Convergence of some Mean Field Games systems to aggregation and flocking models

Martino Bardi* and Pierre Cardaliaguet†

November 4, 2020

Abstract

For two classes of Mean Field Game systems we study the convergence of solutions as the interest rate in the cost functional becomes very large, modeling agents caring only about a very short time-horizon, and the cost of the control becomes very cheap. The limit in both cases is a single first order integro-partial differential equation for the evolution of the mass density. The first model is a 2nd order MFG system with vanishing viscosity, and the limit is an Aggregation Equation. The result has an interpretation for models of collective animal behaviour and of crowd dynamics. The second class of problems are 1st order MFGs of acceleration and the limit is the kinetic equation associated to the Cucker-Smale model. The first problem is analyzed by PDE methods, whereas the second is studied by variational methods in the space of probability measures on trajectories.

Keywords: Mean Field Games, agent based models, aggregation equation, kinetic equations, swarming, flocking, crowd motion, weighted energy-dissipation.

Mathematics Subject Classification: 35Q89, 35Q70, 45K05, 92DXX.

Contents

1 Introduction **2**

2 Convergence for classical MFG systems **5**
- 2.1 The convergence results 6
- 2.2 Proof of Theorem 2.1 7
- 2.3 Examples 9
  - 2.3.1 The Aggregation Equation 10
  - 2.3.2 Models of crowd dynamics 11
  - 2.3.3 On uniqueness of solutions 11

3 Convergence for some MFGs of acceleration towards the Cucker-Smale model **11**
- 3.1 The convergence result 12
- 3.2 Proof of the convergence result 13

*Department of Mathematics "T. Levi-Civita", University of Padova, Via Trieste 63, 35121 Padova, Italy - bardi@math.unipd.it

†Université Paris-Dauphine, PSL Research University, Ceremade, Place du Maréchal de Lattre de Tassigny, 75775 Paris cedex 16 - France - cardaliaguet@ceremade.dauphine.fr

1 Introduction

The aim of this work is to show a rigorous connection between two different mathematical theories modelling the dynamics of large populations of individuals, human or animal, that behave with some degree of rationality. The two theories are Mean Field Games (MFG), introduced by Lasry and Lions [28] and Huang, Caines, and Malhame [27], and agent based models and kinetic equations, which have a large literature, see for instance [7, 9, 10, 19, 23, 25, 34], the survey [18], and the references therein. Both describe the population in a continuous way via its distribution in space. In MFG each agent anticipates the expected future behaviour of the population and solves a stochastic control problem to minimize his cost at a given terminal time. In agent-based models, instead, each individual reacts to the current distribution of the others according to a prescribed rule. Therefore his behaviour is much less rational than in MFG. We consider two classes of problems where the control acts, respectively, on the velocity and on the acceleration. For both the parameter measuring the rationality is the interest rate in the cost functional of a MFG: as it increase the agents become less interested in the future and therefore more myopic and less rational. In the limit the system of PDEs of the MFG degenerates into a single nonlocal equation of the type arising in individual based models.

Motivations. Our results are inspired on one hand by the last part of [8], in which the authors show how to derive a McKean-Vlasov equation from a mean field game system and, on the other hand, by [24] (see also [4, 5]) which discusses how multi-agent control problems where the players have limited anticipation of the future converge to aggregation models. Let us briefly recall the content of both papers. In [8], the authors study MFG systems of the form

$$\begin{cases} -\partial_t u_\lambda - \nu\Delta u_\lambda + H(x, Du_\lambda, m_\lambda(t)) + \lambda u = 0 \text{ in } \mathbb{R}^d \times (0,T), \\ \partial_t m_\lambda - \nu\Delta m_\lambda - \operatorname{div}(m_\lambda D_p H(x, Du_\lambda, m_\lambda(t)) = 0 \text{ in } \mathbb{R}^d \times (0,T), \\ u_\lambda(T,x) = u_T(x),\ m_\lambda(0) = m_0 \text{ in } \mathbb{R}^d. \end{cases} \tag{1}$$

