The master equation and the convergence problems in mean field games with several populations

Giuseppe Di Fabio Unipd 2054736 Dauphine 22200144

Supervisors: Pierre Cardaliaguet and Marco Cirant

Preface

The purpose of this paper is to develop the mean field game theory with several populations. There are several reasons to explain the developing interest for mean field games, one of which is the strong connection within several mathematical areas, including partial differential equations, probability theory, stochastic analysis, optimal control, and optimal transportation. In particular, several results of mean field games can be expressed and proved by analytical or probabilistic tools [5].Another explanation for the interest in the theory is the wide range of applications that it offers. While they were originally inspired by works in economics [10] on heterogeneous agents, MFG models now appear under various forms in several domains, which include, for instance, mathematical finance, study of crowd phenomena, epidemiology, and cybersecurity.
Mean field games should be understood as games with a continuum of players, each of them interacting with the whole statistical distribution of the population. The intrinsic difficulty in proving the convergence of finite player equilibria may be explained as follows: When taken over strategies in closed Markovian form, Nash equilibria of a stochastic differential game with N players in a state of dimension d may be described through a system of N quasilinear parabolic partial differential equations in dimension $N \times d$, which we refer to as the Nash system. The strategy developed in this paper is thus to bypass any detailed study of the Nash system. Instead, we focus directly on the expected limiting form of the Nash system. This limiting form is precisely what we call the master equation. As a result of the symmetry inherent in the mean field structure, this limiting form is no longer a system of equations but reduces to one equation only, which makes it simpler than the Nash system. It describes the equilibrium cost to one representative player in a continuum of players in each populations. To do so we exploit the connection between the MFG system and master equation and at the end of this paper we show that the solution of the multi-popultaion Nash system converges to the solution of the multi-population master equation.
To conclude I would like to thank Marco Cirant and Pierre Cardaliaguet for all the ideas in the numerous stimulating discussions with them.

1 The differential game

In the following chapter we will introduce in a non rigorous way some basic ideas about mean field games.
The existence of time consistent Nash equilibria, based on dynamic programming, requires the solvability of a strongly coupled system of Hamilton-Jacobi equations. This system is denominated the **Nash system**.

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

Where $(v^{N,i})_{i\in\{1,\dots,N\}}$ are called value functions, $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the Hamiltonian and the symbol $\cdot$ denotes the inner product in the $d-$dimensional euclidean space $\mathbb{R}^d$.
In the branch of mathematics denominated differential games theory, differential games are just optimal control problems with N players with state $\big((X_{i,t})_{t\in[0,T]}\big)_{i\in\{1,\dots,N\}}$ and control $(\alpha_{i,t})_{t\in[0,T], i\in\{1,\dots,N\}}$ which evolve following this stochastic differential equation (**SDE**):

$$dX_{i,t} = \alpha_{i,t} dt + \sqrt{2} dB_t^i \quad X_{t_0} = x_{i,0}$$

where $\big((B_t^i)_{t\in[0,T]}\big)$ are $d-$dimensional Brownian motion and controls $(\alpha_{i,t})_{t\in[0,T], i\in\{1,\dots,N\}}$ are progressively measurables with respect to the filtration generated by the $d-$dimensional Brownian motion.
Each player in the differential games aims to minimize the following cost functional

$$J_i^N(t_0, x_0, (\alpha_j)_{j=1,\dots,N}) = \mathbb{E}\left[\int_{t_0}^T L(X_{i,s}, \alpha_{i,s}) + F^{N,i}(X_s) ds + G^{N,i}(X_T)\right].$$

The Hamiltonian of the problem is defined as $H(x,p) = \sup_{\alpha\in\mathbb{R}^d}\{-\alpha \cdot p - L(x,\alpha)\}$ $\forall (x,p) \subset \mathbb{R}^d \times \mathbb{R}^d$.
By Itô' s formula, the optimal controls of the game are:

$$\alpha_i^*(t,x) = -D_p H(x_i, D_{x_i} v^{N,i}(t,x)) \quad i = 1, \dots, N$$

which provide a **Nash equilibrium**:

$$v^{N,i}(t_0, x_0) = J_i^N(t_0, x_0, (\alpha_j^*)_{j=1,\dots,N}) \leq J_i^N(t_0, x_0, \alpha_i, (\alpha_j^*)_{j\neq i}).$$

Therefore in this game the set of "optimal trajectories" solves a system of N stochastic differential equations (**SDE**):

$$dX_{i,t} = -D_p H(X_{i,t}, Dv^{N,i}(t, X_t)) dt + \sqrt{2} dB_t^i \quad t \in [0,T], \; i \in \{1, \dots, N\}.$$

In order to assume that the other players are indistinguishible we will suppose that the maps $F^{N,i}, G^{N,i}$ only depend on x_i and the empirical distribution of the variables $(x_j)_{j\neq i}$:

$$F^{N,i}(x) = F(x_i, m_X^{N,i}) \quad G^{N,i}(x) = G(x_i, m_X^{N,i}),$$

where $m_X^{N,i} = \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j}$ is the empirical distribution of the $(x_j)_{j\neq i}$ and F, G maps from $\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ to $\mathbb{R}$.
With these assumptions $v^{N,i}$ can be written as

$$v^{N,i}(t,x) = v^N(t, x_i, m_X^{N,i}) \;\; t \in [0,T] \; x \in (\mathbb{R})^d.$$

The coupled system of stochastic differential equations (**SDEs**) becomes:

$$dX_{i,t} = b\Big(X_{i,t}, \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j,\tau}\Big)dt + \sqrt{2}dB_t^i, \;\; t\in[0,\tau], \;\; i \in \{1,\dots,N\},$$

where $b : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ and if b is bounded and Lipschitz in both variables both marginal laws of $\big(\,(X_{i,t})_{t\in[0,T]}\big)_{i\in\{1,\dots,N\}}$ with same initial condition as $N \to \infty$ converges to the solution of the **Mckean-Vlasov equation**

$$\partial_t m - \triangle m + div(m \cdot b(\cdot, m)) = 0.$$

Due to the convergence of the trajectories $\big(\,(X_{i,t})_{t\in[0,T]}\big)_{i\in\{1,\dots,N\}}$ with the trajectories $\big(\,(Y_{i,t})_{t\in[0,T]}\big)_{i\in\{1,\dots,N\}}$, which satisfy the following (**SDEs**):

$$dY_{i,t} = b(Y_{i,t}, \mathcal{L}(Y_{i,t}))dt + \sqrt{2}dB_t^i, \;\; t\in[0,\tau], \;\; i\in\{1,\dots,N\},$$

where $\mathcal{L}(Y_{i,t})$ is the law of $Y_{i,t}$, we obtain the **mean field game system** (**MFG** system) which describes the structure of a differential game with infinitely many players:

$$\begin{cases} -\partial_t u - \triangle u + H(x, D_x u) = F(x, m(t)) \quad \text{in } [0,T]\times\mathbb{R}^d \\ \partial_t m - \triangle m - div(m\cdot D_p H(x, D_x U)) = 0 \quad \text{in } [0,T]\times(\mathbb{R}^d)^N \\ u(T,x) = G(x, m(T)), \;\; m(0,\cdot) = m_0, \end{cases} \tag{2}$$

where m_0 denotes the initial state of the population.
The **MFG** system consists in a **Hamilton-Jacobi** equation describing the value function u of the players and a **Kolmogorov** equation describing the dynamics of the distribution of the population $m(t)$.

2 Notations and definitions

The set $\mathcal{P}(\mathbb{R}^d)$ of Borel probability measures is endowed with the **Monge-Kantarovich** distance:

$$d_1(m,m') = \sup_{\phi\, 1-Lip} \int_{\mathbb{R}^d} \phi(y) d(m-m')(y),$$

where the sup is over all $1-Lipschitz$ continuous maps ϕ and the measures have finite first order moment.
When the probability measure m is absolutely continuous with respect to the Lebesque measure, we use the same letter m to identify its **density**. We often consider flows of time dependent measures of the form $(m(t))_{t\in[0,T]}$ with $m(t) \in \mathcal{P}(\mathbb{R}^d)\, \forall t \in [0,T]$.
When, at each time $t \in [0,T]$, $m(t)$ is absolutely continuous with respect to the Lebesque measure on $\mathbb{R}^d$, we identify $m(t)$ with its density.

If $\phi : \mathbb{R}^d \to \mathbb{R}$ is sufficiently smooth and $\ell = (\ell_1, \dots, \ell_d) \in \mathbb{N}^d$, then $D^\ell \phi$ denominates the derivatives $\frac{\partial^{\ell_1}}{\partial^{\ell_1}_{x_1}} \dots \frac{\partial^{\ell_d}}{\partial^{\ell_d}_{x_d}} \phi$. The order of derivation $|\ell|$ denotes $\ell_1 + \ell_2 + \cdots + \ell_d$. Given $e \in \mathbb{R}^d$, we also denote by $\partial_e \phi$ the directional derivative of ϕ in the direction of e. For $n \in \mathbb{N}$ and $\alpha \in (0,1)$, $\mathcal{C}^{n+\alpha}$ is the set of maps ϕ for which $D^\ell \phi$ is defined and $\alpha-$Hölder continuous for any $\ell \in \mathbb{N}^d$ with $|\ell| \leq n$. We set

$$\|\phi\|_{n+\alpha} := \sum_{\ell \leq n} \sup_{x\in\mathbb{R}^d} |D^\ell \phi(x)| + \sum_{|\ell|=n} \sup_{x\neq x'} \frac{|D^\ell\phi(x) - D^\ell\phi(x')|}{|x-x'|^\alpha}.$$

The dual space of $\mathcal{C}^{n+\alpha}$ is denoted by $(\mathcal{C}^{n+\alpha})'$ with norm

$$\forall \rho \in (\mathcal{C}^{n+\alpha})' \quad \|\rho\|_{-(n+\alpha)} := \sup_{\|\phi\|_{n+\alpha}=1} <\rho, \phi>_{(\mathcal{C}^{n+\alpha})', \mathcal{C}^{n+\alpha}} .$$

To simplify the notation we will abbreviate the expression $<\rho,\phi>_{(\mathcal{C}^{n+\alpha})',\mathcal{C}^{n+\alpha}}$ into $<\rho,\phi>_{n+\alpha}$. If a smooth map ψ depends on two variables, e.g. $\psi = \psi(x,y)$, we set

$$\|\psi\|_{(m,n)} := \sum_{|\ell|\leq m, |\ell'|\leq n} \|D^{(\ell,\ell')}\phi\|_\infty,$$

and, if moreover the derivatives are Hölder continuous,

$$\|\psi\|_{(m+\alpha,n+\alpha)} := \|\psi\|_{(m,n)} + \sum_{|\ell|=m,|\ell'|=n} \sup_{(x,y)\neq(x',y')} \frac{|D^{(\ell,\ell')}\phi(x,y) - D^{(\ell,\ell')}\phi(x',y')}{|x-x'|^\alpha + |y-y'|^\alpha}$$

and the notation is generalized in an obvious way to mapping depending on 3 or more variables.

