\begin{aligned}
|U^n(\tilde{t}, x, m) - U^n(t, x, m)| &\leq |U^n(\tilde{t}, x, \mu^n_{\tilde{t}}) - U^n(\tilde{t}, x, m) \; + |u^n(\tilde{t}, x) - u^n(t, x)| \\
&\leq C\sqrt{W_1(\mu^n_{\tilde{t}}, m)} + C\sqrt{\tilde{t} - t} \\
&\leq C(\tilde{t} - t)^{\frac{1}{4}}.
\end{aligned}
$$

Step 6. The fact that the discrete gradient satisfies the same estimate simply follows from remarking that the two previous steps can be made for the discrete gradient exactly in the same way. This comes from using the estimate (93) on representation formulae for the discrete gradient of U^n.

Step 7. It remains to prove (92). This estimate follows from successive uses of Theorem 11 on the discrete HJB equation in the characteristics. Indeed, as we already established the uniform Lipschitz estimate in Lemma 1, we deduce a α-Hölder type estimate on the spatial gradient of u, for $\alpha \in (0, \frac{1}{2})$. Using this new information, we use once again this argument to obtain a higher order regularity on u, and then once more to finally obtain the required boundedness of the discrete Laplacian (uniformly in n of course).

□

Remark 14. *The previous result is the only part of this paper in which the dimension* 1 *plays a particular role. Indeed, even if it is extremely likely that the estimate proved in Theorem 11 can be generalized to other dimension, it is not proved here.*

5.4. **Compactness results for master equations with common noise.** In this section, we explain how the previous estimate can be used to gain compactness on the sequence $(U^n)_{n\geq 0}$ of solutions of (73).

Proposition 15. *Under the assumptions of Proposition 9, there exists a continuous function* $V : [0, T] \times \mathbb{T}^d \times \mathcal{P}(\mathbb{T})$ *such that, extracting a subsequence if necessary*

$$\lim_{n\to\infty} \sup\{|U^n(t, x, m) - V(t, x, m)|, (t, x, m) \in [0, T] \times S^n \times \mathcal{P}(S^n)\} = 0. \tag{94}$$

Moreover, V *is uniformly* $C^{2,\alpha}$ *in* x *for some* $\alpha \in (0, 1)$, $\partial_x V$ *is continuous on* $[0, T] \times \mathbb{T}^d \times \mathcal{P}(\mathbb{T})$ *and*

$$\lim_{n\to\infty} \sup\{|\Delta^n_\pm U^n(t, x, m) - \partial_x V(t, x, m)|, (t, x, m) \in [0, T] \times S^n \times \mathcal{P}(S^n)\} = 0. \tag{95}$$

Proof. This statement is purely a compactness one. It relies on proving an a priori estimate on the solutions of (73) and then using a version of Ascoli-Arzela Theorem. Following the technique of the proof of Proposition 1.3 in [8] (that we do not reproduce here for the sake of clarity), we know that there exists $C > 0$, such that for all $n > 0, (t, x, m) \in [0, T] \times S^n \times \mathcal{P}(S^n), \xi \in \mathbb{R}^n$,

$$\xi \cdot D_m U^n(t, x, m) \cdot \xi \leq C\langle \xi, \xi \rangle. \tag{96}$$

This yields a, uniform in n, Lipschitz estimate on $U(t)$ seen as an operator from $\mathcal{P}(S^n)$ to $\mathbb{R}^n$ when $\mathbb{R}^n$ is equipped with the ℓ_2 norm and $\mathcal{P}(S^n)$ is equipped with the distance[1] $\widetilde{d}(m, m') = \sqrt{n^{-1}\sum_i (m_i - m_i')^2}$. From the properties of the operators $\mathcal{A}$ and $(A^n)_{n>0}$, we deduce that $m \to \lambda A_n^* U(t, \cdot, A_n m)$ is uniformly Lipschitz continuous from $\mathcal{P}(S^n)$ to $\mathbb{R}^n$ when $\mathcal{P}(S^n)$ is equipped with the Wasserstein distance and $\mathbb{R}^n$ with the ℓ_∞ norm. Then, passing this term to the right hand side of the equation, we deduce using Theorem 13 that its conclusion is still satisfied here.