Here $\nu > 0$ is fixed, $\lambda > 0$ is a large parameter which describes the impatience of the players and $H = H(x,p,m)$ is the Hamiltonian of the problem which includes interaction terms among the players. Under suitable assumptions on the data, [8] states that, as λ tends to ∞ and up to subsequences, $u_\lambda, Du_\lambda \to 0$ and m_λ converges to a solution of the McKean-Vlasov equation

$$\begin{cases} \partial_t m - \nu\Delta m - \operatorname{div}(m D_p H(x, 0, m(t)) = 0 \qquad \text{in } \mathbb{R}^d \times (0,T), \\ m(0) = m_0 \qquad \text{in } \mathbb{R}^d. \end{cases} \tag{2}$$

Possible variants and extensions (to MFG models with relative running costs and to higher order approximation) are also discussed in [8].

Although [24] shares some common features with [8], it is quite different. It proposes a deterministic model where the agents have little rationality, as in [8], in the sense that they anticipate the behaviour of the other players only on a short horizon (here through time discretization). On the other hand, and this is in contrast with [8], the agents are supposed to pay little for their move. The paper [24] explains, at least at a heuristic level, that the optimal feedback control of each agent should converge to the gradient descent of the running cost, which the authors call "Best Reply Strategy". They also discuss the limit of the distribution of agents as their number goes to infinity and the related 1st order McKean-Vlasov equation.

In the present paper we consider a continuous time variant of the model in [24] which contains its two main features: the fact that the players minimize a cost on a very short horizon, that we model as in [8] by a large discount factor, and the fact that they pay little for their moves. To fit also better with aggregation or kinetic models, we work with problems with a vanishing viscosity ($\nu = \nu_\lambda \to 0^+$ as $\lambda \to +\infty$) and in infinite horizon. In particular, our result makes rigorous the approach of [24].

MFG with control on the velocity. We prove two convergence results. In the first one, our model (in its simplest version) takes the form of the parabolic system

$$\begin{cases} -\partial_t u_\lambda - \nu_\lambda \Delta u_\lambda + \lambda u_\lambda + \dfrac{\lambda}{2}|Du_\lambda|^2 = F(x, m_\lambda(t)) & \text{in } \mathbb{R}^d \times (0, +\infty) \\ \partial_t m_\lambda - \nu_\lambda \Delta m_\lambda - \operatorname{div}(m_\lambda \lambda D u_\lambda) = 0 & \text{in } \mathbb{R}^d \times (0, +\infty) \\ m_\lambda(0) = m_0, \quad \text{in } \mathbb{R}^d, \quad u_\lambda \text{ bounded}, \end{cases} \tag{3}$$

associated to a stochastic MFG. Under some natural assumptions on F (typically, continuous on $\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$ and uniformly Lipschitz continuous and semi-concave in the space variable), we show that, as λ tends to infinity (meaning that players become more and more myopic in time and that their control is increasingly cheap), $\nu_\lambda \to 0^+$, and along subsequences, m_λ converges to a solution m of the aggregation model

$$\begin{cases} \partial_t m - \operatorname{div}(m D_x F(x, m)) = 0 & \text{in } \mathbb{R}^d \times (0, +\infty) \\ m(0) = m_0, \quad \text{in } \mathbb{R}^d. \end{cases} \tag{4}$$

Moreover, the optimal feedback $-\lambda D u_\lambda$ for the generic agent in the MFG (3) converges a.e. to the vector field $-D_x F(\cdot, m)$, which is the gradient descent of the running cost corresponding to the limit distribution of agents m. Note that this limit is not obvious to guess, whereas in the problem of [8] one expects that (2) is the limit of (1).

The limit equation in (4) covers most examples of the so-called Aggregation Equation

$$\partial_t m + \operatorname{div}\left(m \int_{\mathbb{R}^d} K(x-y) m(y)\, dy \right) = 0,$$

because the kernel of the convolution is usually the gradient of a potential, $K = -Dk$. This equation describes the collective behaviour of various animal populations, its derivation and the choice of the kernel are based on phenomenological considerations, see, e.g., [34, 7] and the references therein. In Subsection 2.3 we show that the examples of Aggregation Equation most studied in the mathematical biology literature fit the assumptions of our convergence theorem, as well as some known models of crowd dynamics taken from [21, 22]. Therefore our result gives a further justification of such models within the framework of dynamic games with a large number of players.