2.1 Derivatives

Definition 2.1.1. *We say that $U : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ is $\mathcal{C}^1$ if there exists a continuous map $\frac{\delta U}{\delta m} : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ such that, for any $m, m' \in \mathcal{P}(\mathbb{R}^d)$,*

$$\lim_{s \in 0^+} \frac{U((1-s)m + sm') - U(m)}{s} = \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m, y) d(m' - m)(y).$$

Note that $\frac{\delta U}{\delta m}$ is defined up to a constant, therefore we will assume without loss of generality that $\int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m, y) d(m)(y) = 0$.
For any $m \in \mathcal{P}(\mathbb{R}^d)$ and any signed measure μ on $\mathbb{R}^d$, we will use equivalently the following notations: $\frac{\delta U}{\delta m}(\mu)$ and $\int_{\mathbb{R}^d} \frac{\delta U}{\delta m} d\mu(y)$. From the above definition we will deduce that

$$U(m) - U(m') = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}((1-s)m + sm', y) d(m' - m)(y) ds \quad \forall\, m, m' \in \mathcal{P}(\mathbb{R}^d)$$

and

$$\begin{aligned} |U(m) - U(m')| &\le \int_0^1 \left\| D_y \frac{\delta U}{\delta m}\Big((1-s)m + sm', \cdot\Big) \right\|_\infty ds\, d_1(m, m') \\ &\le \sup_{m''} \left\| D_y \frac{\delta U}{\delta m}(m'', \psi) \right\|_\infty ds\, d_1(m, m'). \end{aligned}$$

This last inequality leads to the definition of intrinsic derivative:

Definition 2.1.2. *If $\frac{\delta U}{\delta m}$ is of class $\mathcal{C}^1$ with respect to the second variable, the intrinsic derivative $D_m U : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}^d$ is defined by*

$$D_m U(m, y) := D_y \frac{\delta U}{\delta m}(m, y).$$

The expression $D_m U$ can be understood as a derivative of U along vector fields:

Proposition 2.1.3. *Assume that U is $\mathcal{C}^1$, with $\frac{\delta U}{\delta m}$ being $\mathcal{C}^1$ with respect to y, and that $D_m U$ is continuous in both variables. Let $\phi : \mathbb{R}^d \to \mathbb{R}^d$ be a Borel measurable and bounded vector field. Then,*

$$\lim_{h \to 0} \frac{U((id + h\phi)\#m) - U(m)}{h} = \int_{\mathbb{R}^d} D_m U(m, y) \cdot \phi(y) dm(y).$$

Proof. Let us set $m_{h,s} := s(id + h\phi)\#m + (1-s)m$, where $\#$ push-forward. Then,

$$\begin{aligned} &U((id + h\phi)\#m) - U(m) \\ &= \int_0^1 \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m_{h,s}, y) d((id + h\phi)\#m - m)(y) ds \\ &= \int_0^1 \int_{\mathbb{R}^d} \left(\frac{\delta U}{\delta m}(m_{h,s}, y + h\phi(y)) - \frac{\delta U}{\delta m}(m_{h,s}, y) \right) dm(y) ds \\ &= \int_0^1 \int_{\mathbb{R}^d} \left(\frac{\delta U}{\delta m}(m_{h,s}, y + h\phi(y)) - \frac{\delta U}{\delta m}(m_{h,s}, y) \right) dm(y) ds \end{aligned}$$

Dividing by h and letting $h \to 0$ gives the result thanks to the continuity of $D_m U$. □

Note also that, if $U : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ and $\frac{\delta U}{\delta m}$ is $\mathcal{C}^2$ in y, then $D_y D_m U(m, y)$ is a symmetric matrix since

$$D_y D_m U(m, y) = D_y \left(D_y \frac{\delta U}{\delta m} \right)(m, y) = Hess_y \frac{\delta U}{\delta m}(m, y).$$

2.2 Assumptions

Throughout the paper, we will assume that $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is smooth, globally Lipschitz continuous and satisfies the coercivity condition:

$$0 < D^2_{pp} H(x, p) \leq C\mathbb{I}_d.$$

3 The multi-population mean field game system

To describe the structure of a differential game with infinitely many indistinguishable players in various populations, one finds a problem in which each infinitesimal player optimizes his payoff depending on the collective behavior of the others players in his population, and the resulting optimal state of them is exactly distributed according to the state of each of the populations. The resulting system consists in a forward PDE called the Fokker-Planck equation describing the dynamics of the statistical distribution of the population in equilibrium and in a backward PDE called Hamilton-Jacobi-Bellman equation describing the evolution of the optimal expected costs in equilibrium. This is the "**multi-population mean field game system**"

$$\begin{cases} -\partial_t u^\lambda - \triangle u^\lambda + H(x, Du^\lambda) = F(\lambda, x, \mu(t)) & \text{in } [t_0, T] \times \mathbb{R}^d \ \forall \lambda \in \Lambda \\ \partial_t m^\lambda - \triangle m^\lambda - div(m_\lambda \cdot D_p H(x, Du^\lambda)) = 0 & \text{in } [t_0, T] \times \mathbb{R}^d \ \forall \lambda \in \Lambda \\ m^\lambda(t_0, x) = m_0^\lambda(x) & \text{in } \mathbb{R}^d \ \forall \lambda \in \Lambda \\ u^\lambda(T, x) = G(\lambda, x, \mu(T)) & \text{in } \mathbb{R}^d \ \forall \lambda \in \Lambda \\ \mu(t, x) = \int_\Lambda m^\lambda(t, x)\rho(d\lambda) & \text{in } \mathbb{R}^d \end{cases} \tag{3}$$

where m_0^λ denotes the initial state of the population λ which lives in the compact metric space of the populations Λ and $\rho \in \mathcal{P}(\Lambda)$. In this framework, F_λ is the running cost of the population λ and G_λ is the running cost of the population λ.

Theorem 3.0.1 (**Existence and uniqueness**). *Assume that for some $n \geq 1$ and for some $\alpha \in (0, 1)$:*

$$\sup_{\lambda \in \Lambda} \sup_{\mu \in \mathcal{P}(\Lambda \times \mathbb{R}^d)} \left(\|F(\lambda, \cdot, \mu)\|_{n+\alpha} + \left\| \frac{\delta F(\lambda, \cdot, \mu, \cdot)}{\delta \mu} \right\|_{(n+\alpha, n+\alpha)} \right)$$

$$+\sup_{\lambda\in\Lambda} Lip_n\left(\frac{\delta F_\lambda}{\delta\mu}\right) < \infty, \tag{4}$$

and

$$\sup_{\lambda\in\Lambda}\ \sup_{\mu\in\mathcal{P}(\Lambda\times\mathbb{R}^d)} \left(\|G(\lambda,\cdot,\mu)\|_{n+2+\alpha} + \left\|\frac{\delta G(\lambda,\cdot,\mu,\cdot)}{\delta\mu}\right\|_{(n+2+\alpha,n+2+\alpha)}\right)$$

$$+\sup_{\lambda\in\Lambda} Lip_{n+2}\left(\frac{\delta G_\lambda}{\delta\mu}\right) < \infty, \tag{5}$$

where we defined $Lip_n(\frac{\delta F_\lambda}{\delta\mu})$ *by*

$$Lip_n\left(\frac{\delta F_\lambda}{\delta\mu}\right) := \sup_{\mu\neq\tilde{\mu}} (d_1(\mu,\tilde{\mu}))^{-1}\left\|\frac{\delta F}{\delta\mu}(\lambda,\cdot,\mu,\cdot) - \frac{\delta F}{\delta\mu}(\lambda,\cdot,\tilde{\mu},\cdot)\right\|_{(n+\alpha,n+\alpha)}$$

Under the following "multi-population monotonicity assumptions" in which we will assume that the following functions are ρ*-integrable and the resulting integrals non-negative:*

$$\int_\Lambda\int_{\mathbb{R}^d} (F(\lambda,x,\mu_1(t)) - F(\lambda,x,\mu_2(t)))(m_1^\lambda(t) - m_2^\lambda(t))\,dx\,\rho(d\lambda) \geq 0 \tag{6}$$

$$\int_\Lambda\int_{\mathbb{R}^d} (G(\lambda,x,\mu_1(t)) - G(\lambda,x,\mu_2(t)))(m_1^\lambda(t) - m_2^\lambda(t))\,dx\,\rho(d\lambda) \geq 0 \tag{7}$$

for any time t *and measures* m_1^λ, m_2^λ*, where* μ_1 *and* μ_2 *are defined in the following way:* $\mu_1(t,x) = \int_\Lambda m_1^\lambda(t,x)\rho(d\lambda)$*,* $\mu_2(t,x) = \int_\Lambda m_2^\lambda(t,x)\rho(d\lambda)$*.*
For any initial condition $(t_0, m_0^\lambda) \in [0,T]\times\mathcal{P}(\mathbb{R}^d)$*, the multi-population MFG system* (3) *has a set of unique classical solution* $\{(u^\lambda, m^\lambda)\}_{\lambda\in\Lambda}$*, with* $u^\lambda \in \mathcal{C}^{1+\frac{\alpha}{2},2+\alpha}$*.*
Moreover, m^λ *has a continuous, positive density in* $(0,T]\times\mathbb{R}^d$ *and if, in addition,* m_0^λ *is absolutely continuous with a* $\mathcal{C}^{2+\alpha}$ *positive density, then* m^λ *is of class* $\mathcal{C}^{1+\frac{\alpha}{2},2+\alpha}$*.*

Proof. (**Existence**) To prove the existence of a solution we apply Schauder fixed point theorem.
Let C be a big constant and X be the set of time dependent measures $\mu \in \mathcal{C}^0([t_0,T],\mathcal{P}(\Lambda\times\mathbb{R}^d))$ such that:

$$\sup_{t\in[t_0,T],\mathcal{L}(X_t)=\mu(t)} \mathbb{E}[|X_t|^2] \leq C$$

$$d_1(\mu(t),\mu(s)) \leq C|t-s|^{\frac{1}{2}} \quad \forall s,t\in[t_0,T]. \tag{8}$$

Note that, by Ascoli-Arzela theorem, X is a convex compact space for the uniform distance.
Given $\mu\in X$, we consider the solution u^λ to the Hamilton-Jacobi equation

$$\begin{cases} -\partial_t u^\lambda - \triangle u^\lambda + H(x,Du^\lambda) = F(\lambda,x,\mu(t)) & \text{in } [t_0,T]\times\mathbb{R}^d \\ u^\lambda(T,x) = G(\lambda,x,\mu(t)) & \text{in } \mathbb{R}^d \end{cases}$$

Assumption (4) implies that F_λ is Lipschitz continuous in both variables for all $\lambda \in \Lambda$, thus the map $(t,x) \to F(\lambda, x, \mu(t))$ is Hölder continuous in time and space and therefore belongs to $\mathcal{C}^{\frac{1}{2},1}$ with a Hölder constant independent of μ. By assumption (5), the map $x \to G_\lambda(x, \mu(T))$ is of class $\mathcal{C}^{2+\alpha}$ with a constant independent of μ. Thus, by the theory of Hamilton-Jacobi equations with Lipschitz continuous Hamiltonian ([3], Theorem V.6.1), there exists a unique classical solution u^λ to the above equation.
Let now $\tilde{m}^\lambda$ be the weak solution to the Fokker-Planck equation:

$$\begin{cases} \partial_t \tilde{m}^\lambda - \triangle \tilde{m}^\lambda - div(\tilde{m}^\lambda \cdot D_p H(x, Du^\lambda)) = 0 \quad \text{ in } [t_0, T] \times \mathbb{R}^d \\ \tilde{m}^\lambda(t_0, \cdot) = m_0^\lambda \quad \text{ in } \mathbb{R}^d \end{cases}$$

Following [4], the above equation has a unique solution in $\mathcal{C}^0([t_0,T], \mathcal{P}(\mathbb{R}^d))$ in the sense of distribution. Moreover, as D_pH is bounded, $\tilde{\mu} := \int_\Lambda \tilde{m}^\lambda \rho(d\lambda)$ satisfies (8) for a constant C big enough. In particular, $\tilde{\mu}$ belongs to X.This can also be seen by the estimates on a linear estimates in [2] Chapter 3.3(3.16).
This defines a map $\Phi : X \to X$ which, to any $\mu \in X$ associates the map $\tilde{\mu} \in X$.
Next we claim that Φ is continuous. Indeed let $(\mu_\ell)_{\ell \geq 1}$ be a sequence in X converging to $\mu \in X$. For each $\ell \geq 1$, let u_ℓ and $\tilde{\mu}_\ell$ be the corresponding solutions to the Hamilton-Jacobi and the Fokker-Planck equations respectively.
From our previous estimate, the maps $(u_\ell)_{\ell \geq 1}$ are bounded in $\mathcal{C}^{1+\frac{\alpha}{2}, 2+\alpha}$. So, by continuity of F and G, any cluster point of the $(u_\ell)_{\ell \geq 1}$ is a solution associated with μ. By uniqueness of the solution u of this limit problem, the whole sequence $(u_\ell)_{\ell \geq 1}$ converges to u. In the same way, $(\tilde{\mu}_\ell)_{\ell \geq 1}$ converges in X to the unique solution $\tilde{\mu}$ to the Fokker-Planck equation associated with u. This shows the continuity of Φ.
We conclude by Schauder Theorem that Φ has a fixed point which is a solution to the multi-population MFG system (3) restricted to population λ.
Let now (u^λ, m^λ) be the solution to (3) for a fixed λ and assume that m_0^λ has a smooth density. Then m^λ solves the linear parabolic equation

$$\begin{cases} \partial_t m^\lambda - \triangle m^\lambda - Dm^\lambda \cdot D_p H(x, Du) - m^\lambda div(D_p H(x, Du)) = 0 \text{ in } (t_0, T) \times \mathbb{R}^d \\ m^\lambda(0, \cdot) = m_0^\lambda \text{ in } \mathbb{R}^d, \end{cases}$$

with $\mathcal{C}^{\frac{\alpha}{2}, \alpha}$ coefficient and $\mathcal{C}^{\alpha+2}$ initial condition. Thus, by Schauder theory, m^λ is of class $\mathcal{C}^{1+\frac{\alpha}{2}, 2+\alpha}$. If, moreover, m_0^λ is positive, then m^λ remains positive by strong maximum principle.