The rest of the proof is now classical. Let us define $\overline{U}^n$ by

$$\begin{aligned}
&\forall (t, x, m) \in [0, T] \times \mathbb{T}^d \times \mathcal{P}(\mathbb{T}),\\
&\overline{U}^n(t, x, m) = \inf \left\{ U^n(t, y, \mu) + C|x - y| + C\sqrt{W_1(m, \mu)}, (y, \mu) \in S^n \times \mathcal{P}(S^n) \right\},
\end{aligned} \tag{97}$$

where C is a constant given by the use of Theorem 13. The sequence $(\overline{U}^n)_{n>0}$ satisfies the assumptions of Ascoli-Arzela Theorem which concludes the proof of the first part of the statement. The additional regularity of V is simply obtained by remarking that U^n possesses all this regularity, at the discretized level, uniformly in n and in all the variables. Remark that the uniform $C^{2,\alpha}$ estimate can be proved last, by using a representation through characteristics and Theorem 11. ☐

5.5. **Convergence of the discretized problem.** We now state in which sense any function V given by Proposition 15 is indeed the unique monotone solution of (71).

Let us first remark that in the formulation of monotone solutions of (71), one only needs the Laplacian of V to make sense against the test measure ν. Indeed the term $\langle \Delta U, m_* \rangle$ appears on the two sides of the inequality. Hence, if V is not sufficiently regular in x, one can still test its Laplacian against a measure with a regular density. This remark leads us to the following.

Lemma 16. *Let V be any function given by the Proposition 15. For any measure $\nu \in \mathcal{M}(\mathbb{T}) \cap W^{2,\infty}(\mathbb{T})$, C^2 function φ of the space variable and C^1 function ψ of the time variable, for any $(t_*, m_*) \in [0, T) \times \mathcal{P}(\mathbb{T})$ point of strict minimum of $(t, m) \to \langle V(t, m) - \varphi, m - \nu \rangle - \psi(t)$ the following holds*

$$\begin{aligned}
&-\frac{d\psi(t_*)}{dt} + \langle H(x, \nabla_x V) + \lambda(V - \mathcal{A}^* V(\mathcal{A} m_*)), m_* - \nu \rangle \geq \langle f(m_*), m_* - \nu \rangle\\
&- \langle V - \varphi, \partial_x(\partial_p H(\cdot, \partial_x V) m_*) \rangle + \sigma \langle \partial_{xx} \varphi, m_* \rangle - \sigma \langle \partial_{xx} V(t_*, m_*), \nu \rangle.
\end{aligned} \tag{98}$$

Proof. Consider $t_0, \nu, \varphi, \psi, t_*, m_*$ as in the statement. For any $n > 0$, consider ν^n and φ^n suitable discretizations of ν and φ. Thanks to Stegall's Lemma [29, 28], for any $n > 0$, there exists $\delta_n \in \mathbb{R}, a_n \in \mathbb{R}^n$ as small as we want, such that $(t, m) \to \langle U^n(t, m) - \varphi^n, m - \nu^n \rangle - \psi(t) + \delta^n t + \langle a_n, m \rangle$ has a strict minimum at (t_n, m_n) on $[0, t_0] \times \mathcal{P}(S^n)$. Because U^n is a monotone solution of (73) we obtain that

$$\begin{aligned}
-\frac{d\psi(t_n)}{dt} - \delta_n + \lambda \langle U^n(m_n) - A_n^* U(A_n m_n), m_n - \nu \rangle &\geq \langle G^n(m_n, U(t_n, m_n)), m_n - \nu^n \rangle\\
&\quad + \langle F^n(m_*, U(t_*, m_*)), U(t_*, m_*) - \varphi^n - a_n \rangle.
\end{aligned} \tag{99}$$

Passing to the limit $n \to \infty$ in the previous inequality yields the required result.

☐

On the other hand, defining $\mathcal{P}_M = \{m \in \mathcal{P}(\mathbb{T}), \|m\|_{2,\infty} \leq M\}$, the unique monotone solution U of (71) satisfies

[1]This distance can be interpreted as an $L^2(\mathbb{T})$ distance.