MFG with control on the acceleration. Our second result concerns deterministic MFG with control on the acceleration [1, 12], whose first order PDE system is of the form

$$\begin{cases} -\partial_t u_\lambda + \lambda u_\lambda - v \cdot D_x u_\lambda + \dfrac{\lambda}{2}|D_v u_\lambda|^2 = F(x, v, m_\lambda(t)) & \text{in } \mathbb{R}^{2d} \times (0, T) \\ \partial_t m_\lambda + v \cdot D_x m_\lambda - \operatorname{div}_v(m_\lambda \lambda D_v u_\lambda) = 0 & \text{in } \mathbb{R}^{2d} \times (0, T) \\ m_\lambda(0) = m_0, \quad u_\lambda(x, v, T) = 0 & \text{in } \mathbb{R}^{2d}, \end{cases} \tag{5}$$

where the unknowns u_λ and m_λ depend on the position x and the velocity v of the generic agent, and on the time t. In this case we prove the convergence to kinetic equations of the form

$$\begin{cases} \partial_t m + v \cdot D_x m - \operatorname{div}_v(m D_v F(x, v, m)) = 0 & \text{in } \mathbb{R}^{2d} \times (0, T), \\ m(0) = m_0, \quad \text{in } \mathbb{R}^{2d}, \end{cases} \tag{6}$$

as $\lambda \to +\infty$. To fix the ideas we work in the case where the coupling term F corresponds to the celebrated Cucker-Smale model [23, 25]:

$$F(x, v, m(t)) = k * m(x, v, t) = \int_{\mathbb{R}^{2d}} k(x-y, v-v_*) m(y, v_*, t) dy dv_*,$$

where

$$k(x, v) = \frac{|v|^2}{(\alpha + |x|^2)^\beta}, \quad \alpha > 0, \beta \geqslant 0.$$

Note that, in contrast with the first result, the coupling function F is no longer globally Lipschitz continuous: as we explain below, this is a source of major difficulties and it obliges us to change completely the methods of analysis. Other models of swarming and flocking can be treated by these methods.

Methods of the proofs. Let us briefly explain the mechanism of proofs and the differences with the existing literature. In [8], the rough idea is that u_λ converges to 0 and therefore Du_λ converges to 0 as well. In addition, the fact that the diffusion is nondegenerate ($\nu > 0$) provides $C^{2+\alpha,1+\alpha/2}$ bounds on u_λ and m_λ, thanks to which one can pass to the limit.

For our first result, (Theorem 2.1, on the convergence of (3) to (4)), we have to use a different argument. The key idea is that λu_λ behaves like $F(x, m_\lambda)$, because $\lambda^{-1}F(x, m_\lambda)$ is almost a solution to (3). Therefore λDu_λ is close to $DF(x, m_\lambda)$, which explains the limit equation (4). Compared to [8], an additional difficulty comes from the lack of (uniform in λ) smoothness of the solutions, since we have no diffusion term in the limit equation. In particular, the product $m_\lambda \lambda Du_\lambda$ has to be handled with care, since m_λ could degenerate as a measure while Du_λ could become singular. We overcome this issue by proving a uniform semi-concavity of λu_λ, which provides at the same time the L^1_{loc} convergence of λDu_λ and, thanks to an argument going back to [28] (see also [16]) a (locally in time) uniform L^∞ bound on the density of m_λ, and hence a weak-* convergence of m_λ.

For the second result (Theorem 3.2, on the convergence of (5) to (6)), the fact that the coupling function F grows in a quadratic way with respect to the (moment of) the measure prevents us from using fixed point techniques (as in [1, 12]) to show the existence of a solution to the MFG system (5) and to obtain estimates on the solution (this would also be the case in the presence of a viscous term). This obliges us to give up the PDE approach of the previous set-up and to use variational techniques, first suggested for MFG problems in [28] and developed by several authors since then: see, for instance, [6, 14, 15, 17, 30] and the references therein. For that very same reason, we have to work with a finite horizon problem and with initial measure having a compact support. In contrast with the first result, we do not prove the convergence of *all* the solutions of the MFG system, but only for the ones which minimize the energy written formally as