(Uniqueness) Let $\{(u_1^\lambda, m_1^\lambda)\}_{\lambda \in \Lambda}$ and $\{(u_2^\lambda, m_2^\lambda)\}_{\lambda \in \Lambda}$ two set of solutions of the multi-population MFG system. By the smooth and coercitivity assumption on the Hamiltonian H, m_1^λ and m_2^λ are positive and smooth for all λ in Λ.
The result can be obtained computing $\frac{d}{dt} \int_{\mathbb{R}^d} (u_1^\lambda - u_2^\lambda)(m_1^\lambda - m_2^\lambda)$ and applying integration by parts and the fact that $\{(u_1^\lambda, m_1^\lambda)\}_{\lambda \in \Lambda}$ and $\{(u_2^\lambda, m_2^\lambda)\}_{\lambda \in \Lambda}$ are

two set of solutions of the multi-population MFG system :

$$\frac{d}{dt}\int_{\mathbb{R}^d}(u_1^\lambda - u_2^\lambda)(m_1^\lambda - m_2^\lambda)dx =$$
$$\int_{\mathbb{R}^d}\left[\frac{d}{dt}(u_1^\lambda - u_2^\lambda)\right](m_1^\lambda - m_2^\lambda) + (u_1^\lambda - u_2^\lambda)\left[\frac{d}{dt}(m_1^\lambda - m_2^\lambda)\right]dx =$$
$$\int_{\mathbb{R}^d}[-\triangle(u_1^\lambda - u_2^\lambda) + H(x, Du_1^\lambda) - F(\lambda, x, \mu_1) - H(x, Du_2^\lambda) + F(\lambda, x, \mu_2)](m_1^\lambda - m_2^\lambda)$$
$$+ (u_1^\lambda - u_2^\lambda)[\triangle(m_1^\lambda - m_2^\lambda) + div(m_1^\lambda \cdot D_pH(x, Du_1^\lambda)) - div(m_2^\lambda \cdot D_pH(x, Du_2^\lambda))]dx$$
$$= -\int_{\mathbb{R}^d} m_1^\lambda\Big[H(x, Du_2^\lambda) - H(x, Du_1^\lambda) - D_pH(x, Du_1^\lambda)(Du_2^\lambda - Du_1^\lambda)\Big]dx$$
$$-\int_{\mathbb{R}^d} m_2^\lambda\Big[H(x, Du_1^\lambda) - H(x, Du_2^\lambda) - D_pH(x, Du_2^\lambda)(Du_1^\lambda - Du_2^\lambda)\Big]dx$$
$$-\int_{\mathbb{R}^d}\Big(F(\lambda, x, \mu_1(t)) - F(\lambda, x, \mu_2(t))\Big)(m_1^\lambda - m_2^\lambda)dx,$$

where we used the following integration by parts formula:

$$\int_{\mathbb{R}^d} -\triangle(u_1^\lambda - u_2^\lambda)(m_1^\lambda - m_2^\lambda) + \triangle(m_1^\lambda - m_2^\lambda) = 0,$$
$$\int_{\mathbb{R}^d}(u_1^\lambda - u_2^\lambda)\Big(div(m_1^\lambda \cdot D_pH(x, Du_1^\lambda)) - div(m_2^\lambda \cdot D_pH(x, Du_2^\lambda))\Big)$$
$$= \int_{\mathbb{R}^d} -m_1^\lambda D_pH(x, Du_1^\lambda)\cdot(Du_1^\lambda - Du_2^\lambda) + m_2^\lambda D_pH(x, Du_2^\lambda)\cdot(Du_1^\lambda - Du_2^\lambda).$$

Integrating in time $t \in [t_0, T]$ and in population $\lambda \in \Lambda$ the formula

$\frac{d}{dt}\int_{\mathbb{R}^d}(u_1^\lambda - u_2^\lambda)(m_1^\lambda - m_2^\lambda)$ computed before we obtain:

$$
\begin{aligned}
&\int_\Lambda\int_{\mathbb{R}^d}(G(\lambda,x,\mu_1(T)) - G(\lambda,x,\mu_2(T))(m_1^\lambda(T) - m_2^\lambda(T))\rho(d\lambda)dx\\
&-\int_\Lambda\int_{\mathbb{R}^d}(u_1^\lambda(t_0,x) - u_2^\lambda(t_0,x))(m_0^\lambda - m_0^\lambda)\rho(d\lambda)dx =\\
&\int_\Lambda\int_{\mathbb{R}^d}(G(\lambda,x,\mu_1(T)) - G(\lambda,x,\mu_2(T))(m_1^\lambda(T) - m_2^\lambda(T))\rho(d\lambda)dx =\\
&\int_\Lambda\int_{t\in[t_0,T]}\frac{d}{dt}\int_{\mathbb{R}^d}(u_1^\lambda - u_2^\lambda)(m_1^\lambda - m_2^\lambda)\rho(d\lambda)dtdx =\\
&-\int_\Lambda\int_{t_0}^T\int_{\mathbb{R}^d} m_1^\lambda\Big[H(x,Du_2^\lambda) - H(x,Du_1^\lambda) - D_pH(x,Du_1^\lambda)(Du_2^\lambda - Du_1^\lambda)\Big]\rho(d\lambda)dtdx\\
&-\int_\Lambda\int_{t_0}^T\int_{\mathbb{R}^d} m_2^\lambda\Big[H(x,Du_1^\lambda) - H(x,Du_2^\lambda) - D_pH(x,Du_2^\lambda)(Du_1^\lambda - Du_2^\lambda)\Big]\rho(d\lambda)dtdx\\
&-\int_\Lambda\int_{t_0}^T\int_{\mathbb{R}^d}\Big(F(\lambda,x,\mu_1(t)) - F(\lambda,x,\mu_2(t))\Big)(m_1^\lambda - m_2^\lambda)\rho(d\lambda)dtdx,
\end{aligned}
$$

and combining the first and last inequality with the multi-population monotonicity assumptions (6) and (7), we obtain that:

$$
\begin{aligned}
&\int_\Lambda\int_{t_0}^T\int_{\mathbb{R}^d} m_1^\lambda\Big[H(x,Du_2^\lambda) - H(x,Du_1^\lambda) - D_pH(x,Du_1^\lambda)(Du_2^\lambda - Du_1^\lambda)\Big]\\
&+ m_2^\lambda\Big[H(x,Du_1^\lambda) - H(x,Du_2^\lambda) - D_pH(x,Du_2^\lambda)(Du_1^\lambda - Du_2^\lambda)\Big]\rho(d\lambda)dtdx \le 0,
\end{aligned}
$$

and using the fact that H is strictly convex we obtain that $Du_1^\lambda = Du_2^\lambda$ ρ-almost everywhere. This implies the unicity of set of solutions since m_1^λ and m_2^λ are positive measures . □

4 The multi-population Master equation

The importance of the master equation has been acknowledged by several contributions: see, for instance, the monograph [7] and the companion papers [8] and [9], in which Bensoussan, Frehse, and Yam generalize this equation to mean field type control problems and reformulate it as a partial differential equation (PDE) set on an L^2 space, and [5], where Carmona and Delarue interpret this equation as a decoupling field of forward–backward SDE in infinite dimension. The concept of master equation and the use of the terminology master go back to the seminal lectures of P. L. Lions at the Collège de France [6]. The word master emphasizes the fact that all the information needed to describe the equilibria of the game is contained in a single equation, namely the master equation. The resulting "**multi-population master equation**" is:

$$\begin{cases} -\partial_t U(t,(\lambda,x),\mu) - \triangle_x U(t,(\lambda,x),\mu) \\ +H(x, D_x U(t,(\lambda,x),\mu) - F(\lambda,x,\mu) \\ = \int_{\Lambda\times\mathbb{R}^d} \triangle_y \frac{\delta U}{\delta\mu}\Big(t,(\lambda,x),\mu,(\tilde\lambda,y)\Big)\mu(\tilde\lambda,y)d\tilde\lambda dy \\ - \int_{\Lambda\times\mathbb{R}^d} D_y \frac{\delta U}{\delta\mu}\Big(t,(\lambda,x),\mu,(\tilde\lambda,y)\Big)\cdot D_p H\Big(y, DU(t,(\tilde\lambda,y),\mu)\Big) \\ \mu(t)(\tilde\lambda,y)d\tilde\lambda dy \qquad \text{in } [t_0,T]\times(\Lambda\times\mathbb{R}^d)\times\mathcal{P}(\Lambda\times\mathbb{R}^d), \\ U(T,(\lambda,x),\mu) = G(\lambda,x,\mu) \quad \text{in } (\Lambda\times\mathbb{R}^d)\times\mathcal{P}(\Lambda\times\mathbb{R}^d), \end{cases} \tag{9}$$

where $\mu\in\mathcal{P}(\Lambda\times\mathbb{R}^d)$, $t\in[t_0,T]$ and $(\lambda,x)\in\Lambda\times\mathbb{R}^d$.

Definition 4.0.1. *We say that a map $U : [t_0,T]\times\Lambda\times\mathbb{R}^d\times\mathcal{P}(\Lambda\times\mathbb{R}^d)\to\mathbb{R}$ is a classical solution to the first order master equation if:*
(i) U is continuous in time, space and measure (for the d_1 distance on $\mathcal{P}(\Lambda\times\mathbb{R}^d)$), is of class $\mathcal{C}^2$ in x and $\mathcal{C}^1$ in time ;
(ii) U is of class $\mathcal{C}^1$ with respect to the measure, the first order derivative

$$[t_0,T]\times\Lambda\times\mathbb{R}^d\times\mathcal{P}(\Lambda\times\mathbb{R}^d)\ni(t,(\lambda,x),\mu,(\tilde\lambda,y))\to\frac{\delta U}{\delta\mu}(t,(\lambda,x),\mu,(\tilde\lambda,y)),$$

being continuous in all the arguments except the population ones, $\frac{\delta U}{\delta\mu}$ being twice differentiable in y, the derivatives being continuous in all the arguments except the population ones;
(iii) U satisfies the multi-population master equation (9).