Lemma 17. *Fix $C > 0$. There exists a function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ such that $w(M) \to 0$ when $M \to \infty$ and for any measure $\nu \in \mathcal{M}(\mathbb{T})$, $C^{2,\alpha}$ function φ of the space variable and C^1 function ψ of the time variable, both bounded by C, for any $(t_*, m_*) \in [0,T) \times \mathcal{P}_M$ point of strict minimum of $(t,m) \to \langle U(t,m) - \varphi, m - \nu\rangle - \psi(t)$ on $[0,T) \times \mathcal{P}_M$, the following holds*

$$\begin{aligned} &-\frac{d\psi(t_*)}{dt} + \langle H(x, \partial_x U) + \lambda(U - \mathcal{A}^* U(\mathcal{A}m_*)), m_* - \nu\rangle \geq \langle f(m_*), m_* - \nu\rangle \\ &- \langle U - \varphi, \partial_x(\partial_p H(x, \partial_x U)m_*)\rangle + \sigma\langle \partial_{xx}\varphi, m_*\rangle - \sigma\langle \partial_{xx} U(t_*, m_*), \nu\rangle - \omega(M). \end{aligned} \tag{100}$$

In other words, U is almost a solution of (71) on $\mathcal{P}_M$, uniformly in M.

Proof. Assume that it is not the case. Reasoning by contradiction, there exists $\kappa > 0$ and a sequence $(\varphi_M, \psi_M, t_M, m_M, \nu_M)$ such that $\|\psi_M\|_1 + \|\varphi_M\|_{2+\alpha} \leq 2C$, (t_M, m_M) point of strict minimum of $(t,m) \to \langle U(t,m) - \varphi_M, m - \nu_M\rangle - \psi_M(t)$ on $[0,T) \times \mathcal{P}_M$.

$$\begin{aligned} &-\frac{d\psi_M(t_*)}{dt} + \langle H(x, \partial_x U) + \lambda(U - \mathcal{A}^* U(\mathcal{A}m_M)), m_M - \nu_M\rangle \leq \langle f(m_M), m_M - \nu_M\rangle \\ &- \langle U - \varphi, \partial_x(\partial_p H(x, \partial_x U)m_M)\rangle + \sigma\langle \partial_{xx}\varphi_M, m_M\rangle - \sigma\langle \partial_{xx} U(t_M, m_M), \nu_M\rangle - \kappa. \end{aligned} \tag{101}$$

Extract a subsequence if necessary and consider the limit point $(\varphi_*, \psi_*, \nu_*)$ of the sequence $(\varphi_M, \psi_M, \nu_M)_{M\geq 0}$. Using once again Stegall's Lemma, for any $\varepsilon > 0$, there exists $\delta \in (-\varepsilon, \varepsilon)$, $\widetilde{\varphi}$ such that $\|\widetilde{\varphi}\|_{2+\alpha} \leq \varepsilon$ and $(t,m) \to \langle U(t,m) - (\varphi_* + \widetilde{\varphi}), m - \nu_*\rangle - (\psi_*(t) + \delta t)$ has a strict minimum on $[0, t_0] \times \mathcal{P}(\mathbb{T})$ at (t^*, m^*). Because U is a monotone solution of (71)

$$\begin{aligned} &-\frac{d\psi_*(t^*)}{dt} + \delta + \langle H(x, \partial_x U) + \lambda(U - \mathcal{A}^* U(\mathcal{A}m^*)), m^* - \nu_*\rangle \geq \langle f(m^*), m^* - \nu_*\rangle \\ &- \langle U - \varphi_* - \widetilde{\varphi}, \partial_x(\partial_p H(x, \partial_x U)m^*)\rangle + \sigma\langle \partial_{xx}(\varphi_* + \widetilde{\varphi}), m^*\rangle - \sigma\langle \partial_{xx} U(t^*, m^*), \nu_*\rangle. \end{aligned} \tag{102}$$

Consider now that M and ε are fixed. Take $\varepsilon' > 0, \overline{\varphi}$ and $\overline{\delta}$ smaller than ε and consider now a strict minimum $(\overline{t}, \overline{m})$ of $(t,m) \to \langle U(t,m) - (\varphi_M + \widetilde{\varphi} + \overline{\varphi}), m - \nu_M\rangle - (\psi_M(t) + (\delta + \overline{\delta})t)$. Given that M is large enough, if ε and ε' are sufficiently small, then $(\overline{t}, \overline{m})$ is sufficiently close to (t_M, m_M). The uniformity in M large enough comes from the uniform continuity of U and its derivatives and from the convergence of the sequence $(\varphi_M, \psi_M, \nu_M)_{M\geq 0}$. On the other hand, from the same argument, given that M is large enough, $(\overline{t}, \overline{m})$ is sufficiently close to (t^*, m^*). Using once again the uniform continuity of U and its derivatives, and the convergence of $(\varphi_M, \psi_M, \nu_M)_{M\geq 0}$, we obtain a contradiction by comparing (101) and (102).