$$\int_0^T e^{-\lambda t} \int_{\mathbb{R}^{2d}} (\frac{1}{2\lambda}|\alpha(x,v,t)|^2 + \int_{\mathbb{R}^{2d}} k(x-x_*, v-v_*)m(dx_*, dv_*, t))m(dx, dv, t)dt \tag{7}$$

where $\partial_t m + v \cdot D_x m + \mathrm{div}_v(m\alpha) = 0$. We formulate this problem in the space of probability measures on curves $(\gamma, \dot\gamma)$, and the main technique of proof consists in obtaining estimates on the solution based on the dynamic programming principle in such space. This is reminiscent of ideas developed in [33] that we discuss below. Such an approach naturally involves weak solution of the MFG system and does not require the initial measure m_0 to be absolutely continuous. In this case the natural notion of solution for the limit equation (30) is the measure-valued solution developed in [10] for (6).

We could also have developed this second approach for the first type of results (i.e., the convergence of (3) to (4)), assuming that the coupling function F derives from an energy (the so-called potential mean field games), i.e.,

$$F(x,m) = \frac{\delta F}{\delta m}(m,x)$$

(see [2] for the notion of derivative). Then it is known [28] that minimizers $(m_\lambda, \alpha_\lambda)$ of the problem

$$\inf\left\{\int_0^{+\infty} e^{-\lambda t}\left(\int_{\mathbb{R}^d} \frac{1}{2}|\alpha|^2 dx + \lambda\mathcal{F}(m(t))\right)dt, \quad \partial_t m + \mathrm{div}(m\alpha) = 0,\ m(0) = m_0\right\}, \tag{8}$$

are solutions to the MFG system (3) (with $\nu_\lambda = 0$ and if $\mathcal{F}$ is smooth enough) in the sense that there exists u_λ such that (u_λ, m_λ) solves (3) and $\alpha_\lambda = -\lambda Du_\lambda$. The convergence of minimizers, as $\lambda \to +\infty$, is studied in the nice paper [33], where this convergence is called "Weighted Energy-Dissipation": the authors prove that, under suitable assumptions on the function $\mathcal{F}$ (which allow for singular coupling functions), minimizers converge to a solution of the gradient flow associated to $\mathcal{F}$, i.e., at least at a formal level, to a solution of (4). Let us note that, in contrast with our setting, the solution of the limit equation can be singular and that [33] works in general metric spaces. It would be interesting to understand the precise interpretation of the results of [33] in terms of limits of MFGs, but this exceeds the scope of the present paper. Note however that our second result (i.e., the convergence of (5) to (6)) does not fit in the framework of [10]. Indeed, the key idea of [10] is that m_λ is a gradient flow for the value function associated with Problem (8); as this value function converges to $\mathcal{F}$, (m_λ) has to converge to the gradient flow for $\mathcal{F}$, which is precisely m; this gradient flow structure is completely lost in our framework of MFG of acceleration (5): we have therefore to design a different approach.

Notation

We denote by $\mathcal{P}(\mathbb{R}^d)$ the set of Borel probability measures on $\mathbb{R}^d$, and for $p \geqslant 1$ $\mathcal{P}_p(\mathbb{R}^d)$ (or, in short $\mathcal{P}_p$) is the subset of measures with finite $p-$order moment M_p:

$$M_p(m) := \int_{\mathbb{R}^d} |x|^p m(dx).$$

The sets $\mathcal{P}_p(\mathbb{R}^d)$ are endowed with the corresponding Wasserstein distance. Given a positive constant κ, we denote by $\mathcal{M}_{p,\kappa}(\mathbb{R}^d)$ the set of measures $m \in \mathcal{P}_p(\mathbb{R}^d)$ absolutely continuous with respect to the Lebesgue measure and with a density bounded by κ. We set $\mathcal{M}_p(\mathbb{R}^d) := \bigcup_{\kappa>0} \mathcal{M}_{p,\kappa}(\mathbb{R}^d)$. In Section 3 we will also use, for $m \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$,

$$M_{2,v}(m) := \int_{\mathbb{R}^{2d}} |v|^2 m(dx, dv).$$

For functions depending only on x and t we will denote with $Du(x,t) = D_x u(x,t)$ the gradient with respect to the space variable x, and with $\partial_t u(x,t)$ the partial derivative with respect to time.