Theorem 4.0.2 (**Existence and uniqueness**). *Assume that for some $n\geq 2$ and for some $\alpha\in(0,1)$:*

$$\sup_{\lambda\in\Lambda}\sup_{\mu\in\mathcal{P}(\Lambda\times\mathbb{R}^d)}\left(\|F(\lambda,\cdot,\mu)\|_{n+1+\alpha}+\left\|\frac{\delta F(\lambda,\cdot,\mu,\cdot)}{\delta\mu}\right\|_{(n+1+\alpha,n+1+\alpha)}\right)$$
$$+\sup_{\lambda\in\Lambda}\sup_{\mu\in\mathcal{P}(\Lambda\times\mathbb{R}^d)}\left\|\frac{\delta^2 F(\lambda,\cdot,\mu,\cdot,\cdot)}{\delta\mu^2}\right\|_{(n+1+\alpha,n+1+\alpha,n+1+\alpha)}+Lip_{n+1}\left(\frac{\delta^2 F_\lambda}{\delta\mu^2}\right)<\infty.$$

and

$$\sup_{\lambda\in\Lambda}\sup_{\mu\in\mathcal{P}(\Lambda\times\mathbb{R}^d)}\left(\|G(\lambda,\cdot,\mu)\|_{n+2+\alpha}+\left\|\frac{\delta G(\lambda,\cdot,\mu,\cdot)}{\delta\mu}\right\|_{(n+2+\alpha,n+2+\alpha)}\right)$$
$$+\sup_{\lambda\in\Lambda}\sup_{\mu\in\mathcal{P}(\Lambda\times\mathbb{R}^d)}\left\|\frac{\delta^2 G(\lambda,\cdot,\mu,\cdot,\cdot)}{\delta\mu^2}\right\|_{(n+2+\alpha,n+2+\alpha,n+2+\alpha)}+Lip_{n+2}\left(\frac{\delta^2 G_\lambda}{\delta\mu^2}\right)<\infty.$$

where we defined $Lip_n(\frac{\delta^2 F_\lambda}{\delta\mu^2})$ *by*

$$Lip_n\left(\frac{\delta^2 F_\lambda}{\delta\mu^2}\right):=\sup_{\mu\neq\tilde{\mu}}(d_1(\mu,\tilde{\mu}))^{-1}\left\|\frac{\delta^2 F}{\delta\mu^2}(\lambda,\cdot,\mu,\cdot,\cdot)-\frac{\delta^2 F}{\delta\mu^2}(\lambda,\cdot,\tilde{\mu},\cdot,\cdot)\right\|_{(n+\alpha,n+\alpha)}$$

Under the following "multi-population monotonicity assumptions" in which we will assume that the following functions are ρ*-integrable and the resulting integrals non-negative:*

$$\int_\Lambda\int_{\mathbb{R}^d}(F(\lambda,x,\mu_1(t))-F(\lambda,x,\mu_2(t)))(m_1^\lambda(t)-m_2^\lambda(t))\,dx\,\rho(d\lambda)\geq 0 \tag{10}$$

$$\int_\Lambda\int_{\mathbb{R}^d}(G(\lambda,x,\mu_1(t))-G(\lambda,x,\mu_2(t)))(m_1^\lambda(t)-m_2^\lambda(t))\,dx\,\rho(d\lambda)\geq 0 \tag{11}$$

for any time t *and measures* m_1^λ, m_2^λ*, where* μ_1 *and* μ_2 *are defined in the following way:* $\mu_1(t,x)=\int_\Lambda m_1^\lambda(t,x)\rho(d\lambda)$*,* $\mu_2(t,x)=\int_\Lambda m_2^\lambda(t,x)\rho(d\lambda)$ *and for any initial condition* $(t_0,m_0^\lambda)\in[0,T]\times\mathcal{P}(\mathbb{R}^d)$*. Then the multi-population master equation (9) has a unique solution* U*. Moreover,* U *and* $\frac{\delta U}{\delta\mu}$ *are continuous in all variables except the population ones;* $U(t,\cdot,\mu)$ *and* $\frac{\delta U}{\delta\mu}(t,\cdot,\mu,\cdot)$ *are bounded in* $\mathcal{C}^{n+2+\alpha}$ *and* $\mathcal{C}^{n+2+\alpha}\times\mathcal{C}^{n+1+\alpha}$ *respectively , independently of* (t,μ)*.*

Proof. (**Existence**). We will prove that the following map:

$$U(t,(\lambda,x),\mu(t)):=u^\lambda(t,x)$$

is a solution to the multi-population master equation, where $\{(u^\lambda,m^\lambda)\}_{\lambda\in\Lambda}$ is the unique set of solution to the multi-population mean field game system (3) and $\mu(t)\in\mathcal{P}(\Lambda\times\mathbb{R}^d)$ has ρ, $m^\lambda(t)$ as marginal measures, respectively.
To compute $\partial_t U(t,(\lambda,x),\mu(t))$ we will compute the limit of both the quantities in the RHS:

$$\begin{aligned}&\frac{U(t_0+h,(\lambda,x),\mu(t_0))-U(t_0,(\lambda,x),\mu(t_0))}{h}\\&=\frac{U(t_0+h,(\lambda,x),\mu(t_0+h))-U(t_0,(\lambda,x),\mu(t_0))}{h}\\&-\frac{U(t_0+h,(\lambda,x),\mu(t_0+h))-U(t_0+h,(\lambda,x),\mu(t_0))}{h}.\end{aligned}$$

Note from the Fokker-Planck equation satisfied by m^λ and the regularity of U ([3] Corollary 3.4.4) we obtain that:

$$\frac{U(t_0+h,(\lambda,x),\mu(t_0+h))-U(t_0+h,(\lambda,x),\mu(t_0))}{h}$$
$$=\frac{1}{h}\int_0^1\int_{\Lambda\times\mathbb{R}^d}\frac{\delta U}{\delta\mu}\Big(t_0+h,(\lambda,x),s\mu(t_0+h)+(1-s)\mu(t_0),(\tilde\lambda,y)\Big)$$
$$(\mu(t_0+h)-\mu(t_0))d\tilde\lambda dyds$$
$$=\frac{1}{h}\int_0^1\int_{\Lambda\times\mathbb{R}^d}\frac{\delta U}{\delta\mu}\Big(t_0+h,(\lambda,x),s\mu(t_0+h)+(1-s)\mu(t_0),(\tilde\lambda,y)\Big)$$
$$\int_{t_0}^{t_0+h}\partial_t\mu(t)(\tilde\lambda,y)dtd\tilde\lambda dyds$$
$$=\frac{1}{h}\int_0^1\int_{\Lambda\times\mathbb{R}^d}\frac{\delta U}{\delta\mu}\Big(t_0+h,(\lambda,x),s\mu(t_0+h)+(1-s)\mu(t_0),(\tilde\lambda,y)\Big)$$
$$\int_{t_0}^{t_0+h}\rho(\tilde\lambda)\partial_t m^{\tilde\lambda}(t,y)dtd\tilde\lambda dyds$$
$$=\frac{1}{h}\int_0^1\int_{\Lambda\times\mathbb{R}^d}\int_{t_0}^{t_0+h}\frac{\delta U}{\delta\mu}\Big(t_0+h,(\lambda,x),s\mu(t_0+h)+(1-s)\mu(t_0),(\tilde\lambda,y)\Big)$$
$$\rho(\tilde\lambda)[\triangle m^{\tilde\lambda}(t,y)+div(m^{\tilde\lambda}(t,y)\cdot D_pH(y,Du^{\tilde\lambda}(t,y)))]dtd\tilde\lambda dyds$$
$$=\frac{1}{h}\int_0^1\int_{\Lambda\times\mathbb{R}^d}\int_{t_0}^{t_0+h}\triangle_y\frac{\delta U}{\delta\mu}\Big(t_0+h,(\lambda,x),s\mu(t_0+h)+(1-s)\mu(t_0),(\tilde\lambda,y)\Big)$$
$$\rho(\tilde\lambda)m^{\tilde\lambda}(t,y)dtd\tilde\lambda dyds$$
$$-\frac{1}{h}\int_0^1\int_{\Lambda\times\mathbb{R}^d}\int_{t_0}^{t_0+h}D_y\frac{\delta U}{\delta\mu}\Big(t_0+h,(\lambda,x),s\mu(t_0+h)+(1-s)\mu(t_0),(\tilde\lambda,y)\Big)$$
$$\cdot D_pH(y,Du^{\tilde\lambda}(t,y))m^{\tilde\lambda}(t,y)\rho(\tilde\lambda)dtd\tilde\lambda dyds$$
$$\xrightarrow[h\to0^+]{}\int_{\Lambda\times\mathbb{R}^d}\triangle_y\frac{\delta U}{\delta\mu}\Big(t_0,(\lambda,x),\mu(t),(\tilde\lambda,y)\Big)\mu(t_0)(\tilde\lambda,y)d\tilde\lambda dy$$
$$-\int_{\Lambda\times\mathbb{R}^d}D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda,x),\mu(t),(\tilde\lambda,y)\Big)\cdot D_pH\Big(y,DU(t,(\tilde\lambda,y),\mu(t))\Big)$$
$$\mu(t_0)(\tilde\lambda,y)d\tilde\lambda dy.$$

On the other hand:

$$
\begin{aligned}
&\lim_{h\to 0^+} \frac{U(t_0+h,(\lambda,x),\mu(t_0+h)) - U(t_0,(\lambda,x),\mu(t_0)))}{h} \\
&= \lim_{h\to 0^+} \frac{u^\lambda(t_0+h,x) - u^\lambda(t_0,x)}{h} \\
&= \partial_t u^\lambda(t_0,x) = -\triangle_x u^\lambda(t_0,x) + H(x,Du^\lambda) - F(\lambda,x,\mu(t_0)) \\
&= -\triangle_x U(t_0,(\lambda,x),\mu(t_0)) + H(x,D_xU(t_0,(\lambda,x),\mu(t_0)) - F(\lambda,x,\mu(t_0)).
\end{aligned}
$$

Therefore combining the two results $\partial_t U(t_0,(\lambda,x),\mu(t_0))$ exists and it is equal to:

$$
\begin{aligned}
&\partial_t U(t_0,(\lambda,x),\mu(t_0)) \\
&= -\int_{\Lambda\times\mathbb{R}^d} \triangle_y \frac{\delta U}{\delta\mu}\Big(t_0,(\lambda,x),\mu(t),(\tilde\lambda,y)\Big)\mu(t_0)(\tilde\lambda,y)d\tilde\lambda dy \\
&+ \int_{\Lambda\times\mathbb{R}^d} D_y \frac{\delta U}{\delta\mu}\Big(t,(\lambda,x),\mu(t),(\tilde\lambda,y)\Big)\cdot D_pH\Big(y,DU(t,(\tilde\lambda,y),\mu(t))\Big) \\
&\mu(t_0)(\tilde\lambda,y)d\tilde\lambda dy - \triangle_x U(t_0,(\lambda,x),\mu(t_0)) \\
&+ H(x,D_xU(t_0,(\lambda,x),\mu(t_0)) - F(\lambda,x,\mu(t_0)).
\end{aligned}
$$

This means that U has a continuous time derivative at any point $(t_0,(\lambda,x),\mu(t_0))$ and satisfies the multi-population master equation (9) at such point. By continuity of the right-hand side and repeating the same computations at any point we obtain that U has a continuous time derivative and satisfies (9) everywhere. The boundary conditions are satisfied by definition.

(**Uniqueness**). Let V be another solution. By definition of a solution, $D^2_{x,y}\frac{\delta V}{\delta m}(t,(\lambda,x),\tilde\mu(t),(\tilde\lambda,y))$ is bounded and therefore D_xV is Lipschitz continuous with respect to the measure variable.
Let us fix (t_0,m_0^λ) and a population λ with $m_0^\lambda \in \mathcal{C}^\infty(\mathbb{R}^d)$.
In view of the Lipschitz continuity of D_xV, one can uniquely solve in $\mathcal{C}^0([t_0,T],\mathcal{P}(\mathbb{R}^d))$ the Fokker-Planck equation:

$$
\begin{cases}
\partial_t\tilde m^\lambda - \triangle\tilde m^\lambda - div(\tilde m^\lambda\cdot D_pH(x,D_xV(t,(\lambda,x),\tilde\mu)) = 0 \quad \text{in } [t_0,T]\times\mathbb{R}^d \\
\tilde m^\lambda(t_0) = m_0^\lambda \quad \text{in } \mathbb{R}^d.
\end{cases}
$$