□

Theorem 18. *Any function V given by Proposition 15 is equal to the unique monotone solution of* (71).

Proof. Denote by U the unique monotone solution of (71) and by V a function given by Proposition 15. Assume that

$$\inf_{t\in[0,T], m, m' \in \mathcal{P}(S^n)} \langle U(t,m) - V(t,m'), m - m'\rangle < 0. \tag{103}$$

Hence, using the uniform continuity of U and V, there exists $\kappa > 0$, such that for any M large enough, and any $\gamma > 0$

$$\inf_{t,s\in[0,T]^2, m\in\mathcal{P}_M, m'\in\mathcal{P}(\mathbb{T})} \langle U(t,m) - V(s,m'), m - m'\rangle + \gamma(t-s)^2 \leq -\kappa, \tag{104}$$

where $\mathcal{P}_M = \{m \in \mathcal{P}(\mathbb{T}), \|m\|_{2,\infty} \leq M\}$, which is a compact set. Hence, thanks to Stegall's Lemma, for $\varepsilon > 0$ sufficiently small there exists $\delta, \delta' \in ((4T)^{-1}\kappa, (2T)^{-1}\kappa)$, $\varphi, \varphi' \in C^2$ such that $\|\varphi\|_2 + \|\varphi'\|_2 \leq \varepsilon$ and

$$(t,s,m,m') \to \langle U(t,m) - V(s,m'), m - m'\rangle + \gamma(t-s)^2 + \langle \varphi, m\rangle + \langle \varphi', m'\rangle + \delta(T-t) + \delta'(T-s) \tag{105}$$

has a strict minimum on $[0,T]^2 \times \mathcal{P}_M \times \mathcal{P}(\mathbb{T})$ at (t_*, s_*, m_*, m'_*) which is less than $-\frac{\kappa}{2}$. Assume first that $t_*, s_* > 0$. Using Lemma 16 at this point, we obtain that

$$\begin{aligned}&-\delta' - 2\gamma(s_* - t_*) + \langle H(x,\partial_x V) + \lambda(V(s_*, m'_*) - \mathcal{A}^* V(s_*, \mathcal{A}m'_*)), m'_* - m_*\rangle \geq \langle f(m'_*), m'_* - m_*\rangle \\ &- \langle V(s_*, m'_*) - U(t_*, m_*) + \varphi', \partial_x(\partial_p H(x,\partial_x V)) m'_*)\rangle + \sigma\langle \partial_{xx}(U(t_*, m_*) - \varphi'), m'_*\rangle - \sigma\langle \partial_{xx} V(s_*, m'_*), m_*\rangle.\end{aligned} \tag{106}$$

On the other hand, Lemma 17 yields

$$\begin{aligned}-\delta - 2\gamma(t_* - s_*) + &\langle H(x,\partial_x U) + \lambda(U(t_*, m_*) - \mathcal{A}^* U(t_*, \mathcal{A}m_*)), m_* - m'_*\rangle \geq \langle f(m_*), m_* - m'_*\rangle \\ &- \langle U(t_*, m_*) - V(s_*, m'_*) + \varphi, \partial_x(\partial_p H(x,\partial_x U) m_*)\rangle \\ &+ \sigma\langle \partial_{xx}(V(s_*, m'_*) - \varphi), m_*\rangle - \sigma\langle \partial_{xx} U(t_*, m_*), m'_*\rangle - \omega(M).\end{aligned} \tag{107}$$