Acknowledgment

The first-named author is member of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM); he was partially supported by the research project "Nonlinear Partial Differential Equations: Asymptotic Problems and Mean-Field Games" of the Fondazione CaRiPaRo. The second author was partially supported by the ANR (Agence Nationale de la Recherche) project ANR-12-BS01-0008-01, by the CNRS through the PRC grant 1611 and by the Air Force Office for Scientific Research grant FA9550-18-1-0494. This work started during the visit of the second author at Padova University: the University is warmly thanked for this hospitality.

2 Convergence for classical MFG systems

In this section we consider MFG systems of the form

$$\begin{cases} -\partial_t u_\lambda - \nu_\lambda \Delta u_\lambda + \lambda u_\lambda + \lambda^{-1} H(\lambda D u_\lambda, x) = F(x, m_\lambda(t)) \qquad \text{in } \mathbb{R}^d \times (0, +\infty), \\ \partial_t m_\lambda - \nu_\lambda \Delta m_\lambda - \operatorname{div}(m_\lambda D_p H(\lambda D u_\lambda, x)) = 0 \qquad \text{in } \mathbb{R}^d \times (0, +\infty), \\ m_\lambda(0) = m_0, \qquad \text{in } \mathbb{R}^d, \end{cases} \tag{9}$$

where $\lambda > 0$, $\nu_\lambda > 0$ and $\nu_\lambda \to 0$ as $\lambda \to +\infty$. By a solution of (9) we mean a pair (u_λ, m_λ) solving the PDEs in the classical sense, where $m_\lambda(t) \in \mathcal{P}_2(\mathbb{R}^d)$ for all t, it has bounded density, and is continuous up to time $t = 0$.

Our aim is to prove the convergence (up to a subsequence) of m_λ as $\lambda \to +\infty$ to a solution m of

$$\begin{cases} \partial_t m - \operatorname{div}(mD_pH(D_xF(x,m(t)),x)) = 0 & \text{in } \mathbb{R}^d \times (0,+\infty), \\ m(0) = m_0, \quad \text{in } \mathbb{R}^d, \end{cases} \tag{10}$$

and to show also that

$$\lambda u_\lambda(x,t) \to F(x,m(t)) \quad \text{loc. uniformly}, \quad \lambda Du_\lambda(x,t) \to D_xF(x,m(t)) \quad \text{a.e.}$$

2.1 The convergence results

We work under the following conditions: we assume that the initial measure m_0 satisfies

$$m_0 \in \mathcal{P}_2(\mathbb{R}^d) \text{ is absolutely continuous with a bounded density.} \tag{11}$$

The kind of costs we are interested in are non-local and regularizing. A possible assumptions on F is that $F : \mathbb{R}^d \times \mathcal{M}_1(\mathbb{R}^d) \to \mathbb{R}$ is continuous in a suitable topology, has a linear growth and is Lipschitz continuous and semi-concave in x. More precisely, we suppose the existence of a constant $C_o \geqslant 1$ such that:

$$\text{For any } \kappa > 0, \text{ the restrictions of } F \text{ and } D_xF \text{ to } \mathbb{R}^d \times \mathcal{M}_{1,\kappa}(\mathbb{R}^d)$$
$$\text{are continuous in both variables for the topology of } \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d), \tag{12}$$

$$|F(x,m)| \leqslant C_o(1+|x|), \qquad |F(x,m) - F(y,m)| \leqslant C_o|x-y|, \tag{13}$$

$$F(x+h,m) + F(x-h,m) - 2F(x,m) \leqslant C_o|h|^2, \quad \forall\, m \in \mathcal{M}_1(\mathbb{R}^d), h \in \mathbb{R}^d, \tag{14}$$