Then let us set $\tilde u^\lambda(t,x) := V(t,(\lambda,x),\tilde\mu(t))$. Since V is a classical solution by the regularity properties of V, $\tilde{u}^\lambda$ is at least $\mathcal{C}^{1,2}$, with time derivative:

$$\partial_t \tilde{u}^\lambda(t,x) = \lim_{h\to 0} \frac{V(t+h,(\lambda,x),\tilde{\mu}(t+h)) - V(t,(\lambda,x),\tilde{\mu}(t))}{h}$$

$$= \partial_t V(t,(\lambda,x),\tilde{\mu}(t)) + < \frac{\delta V}{\delta \mu}\Big(t,(\lambda,x),\tilde{\mu}(t),(\tilde{\lambda},y)\Big), \partial_t \tilde{\mu}(t)(\tilde{\lambda},y) >_{\mathcal{C}^2,(\mathcal{C}^2)'}$$

$$= \partial_t V(t,(\lambda,x),\tilde{\mu}(t)) + < \frac{\delta V}{\delta \mu}\Big(t,(\lambda,x),\tilde{\mu}(t),(\tilde{\lambda},y)\Big), \partial_t \big(\rho(\tilde{\lambda})\tilde{m}^\lambda(t)(y)\big) >_{\mathcal{C}^2,(\mathcal{C}^2)'}$$

$$= \partial_t V(t,(\lambda,x),\tilde{\mu}(t)) + < \frac{\delta V}{\delta \mu}\Big(t,(\lambda,x),\tilde{\mu}(t),(\tilde{\lambda},y)\Big), \rho(\tilde{\lambda})\partial_t \tilde{m}^\lambda(t)(y) >_{\mathcal{C}^2,(\mathcal{C}^2)'}$$

$$= \partial_t V(t,(\lambda,x),\tilde{\mu}(t)) + < \frac{\delta V}{\delta \mu}\Big(t,(\lambda,x),\tilde{\mu}(t),(\tilde{\lambda},y)\Big), \rho(\tilde{\lambda}) \triangle \tilde{m}^\lambda(t)(y) >_{\mathcal{C}^2,(\mathcal{C}^2)'}$$

$$+ < \frac{\delta V}{\delta \mu}\Big(t,(\lambda,x),\tilde{\mu}(t),(\tilde{\lambda},y)\Big), \rho(\tilde{\lambda}) div(\tilde{m}^\lambda \cdot D_p H(y, D_x V(t,(\tilde{\lambda},y),\tilde{\mu}(t))) >_{\mathcal{C}^2,(\mathcal{C}^2)'}$$

$$= \partial_t V(t,(\lambda,x),\tilde{\mu}(t)) + \int_{\Lambda\times\mathbb{R}^d} \triangle_y \frac{\delta V}{\delta \mu}\Big(t,(\lambda,x),\tilde{\mu}(t),(\tilde{\lambda},y)\Big)\rho(\tilde{\lambda})\tilde{m}^\lambda(t)(y)d\tilde{\lambda}dy$$

$$- \int_{\Lambda\times\mathbb{R}^d} D_y \frac{\delta V}{\delta \mu}\Big(t,(\lambda,x),\tilde{\mu}(t),(\tilde{\lambda},y)\Big) \cdot D_p H(y, D_x V(t,(\tilde{\lambda},y),\tilde{\mu}(t)))$$

$$\rho(\tilde{\lambda})\tilde{m}^\lambda(t)(y)d\tilde{\lambda}dy$$

$$= \partial_t V(t,(\lambda,x),\tilde{\mu}(t)) + \int_{\Lambda\times\mathbb{R}^d} \triangle_y \frac{\delta V}{\delta \mu}\Big(t,(\lambda,x),\tilde{\mu}(t),(\tilde{\lambda},y)\Big)\tilde{\mu}(t)(\tilde{\lambda},y)d\tilde{\lambda}dy$$

$$- \int_{\Lambda\times\mathbb{R}^d} D_y \frac{\delta V}{\delta \mu}\Big(t,(\lambda,x),\tilde{\mu}(t),(\tilde{\lambda},y)\Big) \cdot D_p H(y, D_x V(t,(\tilde{\lambda},y),\tilde{\mu}(t)))$$

$$\tilde{\mu}(t)(\tilde{\lambda},y)d\tilde{\lambda}dy.$$

Recalling that V satisfies the multi-population master equation (9) at population λ, we obtain:

$$\begin{aligned}\partial_t \tilde{u}^\lambda(t,x) &= -\triangle_x V(t,(\lambda,x),\tilde{\mu}(t)) + H(x, D_x V(t,(\lambda,x),\mu(t))) - F(\lambda,x,\tilde{\mu}(t)) \\ &= -\triangle \tilde{u}^\lambda(t,x) + H(x, D\tilde{u}^\lambda(t,x)) - F(\lambda,x,\tilde{\mu}(t))\end{aligned}$$

with $\tilde{\mu}(t,x) = \int_\Lambda \tilde{m}^\lambda(t,x)\rho(d\tilde{\lambda})$ and terminal condition $\tilde{u}^\lambda(T,x) = V(T,(\lambda,x),\tilde{\mu}(T)) = G(\lambda,x,\tilde{\mu}(T))$. Therefore the pair $(\tilde{u}^\lambda, \tilde{m}^\lambda)$ is a solution of the multi-population mean field game system (3) at population λ. As the solution of this system is unique, we get that $V(t_0,(\lambda,x),\mu(t_0)) = U(t_0,(\lambda,x),\mu(t_0))$ if m_0^λ has a smooth density. The equality $V = U$ holds then everywhere by continuity of V and U.
The proof of the differentiability of U with respect to the measure and of the regularities of U and $\frac{\delta U}{\delta \mu}$ is left in [2] and it involves numerous estimates regarding linearized system. □

5 Convergence

Notation. Throughout the rest of this paper we will make use of the classical delta di Dirac δ, defined with the usual density:

$$\delta_{(\lambda_j,x_j)}(\tilde{\lambda},y) = \begin{cases} 1 & \text{if } (\tilde{\lambda},y) = (\lambda_j,x_j), \\ 0 & \text{otherwise.} \end{cases}$$

In this chapter we will consider a set of classical solutions $\{v^{\boldsymbol{\lambda},N,i}\}_{i\in\{1,\dots,N\}}$ of the "**multi-population Nash system**":

$$\begin{cases} -\partial_t v^{\lambda_i,N,i}(t,\boldsymbol{x}) - \sum_{j=1}^{N} \triangle_{x_j} v^{\lambda_i,N,i}(t,\boldsymbol{x}) + H(x_i, D_{x_i} v^{\lambda_i,N,i}(t,\boldsymbol{x})) + \sum_{j\neq i} \\ D_p H(x_j, D_{x_j} v^{\lambda_j,N,j}(t,\boldsymbol{x})) \cdot D_{x_j} v^{\lambda_i,N,i}(t,\boldsymbol{x}) = F(\lambda_i, x_i, \mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}) \text{ in } [t_0,T]\times(\mathbb{R}^d)^N, \\ v^{\lambda_i,N,i}(T,\boldsymbol{x}) = G(\lambda_i, x_i, \mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}) \quad \text{in } (\mathbb{R}^d)^N, \end{cases} \tag{12}$$

where we set $\boldsymbol{x} = (x_1,\dots,x_N) \in (\mathbb{R}^d)^N$, $\boldsymbol{\lambda} = (\lambda_1,\dots,\lambda_N) \in \Lambda^N$, $m_{\boldsymbol{x}}^{N,i} = \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j}$ and $\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i} = \frac{1}{N-1}\sum_{j\neq i}\delta_{(\lambda_j,x_j)}$.

Our aim is to prove that $v^{\boldsymbol{\lambda},N,i}$ converges, in a suitable sense, to the solution of the multi-population master equation $(U(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}))$.

For $N \geq 2$ and $i \in \{1,\dots,N\}$ we define

$$u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}) := U(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}) \quad \text{where } \boldsymbol{x} = (x_1,\dots,x_N) \in (\mathbb{R}^d)^N$$

and $\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i} \in \mathcal{P}(\Lambda\times\mathbb{R}^d)$ is defined with the following density:

$$\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}(\tilde{\lambda},y) := \frac{1}{N-1}\sum_{j\neq i}\delta_{(\lambda_j,x_j)}(\tilde{\lambda},y) = \sum_{j\neq i}\frac{\delta_{\lambda_j}(\tilde{\lambda})\delta_{x_j}(y)}{N-1},$$

which implies that $\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i} \in \mathcal{P}(\Lambda\times\mathbb{R}^d)$ has as marginal measures $\frac{1}{N-1}\sum_{j\neq i}\delta_{\lambda_j} \in \mathcal{P}(\Lambda)$ and $m_{\boldsymbol{x}}^{N,i} = \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j} \in \mathcal{P}(\mathbb{R}^d)$, respectively.

Throughout this chapter we will assume the following assumptions:
For $\alpha' \in (0,\alpha)$, we have for any $(t,x) \in [t_0,T]\times\mathbb{R}^d$, $\mu,\mu' \in \mathcal{P}(\Lambda\times\mathbb{R}^d)$

$$\begin{aligned} &\sup_{\lambda_i\in\Lambda} \|U(t,(\lambda_1,\cdot),\mu)\|_{n+2+\alpha'} + \left\|\frac{\delta U}{\delta\mu}\Big(t,(\lambda_2,\cdot),\mu,(\lambda_3,\cdot)\Big)\right\|_{(n+2+\alpha',n+1+\alpha')} \\ &+ \left\|\frac{\delta U}{\delta\mu}\Big(t,(\lambda_4,\cdot),\mu,(\lambda_5,\cdot),(\lambda_6,\cdot)\Big)\right\|_{(n+2+\alpha',n+\alpha',n+\alpha')} \leq C_0, \end{aligned}$$

and the mapping

$$(t,\mu) \to \frac{\delta U}{\delta\mu}\Big(t,(\lambda,\cdot),\mu,(\tilde{\lambda},\cdot),(\tilde{\tilde{\lambda}},\cdot)\Big) \in \mathcal{C}^{n+2+\alpha'}(\mathbb{R}^d)\times\mathcal{C}^{n+\alpha'}(\mathbb{R}^d)\times\mathcal{C}^{n+\alpha'}(\mathbb{R}^d)$$

is continuous for every $(\lambda,\tilde{\lambda},\tilde{\tilde{\lambda}}) \in \Lambda^3$.

Also we will assume that the solution to the multi-population master equation U is $\mathcal{C}^2$ in space and satisfies the regularities in *Theorem* 3.0.1.

Proposition 5.0.1. *For any $N \geq 2$, $i \in \{1, \dots, N\}$, $u^{\boldsymbol{\lambda},N,i}$ is of class $\mathcal{C}^2$ in the space variable with:*

(i) $D_{x_j} u^{\boldsymbol{\lambda},N,i}(t, \boldsymbol{x}) = \frac{1}{N-1} D_y \frac{\delta U}{\delta \mu}\Big(t, (\lambda_i, x_i), \mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}, (\lambda_j, x_j)\Big)$ $(j \neq i)$.

(ii) $D^2_{x_i,x_j} u^{\boldsymbol{\lambda},N,i}(t, \boldsymbol{x}) = \frac{1}{N-1} D_x D_y \frac{\delta U}{\delta \mu}\Big(t, (\lambda_i, x_i), \mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}, (\lambda_j, x_j)\Big)$ $(j \neq i)$.

$$(iii)\ D^2_{x_j,x_j} u^{\boldsymbol{\lambda},N,i}(t, \boldsymbol{x}) = \frac{1}{N-1} D_y D_y \frac{\delta U}{\delta \mu}\Big(t, (\lambda_i, x_i), \mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}, (\lambda_j, x_j)\Big) + \frac{1}{(N-1)^2} D_z \frac{\delta}{\delta \mu} D_y \frac{\delta U}{\delta \mu}\Big(t, (\lambda_i, x_i), \mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}, (\lambda_j, x_j), (\lambda_j, x_j)\Big)\ (j \neq i).$$

$$(iv)\ D^2_{x_k,x_j} u^{\boldsymbol{\lambda},N,i}(t, x) = \frac{1}{(N-1)^2} D_z \frac{\delta}{\delta \mu} D_y \frac{\delta U}{\delta \mu}\Big(t, (\lambda_i, x_i), \mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}, (\lambda_j, x_j), (\lambda_k, x_k)\Big)$$

(i, j, k) *distinct,*

where D_x (D_y, D_z) is the derivative with respect to the first (second, third) coordinate $\Big(t, (\lambda, x), \mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}, (\lambda, y), (\lambda, z)\Big)$ in the variable $\mathbb{R}^d$.