Combining the two relations, using the convexity of H and the monotonicity of f yield

$$\begin{aligned}&-\delta - \delta' + \lambda(\langle U(t_*, m_*) - V(s_*, m'_*), m_* - m'_*\rangle - \langle U(t_*, \mathcal{A}m_*) - V(s_*, \mathcal{A}m'_*), \mathcal{A}m_* - \mathcal{A}m'_*\rangle) \\ &\quad \geq -\omega(M) - \langle \varphi', \partial_x(\partial_p H(x,\partial_x V) m'_*)\rangle - \sigma\langle \partial_{xx}\varphi', m'_*\rangle - \langle \varphi, \partial_x(\partial_p H(x,\partial_x U) m_*)\rangle - \sigma\langle \partial_{xx}\varphi, m_*\rangle.\end{aligned} \tag{108}$$

Hence, if ε is chosen small enough, we obtain that

$$-\frac{\kappa}{2T} \geq -\omega(M), \tag{109}$$

which is a contradiction if M is large enough.

Consider now the case $t_* = 0$ (the case $s_* = 0$ is similar). In this situation, using the continuity of U and V, taking γ sufficiently large immediately contradicts (104). Hence (104) is false and

$$\inf_{t\in[0,T], m, m' \in \mathcal{P}(S^n)} \langle U(t,m) - V(t,m'), m - m'\rangle \geq 0. \tag{110}$$

From this we deduce, as in [8] for instance, that $\nabla_x U = \nabla_x V$. Once this is established, to obtain the equality between U and V follows exactly as in [8] from the strict monotonicity of f, using Lemmata 16 and 17 instead of the usual definition of monotone solutions. Hence we do not detail this argument here. □

5.6. **Rate of convergence to a classical solution.** In this section we establish a rate for the convergence of $(U^n)_{n\geq 0}$ toward U when U is a classical solution of 71. To simplify the following discussion we assume that the master equations are set on $\mathcal{M}_2(\mathbb{T})$, the set of positive measures of mass at most 2 on $\mathbb{T}$. We assume that f and g are indeed defined and monotone on $\mathcal{M}_2(\mathbb{T})$. We also assume that f and g satisfy the requirements of Proposition 9 where by extension,

$$\mathbf{d}_1(\mu,\nu) = \inf_{\varphi} \langle \varphi, \mu - \nu\rangle, \tag{111}$$

where the supremum is taken over 1-Lipschitz functions φ such that $\varphi(0) = 0$.

We thus assume that there exists U, a classical solution of (71) on $[0,T] \times \mathbb{T} \times \mathcal{M}_2(\mathbb{T})$. By extension we consider the master equation in finite state space (73) on $[0,T] \times S^n \times \mathcal{M}_2(S^n)$.

The associated concepts of monotone solution on $\mathcal{M}_2(\mathbb{T})$ or $\mathcal{M}_2(S^n)$ are exactly the same as before except for replacing $\mathcal{P}$ by $\mathcal{M}_2$ in the Definitions 7 and 8.

We proceed as in the case without common noise and consider V^n defined by $V^n(t,x,m) = U(t,x,m)$ on $[0,T] \times S^n \times \mathcal{M}_2(S^n)$. As in the case without common noise, the following holds.

Proposition 19. *The function V^n satisfies*

$$-\partial_t V^n(t,\cdot,m) + (F^n(m,V^n) \cdot \nabla_m) V^n + \lambda(V^n - A_n^* V^n(t, A_n m)) = G^n(m,V^n) + r^n. \tag{112}$$

with $|r^n(t,x,m)| \leq C\omega(\frac{1}{n})$, where ω is a modulus of continuity of $\partial_x U$, $\partial_x^2 U$, $D^m U$, $\partial_y D^m U(\cdot,y)$. In particular, it is a monotone solution of this equation.

The following holds.

Theorem 20. *There exists $C > 0$ such that*

$$\sup_{t\in[0,T],m\in\mathcal{P}(\mathbb{T})} \|U^n(t,\cdot,m) - V^n(t,\cdot,m)\|_\infty \leq C\left(\omega\left(\frac{1}{n}\right)\right)^{\frac{1}{3}}. \tag{113}$$

Proof. Define W by $W(t,m,m') = \langle U^n(t,m) - V^n(t,m'), m - m'\rangle$ and κ by

$$-\kappa = \inf_{t\in[0,T],m,m'\in\mathcal{M}_2(S^n)} W(t,m,m'). \tag{114}$$

Using the fact that U^n is a monotone solution of (73) and V^n a monotone solution of (112), we arrive at

$$-\frac{\kappa}{2T} \geq -\inf_{t,m,m'\in\mathcal{M}_2(S^n)} \langle r^n(t,m), m - m'\rangle \geq -C\omega\left(\frac{1}{n}\right). \tag{115}$$