(recall that $\mathcal{M}_{p,\kappa}(\mathbb{R}^d)$ and $\mathcal{M}_p(\mathbb{R}^d)$ are defined in the introduction). We assume that $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is convex with respect to the first variable and satisfies,

$$-C_o \leqslant H(p,x) \leqslant C_o(1+|p|^2), \qquad D^2_{pp}H(p,x) \geqslant C_o^{-1} I_d, \tag{15}$$

$$|H(p,x) - H(p,y)| + |D_pH(p,x) - D_pH(p,y)| \leqslant C_o|x-y|(1+|p|), \tag{16}$$

$$|H(p,x) - H(q,x)| \leqslant C_o|p-q|(1+|p|+|q|), \tag{17}$$

$$H(p,x+h) + H(p,x-h) - 2H(p,x) \geqslant -C_o|h|^2(1+|p|). \tag{18}$$

Note that, if H is smooth, then conditions (16), (17) and (18) can be equivalently rewritten as

$$|D_xH(p,x)| + |D_{px}H(p,x)| \leqslant C_o(1+|p|),$$
$$|D_pH(p,x)| \leqslant C_o(1+|p|), \qquad D^2_{xx}H(p,x) \geqslant -C_o(1+|p|).$$

Similarly, if F is smooth, the second condition in (13) and (14) can be rewritten as

$$|D_xF(x,m)| \leqslant C_o, \qquad D^2_{xx}F(x,m) \leqslant C_o.$$

Theorem 2.1. *Assume* (11), (12), (13), (14), (15), (16), (17) *and* (18). *Let* (u_λ, m_λ) *be a solution to* (9). *Then* (m_λ) *is relatively compact in* $C^0([0,T], \mathcal{P}_1(\mathbb{R}^d))$ *and is bounded in* $L^\infty(\mathbb{R}^d \times [0,T])$ *for any* $T > 0$. *Moreover, the limit* m*, as* $\lambda_n \to +\infty$*, of any converging subsequence* (m_{λ_n}) *in* $C^0([0,T], \mathcal{P}_1(\mathbb{R}^d))$ *is a solution of* (10) *in the sense of distributions and*

$$\lambda_n u_{\lambda_n}(x,t) \to F(x,m(t)) \quad \textit{locally uniformly and} \quad \lambda_n Du_{\lambda_n}(x,t) \to D_xF(x,m(t)) \quad \textit{a.e.}$$

The existence of a solution to (9) under the assumptions above can be established by standard arguments, using the estimates in Section 2.2 below, see Remark 2.4. A typical example of a Hamiltonian satisfying our assumptions is

$$H(p,x) = -v(x)\cdot p + \frac{1}{2}|p|^2,$$

where the vector field $v : \mathbb{R}^d \to \mathbb{R}^d$ is bounded and with bounded first and second order derivatives.

Remark 2.1. Under somewhat stronger assumptions than in Theorem 2.1 we can also prove that the whole family m_λ converges to m as $\lambda \to +\infty$. This is achieved if the solution m of (10) is unique, as in the problem of Section 3.

The first additional assumption is that m_0 has compact support. Then the support of any solution of (10) is compact in x, as it can be proved, e.g., by the superposition principle in [2] as in [13].

Next we assume that the vector field G appearing in the limit equation (10), $G(x, m) := -D_pH(D_xF(x, m), x)$, is such that, for all $m \in \mathcal{M}_1(\mathbb{R}^d)$, $x \mapsto G(x, m)$ is C^1 and

$$|G(x,m) - G(y,m)| \leqslant C_1|x-y| \,, \quad \|G(\cdot,m) - G(\cdot,\bar{m})\|_\infty \leqslant C_1\mathbf{d}_1(m,\bar{m}) \,,$$

where $\mathbf{d}_1$ is the 1-Wasserstein distance. Then it is proved in [31] that there is a unique solution m of (10) with compact support in x.

Remark 2.2. In the case of deterministic MFGs, $\nu_\lambda = 0$ for all λ, the solution (u_λ, m_λ) is not smooth and the proof of convergence by PDE methods is harder. We can prove a result analogous to Theorem 2.1 under the additional assumption that $\|F(\cdot, m)\|_{C^2} \leqslant C$ for all $m \in \mathcal{M}_1(\mathbb{R}^d)$, and the support of m_0 is compact, using the methods of [13]. Moreover the whole family m_λ converges to m under the additional condition on G of the preceding remark.