Proof.

$$(i)\ \ D_{x_j} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}) = \lim_{h\to 0} \frac{u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}+hx_j\vec{e_j}) - u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x})}{h}$$

$$= \lim_{h\to 0} \frac{U(t,(\lambda_i,x_i),\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_j\vec{e_j}}) - U(t,(\lambda_i,x_i),\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}})}{h}$$

$$= \lim_{h\to 0} \frac{1}{h}\int_0^1 \int_{\Lambda\times\mathbb{R}^d} \frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i), s\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_j\vec{e_j}} + (1-s)\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}}, (\tilde{\lambda},y)\Big)$$

$$\Big[\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_j\vec{e_j}}(\tilde{\lambda},y) - \mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}}(\tilde{\lambda},y)\Big] d\tilde{\lambda}dyds$$

$$= \lim_{h\to 0} \frac{1}{h}\int_0^1 \int_{\Lambda\times\mathbb{R}^d} \frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i), s\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_j\vec{e_j}} + (1-s)\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}}, (\tilde{\lambda},y)\Big)$$

$$\frac{1}{N-1}\Big[\sum_{k\neq i} \delta_{(\lambda_k,x_k+hx_j\delta_{jk})}(\tilde{\lambda},y) - \delta_{(\lambda_k,x_k)}(\tilde{\lambda},y)\Big] d\tilde{\lambda}dyds$$

$$= \lim_{h\to 0} \frac{1}{h}\int_0^1 \int_{\Lambda\times\mathbb{R}^d} \frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i), s\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_j\vec{e_j}} + (1-s)\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}}, (\tilde{\lambda},y)\Big)$$

$$\frac{1}{N-1}\Big[\delta_{(\lambda_j,x_j+hx_j)}(\tilde{\lambda},y) - \delta_{(\lambda_j,x_j)}(\tilde{\lambda},y)\Big] d\tilde{\lambda}dyds$$

$$= \lim_{h\to 0} \frac{1}{h(N-1)}\int_0^1 \frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i), s\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_j\vec{e_j}} + (1-s)\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}}, (\lambda_j,x_j+hx_j)\Big)$$

$$- \frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i), s\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_j\vec{e_j}} + (1-s)\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}}, (\lambda_j,x_j)\Big) ds$$

$$= \frac{1}{N-1} D_y \frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}},(\lambda_j,x_j)\Big).$$

$$(ii)\ \ D^2_{x_i,x_j} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}) = D_{x_i}\Big(\frac{1}{N-1} D_y \frac{\delta U}{\delta\mu}(t,(\lambda_i,x_i),\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}},(\lambda_j,x_j))\Big)$$

$$= \frac{1}{N-1} D_x D_y \frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}},(\lambda_j,x_j)\Big).$$

$$(iii)\ \ D_{x_j}D_{x_j}u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}) = D_{x_j}\Big(\frac{1}{N-1}D_y\frac{\delta U}{\delta\mu}(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j)\Big)$$

$$= \frac{1}{N-1}\lim_{h\to 0}\frac{1}{h}\Big[D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}+hx_j\vec{e_j}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j+hx_j)\Big)$$

$$- D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j)\Big)\Big]$$

$$= \frac{1}{N-1}\lim_{h\to 0}\frac{1}{h}\Big[D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}+hx_j\vec{e_j}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j+hx_j)\Big)$$

$$- D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}+hx_j\vec{e_j}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j)\Big) + D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}+hx_j\vec{e_j}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j)\Big)$$

$$- D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j)\Big)\Big]$$

$$= \frac{1}{N-1}D_yD_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j)\Big)$$

$$+ \frac{1}{N-1}\lim_{h\to 0}\int_0^1\int_{\Lambda\times\mathbb{R}^d}\frac{\delta}{\delta\mu}D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),s\mu_{\boldsymbol{x}+hx_j\vec{e_j}}^{\boldsymbol{\lambda},N,i}+(1-s)\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j),$$

$$(\tilde{\tilde{\lambda}},z)\Big)\frac{1}{N-1}\Big[\sum_{k\neq i}\delta_{(\lambda_k,x_k+hx_j\delta_{jk})}(\tilde{\tilde{\lambda}},y)-\delta_{(\lambda_k,x_k)}(\tilde{\tilde{\lambda}},y)\Big]d\tilde{\tilde{\lambda}}dyds$$

$$= \frac{1}{N-1}D_yD_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j)\Big)$$

$$+ \frac{1}{N-1}\lim_{h\to 0}\int_0^1\int_{\Lambda\times\mathbb{R}^d}\frac{\delta}{\delta\mu}D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),s\mu_{\boldsymbol{x}+hx_j\vec{e_j}}^{\boldsymbol{\lambda},N,i}+(1-s)\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j),$$

$$(\tilde{\tilde{\lambda}},z)\Big)\Big(\frac{1}{N-1}\delta_{(\lambda_j,x_j+hx_j)}(\tilde{\tilde{\lambda}},z)-\frac{1}{N-1}\delta_{(\lambda_j,x_j)}(\tilde{\tilde{\lambda}},z)\Big)d\tilde{\tilde{\lambda}}dzds$$

$$= \frac{1}{N-1}D_yD_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j)\Big)$$

$$+ \frac{1}{(N-1)^2}\lim_{h\to 0}\frac{1}{h}\int_0^1\Big[\frac{\delta}{\delta\mu}D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),s\mu_{\boldsymbol{x}+hx_j\vec{e_j}}^{\boldsymbol{\lambda},N,i}+(1-s)\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j),$$

$$(\lambda_j,x_j+hx_j)\Big)-\frac{\delta}{\delta\mu}D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),s\mu_{\boldsymbol{x}+hx_j\vec{e_j}}^{\boldsymbol{\lambda},N,i}+(1-s)\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j),(\lambda_j,x_j)\Big)\Big]ds$$

$$= \frac{1}{N-1}D_yD_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j)\Big)$$

$$+ \frac{1}{(N-1)^2}D_z\frac{\delta}{\delta\mu}D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j),(\lambda_j,x_j)\Big).$$

(iv) $D_{x_k} D_{x_j} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}) = D_{x_k}\Big[\frac{1}{N-1} D_y \frac{\delta U}{\delta \mu}\Big(t,(\lambda_i,x_i),\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}},(\lambda_j,x_j)\Big)\Big]$

$$= \frac{1}{N-1}\lim_{h\to 0}\frac{1}{h}\Big[D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_k\vec{e_k}},(\lambda_j,x_j)\Big) - \frac{\delta U}{\delta \mu}\Big(t,(\lambda_i,x_i),\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}},(\lambda_j,x_j)\Big)\Big]$$

$$= \frac{1}{N-1}\lim_{h\to 0}\frac{1}{h}\int_0^1\int_{\Lambda\times\mathbb{R}^d}\frac{\delta}{\delta\mu}D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),s\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_k\vec{e_k}}+(1-s)\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}},(\lambda_j,x_j),$$
$$(\tilde{\tilde{\lambda}},z)\Big)\Big(\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_k\vec{e_k}}(\tilde{\tilde{\lambda}},z)-\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}}(\tilde{\tilde{\lambda}},z)\Big)d\tilde{\tilde{\lambda}}dzds$$

$$= \frac{1}{N-1}\lim_{h\to 0}\frac{1}{h}\int_0^1\int_{\Lambda\times\mathbb{R}^d}\frac{\delta}{\delta\mu}D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),s\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_k\vec{e_k}}+(1-s)\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}},(\lambda_j,x_j),$$
$$(\tilde{\tilde{\lambda}},z)\Big)\frac{1}{N-1}\Big(\sum_{h\neq i}\delta_{(\lambda_h,x_h+hx_k\delta_{hk})}(\tilde{\tilde{\lambda}},z)-\delta_{(\lambda_h,x_h)}(\tilde{\tilde{\lambda}},z)\Big)d\tilde{\tilde{\lambda}}dzds$$

$$= \frac{1}{N-1}\lim_{h\to 0}\frac{1}{h}\int_0^1\int_{\Lambda\times\mathbb{R}^d}\frac{\delta}{\delta\mu}D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),s\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_k\vec{e_k}}+(1-s)\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}},(\lambda_j,x_j),$$
$$(\tilde{\tilde{\lambda}},z)\Big)\frac{1}{N-1}\Big(\delta_{(\lambda_k,x_k+hx_k)}(\tilde{\tilde{\lambda}},z)-\delta_{(\lambda_k,x_k)}(\tilde{\tilde{\lambda}},z)\Big)d\tilde{\tilde{\lambda}}dzds$$

$$= \frac{1}{(N-1)^2}\lim_{h\to 0}\frac{1}{h}\int_0^1\Big[\frac{\delta}{\delta\mu}D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),s\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_k\vec{e_k}}+(1-s)\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}},(\lambda_j,x_j),$$
$$(\lambda_k,x_k+hx_k)\Big)-\frac{\delta}{\delta\mu}D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),s\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}+hx_k\vec{e_k}}+(1-s)\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}},(\lambda_j,x_j),(\lambda_k,x_k)\Big)\Big]ds$$

$$= \frac{1}{(N-1)^2}D_z\frac{\delta}{\delta\mu}D_y\frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}},(\lambda_j,x_j),(\lambda_k,x_k)\Big).$$

□

Proposition 5.0.2. *One has, for any* $i\in\{1,\dots,N\},\lambda_i\in\Lambda$,

$$\begin{cases} -\partial_t u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x})-\sum_{j=1}^N\triangle_{x_j}u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x})+H(x_i,D_{x_i}u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x})) \\ +\sum_{j\neq i}D_pH(x_j,D_{x_j}u^{\boldsymbol{\lambda},N,j}(t,\boldsymbol{x}))\cdot D_{x_j}u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}) \\ =F(\lambda_i,x_i,\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}})+r^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}) \ \ in \ [t_0,T]\times(\mathbb{R}^d)^N, \\ u^{\boldsymbol{\lambda},N,i}(T,\boldsymbol{x})=G(\lambda_i,x_i,\mu^{\boldsymbol{\lambda},N,i}_{\boldsymbol{x}}) \quad in \ (\mathbb{R}^d)^N, \end{cases} \tag{13}$$

where $r^{\boldsymbol{\lambda},N,i}\in\mathcal{C}^0([t_0,T]\times\mathbb{R}^{Nd})$ *with* $\sup_{\boldsymbol{\lambda}\in\Lambda^N}\|r^{\boldsymbol{\lambda},N,i}\|_\infty\leq\frac{C}{N}$.

Proof. The terminal conditions of the system follow from the terminal conditions of the multi-population master equation (9).
As U solves the multi-population master equation (9), computing at point

$(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i})$, we obtain:

$$
\begin{aligned}
&-\partial_t U(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}) - \triangle_x U(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i})\\
&+H(x_i, D_x U(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}) - F(\lambda_i,x_i,\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i})\\
&=\int_{\Lambda\times\mathbb{R}^d} \triangle_y \frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\tilde\lambda,y)\Big)\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}(\tilde\lambda,y)d\tilde\lambda dy\\
&-\int_{\Lambda\times\mathbb{R}^d} D_y \frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\tilde\lambda,y)\Big)\cdot D_pH\Big(y, DU(t,(\tilde\lambda,y),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i})\Big)\\
&\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}(\tilde\lambda,y)d\tilde\lambda dy
\end{aligned}
$$

which by definition of $\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}$ is equal to:

$$
\begin{aligned}
&-\partial_t U(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}) - \triangle_x U(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i})\\
&+H(x_i, D_x U(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}) - F(\lambda_i,x_i,\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i})\\
&=\frac{1}{N-1}\sum_{j\neq i} \triangle_y \frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda,x_j)\Big)\\
&-\frac{1}{N-1}\sum_{j\neq i} D_y \frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j)\Big)\cdot D_pH\Big(x_j, DU(t,(\lambda_j,x_j),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i})\Big)
\end{aligned}
$$

Note by previous Proposition

$$
\frac{1}{N-1} D_y \frac{\delta U}{\delta\mu}\Big(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j)\Big) = D_{x_j} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}).
$$

Therefore $\|D_{x_j} u^{\boldsymbol{\lambda},N,i}\|_\infty \leq \frac{C}{N}$, and by the Lipschitz continuity of D_xU with respect to the measure variable, we have

$$
|D_xU(t,(\lambda_j,x_j),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}) - D_xU(t,(\lambda_j,x_j),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,j})| \leq C d_1(\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,j}) \leq \frac{C}{N-1},
$$

which implies by the Lipschitz continuity of D_pH that

$$
\Big| D_pH(x_j, D_xU(t,(\lambda_j,x_j),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}) - D_pH(x_j, D_{x_j}u^{\boldsymbol{\lambda},N,j}(t,\boldsymbol{x}))\Big| \leq \frac{C}{N}.
$$

Applying the results above we obtain that

$$
\begin{aligned}
&\frac{1}{N-1}\sum_{j\neq i} D_y \frac{\delta U}{\delta\mu}(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j))\cdot D_pH(x_j, D_xU(t,(\lambda_j,x_j,\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}))\\
&=\sum_{j\neq i} D_{x_j}u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x})\cdot D_pH(x_j,\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}))\\
&=\sum_{j\neq i} D_{x_j}u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x})\cdot D_pH(x_j, D_{x_j}u^{\boldsymbol{\lambda},N,j}(t,\boldsymbol{x})) + O\Big(\frac{1}{N}\Big).
\end{aligned}
$$