Take $t\in[0,T], m\in\mathcal{M}_2(S^n)$ and $z\in\mathcal{M}_2(S^n)$. From the previous estimate we obtain for any $h\in(0,1)$

$$\langle V^n(t,(1-h)m+hz) - U^n(t,m), h(z-m)\rangle \geq -C\omega\left(\frac{1}{n}\right). \tag{116}$$

Thus,

$$h\langle V^n(t,m)-U^n(t,m), z-m\rangle + \langle V^n(t,(1-h)m+hz) - V^n(t,m), h(z-m)\rangle \geq -C\omega\left(\frac{1}{n}\right). \tag{117}$$

Using the Lipschitz regularity of U (hence of V^n)

$$\langle V^n(t,m) - U^n(t,m), z-m\rangle \geq -\frac{C}{h}\omega\left(\frac{1}{n}\right) - Ch. \tag{118}$$

It follows that

$$\inf_{z\in\mathcal{M}_2(S^n)} \langle V^n(t,m) - U^n(t,m), z-m\rangle \geq -C\sqrt{\omega\left(\frac{1}{n}\right)}. \tag{119}$$

Arguing similarly for U^n and using the uniform Hölder estimate established, we arrive at

$$\inf_{z\in\mathcal{M}_2(S^n)} \langle V^n(t,m) - U^n(t,m), m-z\rangle \geq -\frac{C}{h}\omega\left(\frac{1}{n}\right) - C\sqrt{h}. \tag{120}$$

Hence we deduce, using the previous estimate, that

$$\sup_{z,m\in\mathcal{M}_2(S^n)} |\langle V^n(t,m) - U^n(t,m), z-m\rangle| \leq C\left(\omega\left(\frac{1}{n}\right)\right)^{\frac{1}{3}} \tag{121}$$

Now take $m, m'\in\mathcal{P}(\mathbb{T})$. From the previous estimate we obtain by choosing $z = m + m'$

$$|\langle V^n(t,m) - U^n(t,m), m'\rangle| \leq C\left(\omega\left(\frac{1}{n}\right)\right)^{\frac{1}{3}}, \tag{122}$$

from which we obtain the required result.

□

5.7. **A weaker notion of monotone solution.** We conclude this part on the common noise by indicating another definition of monotone solution which could have been used here and that we believe to have an interest in itself. This concept allows to deal with monotone solution of (71) which are $C^{1,\alpha}$ for $\alpha\in(0,1)$ with respect to the space variable x. This allows to avoid the assumption on the uniform continuity of the space Laplacian of the solution with respect to the measure. The following is very much in the flavor of the work done in [14] in which this method is introduced to deal with first order MFG.

The definition at interest here is

Definition 21. *A continuous function U, uniformly $C^{1,\alpha}$ in space, is a monotone solution of* (71) *if there exists a constant $C > 0$ such that for any $\varepsilon > 0, \nu \in \mathcal{M}(\mathbb{T}) \cap W^{1,\infty}$, $C^{1,\alpha}$ function φ of the space variable and C^1 function ψ of the time variable, for any $(t_*, m_*) \in [0,T) \times \mathcal{P}(\mathbb{T})$ point of strict minimum of $(t,m) \to \langle U(t,m) - \varphi, m - \nu\rangle - \psi(t) + \varepsilon\|m\|_{1,\infty}$, the following holds*

$$\begin{aligned} &-\frac{d\psi(t_*)}{dt} + \langle -\sigma\partial_{xx}U + H(x,\partial_x U) + \lambda(U - \mathcal{A}^*U(\mathcal{A}m_*)), m_* - \nu\rangle \geq \langle f(m_*), m_* - \nu\rangle \\ &- \langle U - \varphi, \partial_x(\partial_p H(x,\partial_x U)m_*)\rangle - \sigma\langle \partial_{xx}(U-\varphi), m_*\rangle - C\varepsilon. \end{aligned} \tag{123}$$

The main idea of this definition is to use the fact that, independently of the strategies of the players, the evolution of the underlying repartition of players is continuous in a space of regular repartition of players. The constant C in the previous is directly related to this smoothness.