Remark 2.3. The case of $\nu_\lambda \to \nu_\infty > 0$ can be treated as in the proof of Theorem 2.1 and leads in the limit to the viscous Fokker-Planck equation

$$\partial_t m - \nu_\infty \Delta m - \operatorname{div}(mD_pH(D_xF(x, m(t)), x)) = 0 \qquad \text{in } \mathbb{R}^d \times (0, +\infty).$$

2.2 Proof of Theorem 2.1

In this part, assumptions (11), (12), (13), (14), (15), (16), (17) and (18) are in force. We start with some estimates for a solution to (9).

Proposition 2.2. *Let* (u_λ, m_λ) *be a solution of* (9). *Then* $|u_\lambda(x,t)| \leqslant \lambda^{-1}\tilde{C}(1 + |x|)$ *for some constant* $\tilde{C}$ *independent of* $\lambda \geqslant 1 + 4d + 16C_0^2$ *such that* $\nu_\lambda \leqslant 1$.

Proof. We note that $w^\pm(x,t) := \pm\lambda^{-1}\tilde{C}(1 + |x|^2)^{1/2}$ is a supersolution (for +) and a subsolution (for -) of (9) for a suitable $\tilde{C}$. Let us determine C such that $w = -\lambda^{-1}C(1 + |x|^2)^{1/2}$ is a subsolution, the other case being easier. By the growth assumptions (13) and (15), and for $\nu_\lambda \leqslant 1$,

$$\begin{aligned} &- \partial_t w - \nu_\lambda \Delta w + \lambda w + \lambda^{-1} H(\lambda Dw, x) - F(x, m_\lambda(t)) \leqslant \\ &\frac{Cd\nu_\lambda}{\lambda(1 + |x|^2)^{1/2}} - C(1 + |x|^2)^{1/2} + \frac{1}{\lambda} H\left(\frac{-Cx}{(1 + |x|^2)^{1/2}}, x\right) - F(x, m_\lambda(t)) \leqslant \\ &\quad \frac{Cd}{\lambda} - \frac{C}{2}(1 + |x|) + \frac{C_0}{\lambda}\left(1 + \frac{C^2|x|^2}{(1 + |x|^2)}\right) + C_0(1 + |x|). \end{aligned}$$

If we choose $C = 4C_0$ the right hand side can be bounded above by

$$\frac{4C_0 d}{\lambda} - C_0 + C_0\frac{1 + 16C_0^2}{\lambda} \leqslant 0$$

if $\lambda \geqslant 1 + 4d + 16C_0^2$. ∎

Proposition 2.3. *Let* (u_λ, m_λ) *be a solution of* (9). *Then* $\|Du_\lambda\|_\infty \leqslant 4\lambda^{-1}C_o$ *for* $\lambda \geqslant 2C_o$.

Proof. We use an a priori estimate, proving that, if u_λ is Lipschitz continuous and if H and $(x,t) \to F(x, m_\lambda(t))$ are smooth, then u_λ satisfies the required estimate. One can then complete the proof easily, approximating the Hamilton-Jacobi (HJ) equation by HJ equations with smooth and globally Lipschitz continuous Hamiltonians and right-hand sides and passing to the limit. We omit this last part which is standard and proceed with the argument.

Given a direction $\xi \in \mathbb{R}^d$ with $|\xi| \leqslant 1$, let $w := Du_\lambda \cdot \xi$. Then w satisfies

$$-\partial_t w - \nu_\lambda \Delta w + \lambda w + D_p H(\lambda Du_\lambda, x) \cdot Dw + \lambda^{-1} D_x H(\lambda Du_\lambda, x) \cdot \xi = D_x F(x, m(t)) \cdot \xi.$$

In view of assumptions (13) and (16) we have therefore

$$-\partial_t w - \nu_\lambda \Delta w + \lambda w + D_p H(\lambda Du_\lambda, x) \cdot Dw - C_o \lambda^{-1}(1 + \lambda \|Du_\lambda\|_\infty) \leqslant C_o.$$

So by the maximum principle we have

$$Du_\lambda \cdot \xi = w \leqslant \lambda^{-1} C_o (1 + \lambda^{-1} + \|Du_\lambda\|_\infty).$$

Taking the supremum over $|\xi| \leqslant 1$, gives the result for λ larger than $2C_o$. ■