On the other hand we have that

$$\sum_{j=1}^{N} \triangle_{x_j} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}) = \triangle_{x_i} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}) + \sum_{j\neq i} \triangle_{x_j} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x})$$

$$= \triangle_{x_i} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}) + \frac{1}{N-1}\sum_{j\neq i} \triangle_y \frac{\delta U}{\delta \mu}(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j))$$

$$+ \frac{1}{(N-1)^2}\sum_{j\neq i} D_z \frac{\delta}{\delta\mu} D_y \frac{\delta U}{\delta \mu}(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j),(\lambda_j,x_j))$$

$$= \triangle_{x_i} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}) + \frac{1}{N-1}\sum_{j\neq i} \triangle_y \frac{\delta U}{\delta \mu}(t,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i},(\lambda_j,x_j)) + O\Big(\frac{1}{N}\Big).$$

Therefore combining the previous results

$$-\partial_t u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}) - \sum_{j=1}^{N} \triangle_{x_j} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}) + H(x_i, D_{x_i} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x}))$$

$$+ \sum_{j\neq i} D_p H(x_j, D_{x_j} u^{\boldsymbol{\lambda},N,j}(t,\boldsymbol{x})) \cdot D_{x_j} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{x})$$

$$= F(\lambda_i, x_i, \mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}) + O\Big(\frac{1}{N}\Big).$$

□

A key idea to prove convergence is comparing the "optimal trajectories" of $v^{\boldsymbol{\lambda},N,i}$ and $u^{\boldsymbol{\lambda},N,i}$ for any $i \in \{1,\dots,N\}$. To do this, let us fix $t_0 \in [0,T)$, $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$ and let $(Z_j)_{j\in\{1,\dots,N\}}$ be an i.i.d family of N random variables of law μ_0. We set $\boldsymbol{Z} = (Z_j)_{j\in\{1,\dots,N\}}$. Let also $\Big((B_t^i)_{t\in[t_0,T]}\Big)_{i\in\{1,\dots,N\}}$ be a family of N independent d- dimensional Brownian motion that is also independent of $(Z_j)_{j\in\{1,\dots,N\}}$.
We consider the systems of stochastic differential equations (SDEs) with variables $\Big(\boldsymbol{X}_t = (X_{i,t})_{i\in\{1,\dots,N\}}\Big)_{t\in[t_0,T]}$ and $\Big(\boldsymbol{Y}_t = (Y_{i,t})_{i\in\{1,\dots,N\}}\Big)_{t\in[t_0,T]}$

$$\begin{cases} dX_{i,t} = -D_p H(X_{i,t}, D_{x_i} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{X}_t))dt + \sqrt{2}dB_t^i, \quad t\in[t_0,T], \\ X_{i,t_0} = Z_i, \end{cases} \tag{14}$$

and

$$\begin{cases} dY_{i,t} = -D_p H(Y_{i,t}, D_{x_i} v^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{Y}_t))dt + \sqrt{2}dB_t^i, \quad t\in[t_0,T], \\ Y_{i,t_0} = Z_i. \end{cases} \tag{15}$$

Theorem 5.0.3. *Assume that H, F and G satisfy the assumption of Theorem 4.0.2 with $n \geq 2$ and for $\alpha' \in (0,\alpha)$, we have for any $(t,x) \in [t_0,T]\times\mathbb{R}^d$,*

$(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) \in \Lambda^6,\ \mu, \mu' \in \mathcal{P}(\Lambda \times \mathbb{R}^d)$

$$\|U(t, (\lambda_1, \cdot), \mu)\|_{n+2+\alpha'} + \left\| \frac{\delta U}{\delta \mu}(t, (\lambda_2, \cdot), \mu, (\lambda_3, \cdot)) \right\|_{(n+2+\alpha', n+1+\alpha')}$$
$$+ \left\| \frac{\delta^2 U}{\delta \mu^2}(t, (\lambda_4, \cdot), \mu, (\lambda_5, \cdot), (\lambda_6, \cdot)) \right\|_{(n+2+\alpha', n+\alpha', n+\alpha')} \leq C_0,$$

and the mapping

$$(t, \mu) \to \frac{\delta^2 U}{\delta \mu^2}(t, (\lambda, \cdot), \mu, (\tilde{\lambda}, \cdot), (\tilde{\tilde{\lambda}}, \cdot))$$

from $[t_0, T] \times \mathcal{P}(\Lambda \times \mathbb{R}^d)$ *to* $\mathcal{C}^{n+2+\alpha'}(\mathbb{R}^d) \times [\mathcal{C}^{n+\alpha'}(\mathbb{R}^d)]^2$ *is continuous for any* $(\lambda, \tilde{\lambda}, \tilde{\tilde{\lambda}}) \in \Lambda^3$.
Then we have, for any $i \in \{1, \ldots, N\}$,

$$\mathbb{E}[\sup_{t \in [t_0, T]} |Y_{i,t} - X_{i,t}|] \leq \frac{C}{N}, \quad \forall t \in [t_0, T], \tag{16}$$

$$\mathbb{E}\Big[\int_{t_0}^T |D_{x_i} v^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t) - D_{x_i} u^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t)|^2 dt] \leq \frac{C}{N^2}, \tag{17}$$

and $\mathbb{P}$*-almost surely, for all* $i = 1, \ldots, N$,

$$|u^{\boldsymbol{\lambda}, N, i}(t_0, \boldsymbol{Z}) - v^{\boldsymbol{\lambda}, N, i}(t_0, \boldsymbol{Z})| \leq \frac{C}{N}, \tag{18}$$

where C *is a constant that doesn' t depend on* t_0, m_0 *and* N.

Proof. We start proving (17). WLOG $t_0 = 0$ and let us introduce new notations:

$$U_t^{\boldsymbol{\lambda}, N, i} = u^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t), \quad V_t^{\boldsymbol{\lambda}, N, i} = v^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t),$$
$$DU_t^{\boldsymbol{\lambda}, N, i, j} = D_{x_j} u^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t),\ DV_t^{\boldsymbol{\lambda}, N, i, j} = D_{x_j} v^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t),\ \ t \in [0, T].$$

Applying (12) on $(v^{\boldsymbol{\lambda}, N, i})_{i \in \{1, \ldots, N\}}$ and Itô's formula we deduce that

$$\begin{aligned}
dV_t^{N,i} &= \Big[\partial_t v^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t) - \sum_{j=1}^N D_{x_j} v^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t) \cdot D_p H(Y_{j,t}, D_{x_j} v^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t)) \\
&\quad + \sum_{j=1}^N \triangle_{x_j} v^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t)\Big] dt \\
&\quad + \sqrt{2} \sum_{j=1}^N D_{x_j} v^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t) \cdot dB_t^j \\
&= \Big[H(Y_{i,t}, D_{x_i} v^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t)) - D_{x_i} v^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t) \cdot D_p H(Y_{i,t}, D_{x_i} v^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t)) \\
&\quad - F(\lambda_i, Y_{i,t}, \mu_{\boldsymbol{Y}_t}^{\boldsymbol{\lambda}, N, i})\Big] dt + \sqrt{2} \sum_{j=1}^N D_{x_j} v^{\boldsymbol{\lambda}, N, i}(t, \boldsymbol{Y}_t) \cdot dB_t^j.
\end{aligned}$$

In a similar way, since $(u^{\boldsymbol{\lambda},N,i})_{i\in\{1,\dots,N\}}$ satisfies Proposition 5.0.2, we obtain that:

$$\begin{aligned}
dU_t^{\boldsymbol{\lambda},N,i} =& \Big[H(Y_{i,t}, D_{x_i} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{Y}_t)) \\
& \cdot D_p H(Y_{i,t}, D_{x_i} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{Y}_t)) - F(\lambda_i, Y_{i,t}, \mu_{\boldsymbol{Y}_t}^{\boldsymbol{\lambda},N,i}) - r^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{Y}_t)\Big] dt \\
& - \sum_{j=1}^N D_{x_j} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{Y}_t) \cdot \Big(D_p H(Y_{j,t}, D_{x_j} v^{\boldsymbol{\lambda},N,j}(t,\boldsymbol{Y}_t)) \\
& - D_p H(Y_{j,t}, D_{x_j} u^{\boldsymbol{\lambda},N,j}(t,\boldsymbol{Y}_t))\Big) dt + \sqrt{2} \sum_{j=1}^N D_{x_j} u^{\boldsymbol{\lambda},N,i}(t,\boldsymbol{Y}_t) \cdot dB_t^j.
\end{aligned}$$

Computing the difference between the previous two exspressions, taking the square and applying Itô's formula we compute that:

$$
\begin{aligned}
d[U_t^{\boldsymbol{\lambda},N,i} - V_t^{\boldsymbol{\lambda},N,i}]^2 = \Big[& 2(U_t^{\boldsymbol{\lambda},N,i} - V_t^{\boldsymbol{\lambda},N,i})\big(H(Y_{i,t}, DU_t^{\boldsymbol{\lambda},N,i,i}) \\
& - H(Y_{i,t}, DV_t^{\boldsymbol{\lambda},N,i,i})\big) - 2(U_t^{\boldsymbol{\lambda},N,i} - V_t^{\boldsymbol{\lambda},N,i})\big(DU_t^{\boldsymbol{\lambda},N,i,i} \cdot [D_pH(Y_{i,t}, DU_t^{\boldsymbol{\lambda},N,i,i}) \\
& - D_pH(Y_{i,t}, DV_t^{\boldsymbol{\lambda},N,i,i})]\big) - 2(U_t^{\boldsymbol{\lambda},N,i} - V_t^{\boldsymbol{\lambda},N,i})\big([DU_t^{\boldsymbol{\lambda},N,i,i} - DV_t^{\boldsymbol{\lambda},N,i,i}] \\
& \cdot D_pH(Y_{i,t}, DV_t^{N,i,i})\big) - 2(U_t^{\boldsymbol{\lambda},N,i} - V_t^{\boldsymbol{\lambda},N,i}) r^{\boldsymbol{\lambda},N,i}(t, \boldsymbol{Y}_t)\Big] dt \\
& - 2(U_t^{\boldsymbol{\lambda},N,i} - V_t^{\boldsymbol{\lambda},N,i}) \sum_{j=1}^N DU_t^{\boldsymbol{\lambda},N,i,j} \cdot \big(D_pH(Y_{j,t}, DV_t^{\boldsymbol{\lambda},N,j,j}) \\
& - D_pH(Y_{j,t}, DU_t^{\boldsymbol{\lambda},N,j,j})\big) dt + 2\sum_{j=1}^N |DU_t^{\boldsymbol{\lambda},N,i,j} - DV_t^{\boldsymbol{\lambda},N,i,j}|^2 \\
& + \sqrt{2}(U_t^{\boldsymbol{\lambda},N,i} - V_t^{\boldsymbol{\lambda},N,i}) + \sum_{j=1}^N (DU_t^{\boldsymbol{\lambda},N,i,j} - DV_t^{\boldsymbol{\lambda},N,i,j}) \cdot dB_t^j.
\end{aligned}
$$

Note that H and D_pH are Lipschitz continuous in the variable p.
We also recall that $DU_t^{\boldsymbol{\lambda},N,i,i} = D_xU(t, Y_{i,t}, \mu_{\boldsymbol{Y}_t}^{\boldsymbol{\lambda},N,i})$ is bounded independently of i, N and t and that $DU_t^{\boldsymbol{\lambda},N,i,j}$ is bounded by $\frac{C}{N}$ (due to Proposition 5.0.1) when $i \neq j$ for C independent of i, j, N and t.
By Proposition 5.0.2 $r^{\boldsymbol{\lambda},N,i}$ is bounded by $\frac{C}{N}$, therefore integrating from t to T the above formula and taking the conditional expectation given $\boldsymbol{Z}$ we obtain that:

$$
\begin{aligned}
& \mathbb{E}\Big[|U_t^{\boldsymbol{\lambda},N,i} - V_t^{\boldsymbol{\lambda},N,i}|^2 \,\Big|\, \boldsymbol{Z}\Big] + 2\sum_{j=1}^N \mathbb{E}\Big[\int_t^T |DU_s^{\boldsymbol{\lambda},N,i,j} - DV_s^{\boldsymbol{\lambda},N,i,j}|^2 ds \,\Big|\, \boldsymbol{Z}\Big] \\
& \leq \mathbb{E}\Big[|U_T^{\boldsymbol{\lambda},N,i} - V_T^{\boldsymbol{\lambda},N,i}|^2 |\boldsymbol{Z}\Big] + \frac{C}{N}\int_t^T \mathbb{E}\Big[|U_s^{\boldsymbol{\lambda},N,i} - V_s^{\boldsymbol{\lambda},N,i}| \,\big|\, \boldsymbol{Z}\Big] ds \\
& + C\int_t^T \mathbb{E}\Big[|U_s^{\boldsymbol{\lambda},N,i} - V_s^{\boldsymbol{\lambda},N,i}| \cdot |DU_s^{\boldsymbol{\lambda},N,i,i} - DV_s^{\boldsymbol{\lambda},N,i,i}| \,\Big|\boldsymbol{Z}\Big] ds \\
& + \frac{C}{N}\sum_{j\neq i}\int_t^T \mathbb{E}\Big[|U_s^{\boldsymbol{\lambda},N,i} - V_s^{\boldsymbol{\lambda},N,i}| \cdot |DU_s^{\boldsymbol{\lambda},N,j,j} - DV_s^{\boldsymbol{\lambda},N,j,j}| \,\Big|\boldsymbol{Z}\Big] ds.
\end{aligned}
$$