In some sense, the penalization term in ε in the minimization of the function constrains the minima to be in $W^{1,\infty}$, and this penalization only has a cost $C\varepsilon$ because of the smoothness of the evolution of the repartition of players. To derive this formulation, consider a classical solution U of (71). To lighten notation, we do not come back on the interpretation of the time derivative or of the common noise. Hence, we take $\lambda = 0$ and consider $t \geq 0$ and a point m_* of minimum of $m \to \langle U(t,m) - \varphi, m - \nu\rangle + \varepsilon\|m\|_\infty$. Denote by $(m_s)_{s\geq 0}$ the solution of

$$\partial_s m - \sigma\partial_{xx}m - \partial_x(m\partial_p H(x, \partial_x U(t,x,m_s))) = 0 \text{ in } (0,\infty)\times\mathbb{T} \tag{124}$$

with initial condition $m_0 = m_*$. By definition of m_*, for any $s \geq 0$:

$$\langle U(t,m_*) - \varphi, m_* - \nu\rangle + \varepsilon\|m_*\|_\infty \leq \langle U(t,m_s) - \varphi, m_s - \nu\rangle + \varepsilon\|m_s\|_\infty. \tag{125}$$

Hence using the fact that U is a classical solution of (71), we deduce by dividing the previous inequality by s and letting $s \to 0$ that

$$\begin{aligned} \langle -\partial_t U, m_* - \nu\rangle + \langle -\sigma\partial_{xx}U + H(x,\partial_x U), m_* - \nu\rangle \geq \langle f(m_*), m_* - \nu\rangle - \langle U - \varphi, \partial_x(\partial_p H(x,\partial_x U)m_*)\rangle \\ - \sigma\langle\partial_{xx}(U-\varphi), m_*\rangle + \varepsilon \liminf_{s\to 0} s^{-1}(\|m_*\|_{1,\infty} - \|m_s\|_{1,\infty}). \end{aligned} \tag{126}$$

However, since U is $C^{1,\alpha}$ in x, uniformly in t and m, there exists $C > 0$ such that, for any m_*

$$\liminf_{s\to 0} s^{-1}(\|m_*\|_{1,\infty} - \|m_s\|_{1,\infty}) \geq -C. \tag{127}$$

This last inequality is a consequence of propagation of $\|\cdot\|_{1,\infty}$ norms by the Fokker-Planck equation

$$\partial_t m - \sigma\Delta m + \text{div}(bm) = 0 \text{ in } (0,\infty)\times\mathbb{T}^d, \tag{128}$$

for a vector field b in $L^\infty((0,\infty), C^{0,\alpha})$. The proof is trivial in dimension 1 as one can simply integrate the Fokker-Planck equation and use standard parabolic estimates in Hölder norms. In dimension $d \geq 1$ the proof of such a regularity is more involved but it remains true. As this question is far from the main topic of this article we do not detail such a proof here.

As a consequence of the previous remark, results of existence and uniqueness of such monotone solutions can be established quite easily following [8, 14].

References

[1] Achdou, Y., and Capuzzo-Dolcetta, I. (2010). Mean field games: numerical methods. SIAM Journal on Numerical Analysis, 48(3), 1136-1162.

[2] Achdou, Y., Camilli, F., and Capuzzo-Dolcetta, I. (2012). Mean field games: numerical methods for the planning problem. SIAM Journal on Control and Optimization, 50(1), 77-109.

[3] Achdou, Y., Camilli, F., and Capuzzo-Dolcetta, I. (2013). Mean field games: convergence of a finite difference method. SIAM Journal on Numerical Analysis, 51(5), 2585-2612.

[4] Almulla, N., Ferreira, R., and Gomes, D. (2017). Two numerical approaches to stationary mean-field games. Dynamic Games and Applications, 7(4), 657-682.

[5] Bayraktar, E., Cecchin, A., Cohen, A., and Delarue, F. (2021). Finite state mean field games with wright?fisher common noise. Journal de Mathématiques Pures et Appliquées, 147, 98-162.

[6] Benamou, J. D., and Carlier, G. (2015). Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations. Journal of Optimization Theory and Applications, 167(1), 1-26.