Note that the boundary condition of $U_T^{\boldsymbol{\lambda},N,i} - V_T^{\boldsymbol{\lambda},N,i}$ is 0 and by a convexity argument we obtain that

$$\begin{aligned}
&\mathbb{E}\Big[|U_t^{\boldsymbol{\lambda},N,i} - V_t^{\boldsymbol{\lambda},N,i}|^2 \,\big|\boldsymbol{Z}\Big]+\mathbb{E}\Big[\int_t^T |DU_s^{\boldsymbol{\lambda},N,i,i} - DV_s^{\boldsymbol{\lambda},N,i,i}|^2 ds \,\big|\, \boldsymbol{Z}\Big] \\
&\quad\leq \frac{C}{N^2} + C\int_t^T \mathbb{E}\Big[|U_s^{\boldsymbol{\lambda},N,i} - V_s^{\boldsymbol{\lambda},N,i}| \,\big|\, \boldsymbol{Z}\Big] ds \\
&\quad+ \frac{1}{2N}\sum_{n=1}^N \mathbb{E}\Big[\int_t^T |DU_s^{\boldsymbol{\lambda},N,i} - DV_s^{\boldsymbol{\lambda},N,j,j}|^2 ds \big|\, \boldsymbol{Z}\Big].
\end{aligned} \tag{19}$$

Taking the mean of the expression above over $i \in \{1, \dots, N\}$ we get

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N \mathbb{E}\Big[|U_t^{\boldsymbol{\lambda},N,i} - V_t^{\boldsymbol{\lambda},N,i}|^2\big|\boldsymbol{Z}\Big] \\
&\leq \frac{C}{N^2} + \int_t^T \frac{C}{N}\sum_{i=1}^N \mathbb{E}\Big[|U_s^{\boldsymbol{\lambda},N,i} - V_s^{\boldsymbol{\lambda},N,i}|^2\big|\boldsymbol{Z}\Big] \\
&- \frac{1}{2N}\sum_{n=1}^N \mathbb{E}\Big[\int_t^T |DU_s^{\boldsymbol{\lambda},N,i} - DV_s^{\boldsymbol{\lambda},N,j,j}|^2 ds \big|\, \boldsymbol{Z}\Big] \\
&\leq \frac{C}{N^2} + \int_t^T \frac{C}{N}\sum_{i=1}^N \mathbb{E}\Big[|U_s^{\boldsymbol{\lambda},N,i} - V_s^{\boldsymbol{\lambda},N,i}|^2\big|\boldsymbol{Z}\Big],
\end{aligned}$$

which by Gronwall's lemma implies that there exists a costant $\tilde{C}$ such that

$$\sup_{0\leq t\leq T}\left[\frac{1}{N}\sum_{i=1}^N \mathbb{E}\Big[|U_t^{\boldsymbol{\lambda},N,i} - V_t^{\boldsymbol{\lambda},N,i}|^2\big|\boldsymbol{Z}\Big]\right] \leq \frac{\tilde{C}}{N^2}. \tag{20}$$

Plugging (20) into (19) we deduce that there exists a constant $\tilde{\tilde{C}}$ for which

$$\frac{1}{N}\sum_{j=1}^N \mathbb{E}\Big[\int_0^T |DU_s^{\boldsymbol{\lambda},N,i} - DV_s^{\boldsymbol{\lambda},N,j,j}|^2 ds \big|\, \boldsymbol{Z}\Big] \leq \frac{\tilde{\tilde{C}}}{N^2}.$$

Inserting this bound on the RHS of (19) and applying Gronwall's lemma again we deduce that there exists a constant $\tilde{\tilde{\tilde{C}}}$ for which

$$\sup_{t\in[0,T]}\mathbb{E}\Big[|U_t^{\boldsymbol{\lambda},N,i} - V_t^{\boldsymbol{\lambda},N,i}|^2\big|\boldsymbol{Z}\Big]\Big]+\mathbb{E}\Big[\int_0^T |DU_s^{\boldsymbol{\lambda},N,i} - DV_s^{\boldsymbol{\lambda},N,j,j}|^2 ds \big|\, \boldsymbol{Z}\Big] \leq \frac{\tilde{\tilde{\tilde{C}}}}{N^2}, \tag{21}$$

which proves (17).
Now we will prove (16) and (18). We start with (18).Noticing that $U_0^{\boldsymbol{\lambda},N,i}$ −

$V_0^{\boldsymbol{\lambda},N,i} = u^{\boldsymbol{\lambda},N,i}(0,\boldsymbol{Z}) - v^{\boldsymbol{\lambda},N,i}(0,\boldsymbol{Z})$, we deduce from (21) that with probability 1 for all $i \in \{1,\dots,N\}$,

$$|u^{\boldsymbol{\lambda},N,i}(0,\boldsymbol{Z}) - v^{\boldsymbol{\lambda},N,i}(0,\boldsymbol{Z})| \leq \frac{\tilde{\tilde{C}}}{N}$$

which concludes the proof of (18).
To prove (16) we estimate the difference $X_{i,t} - Y_{i,t}$, for $t \in [0,T]$ and $i \in \{1,\dots,N\}$. Due to the equation satisfied by the processes $(X_{i,t})_{t\in[0,T]}$ and $(Y_{i,t})_{t\in[0,T]}$, we have

$$\begin{aligned}|X_{i,t} - Y_{i,t}| \leq \int_0^t &|D_pH(X_{i,s}, D_{x_i}u^{\boldsymbol{\lambda},N,i}(s,\boldsymbol{X}_s))\\ &- D_pH(Y_{i,s}, D_{x_i}v^{\boldsymbol{\lambda}}(s,\boldsymbol{Y}_s))|ds\end{aligned}$$

Using the Lipschitz regularity of D_pH, the regularity of U and Proposition 5.0.1, we obtain

$$\begin{aligned}|X_{i,t} - Y_{i,t}| \leq &C\int_0^t \Big(\big|X_{i,s} - Y_{i,s}\big| + \frac{1}{N}\sum_{j\neq i}\big|X_{j,s} - Y_{j,s}\big|\Big)ds\\ &+ \int_0^t \big|D_pH(Y_{i,s}, D_{x_i}u^{\boldsymbol{\lambda},N,i}(s,\boldsymbol{Y}_s)\\ &- D_pH(Y_{i,s}, D_{x_i}v^{\boldsymbol{\lambda},N,i}(s,\boldsymbol{Y}_s)\big|ds.\end{aligned}$$

Taking the sup over $t \in [0,\tau]$ (for $\tau \in [0,T]$) and the conditional expectation with respect to $\boldsymbol{Z}$ we find out that

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in[0,\tau]}|X_{i,t} - Y_{i,t}|\big|\boldsymbol{Z}\Big] \leq &C\int_0^\tau \Bigg(\mathbb{E}\Big[\sup_{t\in[0,s]}\big|X_{i,t} - Y_{i,t}\big|\big|\boldsymbol{Z}\Big]\\ &+ \frac{1}{N}\sum_{j\neq i}\mathbb{E}\Big[\sup_{t\in[0,s]}\big|X_{j,t} - Y_{j,t}\big|\big|\boldsymbol{Z}\Big]\Bigg)ds \qquad (22)\\ &+ \mathbb{E}\Bigg[\int_0^T |DU_s^{\boldsymbol{\lambda},N,i,i} - DV_s^{\boldsymbol{\lambda},N,i,i}|ds\big|\boldsymbol{Z}\Bigg].\end{aligned}$$

Summing over $i \in \{1,\dots,N\}$ and using (21), we obtain by Gronwall' s inequality that there exists a constant K for which:

$$\sum_{j=1}^N \mathbb{E}\Big[\sup_{t\in[0,T]}|X_{i,t} - Y_{i,t}|\big|\boldsymbol{Z}\Big] \leq K$$

which concludes the proof of (16). □

Theorem 5.0.4. *Let the assumption of Theorem* 4.0.2 *of the existence and uniqueness of the multi-population master equation stand for* $n \geq 2$ *and let* $v^{\boldsymbol{\lambda},N,i}$

be the solution to the multi-population Nash system (12) *and* U *be the classical solution to the multi-population master equation. Fix* $N \geq 1$ *and* $(t_0, m_0) \in [0,T] \times \mathcal{P}(\mathbb{R}^d)$.
For any $\boldsymbol{x} \in (\mathbb{R}^d)^N$ *and* $\boldsymbol{\lambda} \in (\Lambda)^N$, *let* $\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N} := \frac{1}{N}\sum_{j=1}^N \delta_{(\lambda_j,x_j)}$. *Then*

$$\sup_{i=1,\dots,N} \left| v^{\boldsymbol{\lambda},N,i}(t_0,\boldsymbol{x}) - U(t_0,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N}) \right| \leq \frac{C}{N}.$$

Proof. Applying (18) we obtain that

$$\left| U(t_0,(\lambda_i,Z_i),\mu_{\boldsymbol{Z}}^{\boldsymbol{\lambda},N,i}) - v^{\boldsymbol{\lambda},N,i}(t_0,\boldsymbol{Z}) \right| \leq \frac{C}{N} \quad \text{a.e.,} \quad i \in \{1,\dots,N\},$$

where $\boldsymbol{Z} = (Z_1,\dots,Z_N)$ with $Z_1,\dots,Z_N$ i.i.d. random variables with uniform density on $\mathbb{R}^d$. The support of $\boldsymbol{Z}$ being $(\mathbb{R}^d)^N$, we derive from the continuity of U and of the $(v^{\boldsymbol{\lambda},N,i})_{i\in\{1,\dots,N\}}$ that the above inequality holds for any $\boldsymbol{x} \in (\mathbb{R}^d)^N$:

$$\left| U(t_0,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i}) - v^{\boldsymbol{\lambda},N,i}(t_0,\boldsymbol{x}) \right| \leq \frac{C}{N} \quad \forall x \in (\mathbb{R}^d)^N, \;\; i \in \{1,\dots,N\}.$$

Then we use the Lipschitz continuity of U with respect to μ to replace $U(t_0,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N,i})$ by $U(t_0,(\lambda_i,x_i),\mu_{\boldsymbol{x}}^{\boldsymbol{\lambda},N})$ in the above inequality with additional error a term of order $\frac{1}{N}$. □

References

[1] Martino Bardi, Marco Cirant. *Uniqueness of solutions in Mean Field Games with Several Populations and Neumann Conditions*. Springer Indam Series, 2019.

[2] P.Cardaliaguet, F.Delarue, J.-M. Lasry, P.-L.Lions. *The master equation and the convergence problems in mean field games*. Princenton and Oxford, 2017.

[3] Kolokoltsov, V. N., Li, J., and Yang, W.*Mean field games and nonlinear Markov processes*.Preprint arXiv:1112.3744, 2011.

[4] Bensoussan, A., Frehse J., and Yam.*The master equation in mean field theory*. J. Math. Pures Appl, 2015.

[5] René Carmona, François Delaruè.*Probabilistic Theory of Mean Field Games with application II, Mean field games with common noise and Master Equation*. Springer, 2018.

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

[7] Bensoussan, A., and Frehse, J.*Mean field games and mean field type control theory*. Springer, 2013.

[8] Bensoussan, A., and Frehse, J.*On the interpretation of the master equation*. Stoc. Proc. App., 127, 2093–2137, 2017.

[9] Bensoussan, A., and Frehse, J.*The master equation in mean field theory*. J. Math. Pures Appl., 103, 1441–1474, 2015.

[10] Achdou, Y., Buera, F. J., Lasry, J. M., Lions, P. L., and Moll, B.*Partial differential equation models in macroeconomics*.Philos. Trans. R. Soc. A: Math. Phys. Engin. Sci., 2014.