[7] Bertucci, C. (2021). Monotone solutions for mean field games master equations: finite state space and optimal stopping. Journal de l'École polytechnique–Mathématiques, 8, 1099-1132.
[8] Bertucci, C. (2021). Monotone solutions for mean field games master equations: continuous state space and common noise. arXiv preprint arXiv:2107.09531.
[9] Bertucci, C. (2020). A remark on Uzawa?s algorithm and an application to mean field games systems. ESAIM: Mathematical Modelling and Numerical Analysis, 54(3), 1053-1071.
[10] Bertucci, C., Lasry, J. M., and Lions, P. L. (2019). Some remarks on mean field games. Communications in Partial Differential Equations, 44(3), 205-227.
[11] Briceno-Arias, L. M., Kalise, D., and Silva, F. J. (2018). Proximal methods for stationary mean field games with local couplings. SIAM Journal on Control and Optimization, 56(2), 801-836.
[12] V. I. Bogachev, A. N. Doledenok , and S. V. Shaposhnikov. Weighted Zolotarev Metrics and the Kantorovich Metric. Doklady Mathematics, 2017, Vol. 95, No. 2, pp. 113-117.
[13] Cardaliaguet, P., Delarue, F., Lasry, J. M., and Lions, P. L. (2019). The master equation and the convergence problem in mean field games. Princeton University Press.
[14] Cardaliaguet, P., and Souganidis, P. (2021). Weak solutions of the master equation for Mean Field Games with no idiosyncratic noise. arXiv preprint arXiv:2109.14911.
[15] Carmona, R., and Delarue, F. (2018). Probabilistic Theory of Mean Field Games with Applications I-II. Springer Nature.
[16] Cecchin, A., and Delarue, F. (2021). Selection by vanishing common noise for potential finite state mean field games. Communications in Partial Differential Equations, 1-76.
[17] Cecchin, A., Pra, P. D., Fischer, M., and Pelino, G. (2019). On the convergence problem in mean field games: a two state model without uniqueness. SIAM Journal on Control and Optimization, 57(4), 2443-2466.
[18] Cecchin, A., and Pelino, G. (2019). Convergence, fluctuations and large deviations for finite state mean field games via the master equation. Stochastic Processes and their Applications, 129(11), 4510-4555.
[19] Gangbo, W., Mészáros, A. R., Mou, C., and Zhang, J. (2021). Mean field games master equations with non-separable hamiltonians and displacement monotonicity. arXiv preprint arXiv:2101.12362.
[20] Gangbo, W., and Mészáros, A. R. (2020). Global well-posedness of Master equations for deterministic displacement convex potential mean field games. arXiv preprint arXiv:2004.01660.
[21] Hadikhanloo, S., and Silva, F. J. (2019). Finite mean field games: fictitious play and convergence to a first order continuous mean field game. Journal de Mathématiques Pures et Appliquées, 132, 369-397.
[22] Huang, M., Malhamé R. P., and Caines, P. E. (2006). Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. Communications in Information and Systems, 6(3), 221-252.
[23] Katznelson, Y. (2004). An introduction to harmonic analysis. Cambridge University Press.
[24] Lasry, J. M., and Lions, P. L. (2007). Mean field games. Japanese journal of mathematics, 2(1), 229-260.
[25] Lions, P. L. (2007). Cours au college de france. Available at www.college-de-france.fr
[26] Mou, C., and Zhang, J. (2019). Wellposedness of second order master equations for mean field games with nonsmooth data. arXiv preprint arXiv:1903.09907.
[27] Mou, C., and Zhang, J. (2022). Mean Field Game Master Equations with Anti-monotonicity Conditions. arXiv preprint arXiv:2201.10762.
[28] Phelps, R. R. (2009). Convex functions, monotone operators and differentiability (Vol. 1364). Springer.
[29] Stegall, C. (1978). Optimization of functions on certain subsets of Banach spaces. Mathematische Annalen, 236(2), 171-176.

(C. Bertucci)
Centre de Mathématiques Appliquées, UMR 7641, École Polytechnique
Route de Saclay, 91124 Palaiseau Cedex, France

(A. Cecchin)
Centre de Mathématiques Appliquées, UMR 7641, École Polytechnique
Route de Saclay, 91124 Palaiseau Cedex, France