Proposition 2.4. *Let (u_λ, m_λ) be a solution of* (9)*. Then $D^2 u_\lambda \leqslant \lambda^{-1}\tilde{C}$, where $\tilde{C}$ does not depend on $\lambda \geqslant 2C_o$.*

Proof. Here again we focus on a priori estimates for smooth data. Given a direction $\xi \in \mathbb{R}^d$ with $|\xi| \leqslant 1$, let $w := Du_\lambda \cdot \xi$ and $z := D^2 u_\lambda \xi \cdot \xi$. Then

$$\begin{aligned} &-\partial_t z - \nu_\lambda \Delta z + \lambda z + D_p H(\lambda Du_\lambda, x) \cdot Dz + 2D^2_{px} H(\lambda Du_\lambda, x)\xi \cdot Dw \\ &\quad + \lambda D^2_{pp} H(\lambda Du_\lambda, x) Dw \cdot Dw + \lambda^{-1} D^2_{xx} H(\lambda Du_\lambda, x)\xi \cdot \xi = D^2_{xx} F(x, m_\lambda(t))\xi \cdot \xi. \end{aligned}$$

Since F is semiconcave in x (14), the right hand side $D^2_{xx} F\xi \cdot \xi$ is bounded above by C_0. The uniform bound on λDu_λ proved in Proposition 2.3 and the assumption (16) imply

$$2D^2_{px} H(\lambda Du_\lambda, x)\xi \cdot Dw \geqslant -2C_0(1 + \lambda |Du_\lambda|)|Du_\lambda| \geqslant -C_1.$$

The same bound on λDu_λ and the assumption (18) imply

$$\lambda^{-1} D^2_{xx} H(\lambda Du_\lambda, x)\xi \cdot \xi \geqslant -C_0(1 + \lambda |Du_\lambda|) \geqslant -C_2.$$

Since H is convex in p, $D^2_{pp} H \geqslant 0$ and we infer that z satisfies

$$-\partial_t z - \nu_\lambda \Delta z + \lambda z + D_p H(\lambda Du_\lambda, x) \cdot Dz \leqslant \tilde{C},$$

where the constant $\tilde{C}$ does not depend on λ and $|\xi| \leqslant 1$. We conclude again by the maximum principle. ■

Proposition 2.5. *Let (u_λ, m_λ) be a solution of* (9)*. For any $T > 0$, the family (m_λ) satisfies*

$$\sup_{\lambda \geqslant 2C_o} \sup_{t \in [0,T]} \int_{\mathbb{R}^d} |x|^2 m_\lambda(x,t) dx < +\infty,$$

is relatively compact in $C^0([0,T], \mathcal{P}_1)$, and bounded in $L^\infty(\mathbb{R}^d \times [0,T])$.

Proof. We do the proof again for smooth data. For the bound on the second order moment of $m_\lambda(t)$ on $[0,T]$ we recall that $m_\lambda(t)$ is the law $\mathcal{L}(X_t)$ of the solution X_t of the SDE

$$dX_t = -D_p H(\lambda Du_\lambda(X_t), X_t)dt + \sqrt{2\nu_\lambda} dW_t, \qquad \mathcal{L}(X_0) = m_0,$$

where W_t is a standard Brownian motion. Since the vector field $D_p H(\lambda Du_\lambda, x)$ is uniformly bounded by Proposition 2.3, we have $\mathbb{E}[|X_t|^2] \leqslant C(\mathbb{E}[|X_0|^2] + 1)e^{Ct}$. Then

$$M_2(m_\lambda(t)) = \int_{\mathbb{R}^d} |x|^2 m_\lambda(x,t) dx = \mathbb{E}|X_t|^2 \leqslant C(M_2(m_0) + 1)e^{CT}, \forall t \leqslant T.$$

For the L^∞ bound on m_λ, we rewrite the equation of m_λ as

$$\begin{aligned} \partial_t m_\lambda - \nu_\lambda \Delta m_\lambda - m_\lambda \mathrm{Tr}\left(D_{pp}H(\lambda Du_\lambda, x)D^2 u_\lambda + D_{px}H(\lambda Du_\lambda, x)\right) \\ - Dm_\lambda \cdot D_p H(Du_\lambda, x) = 0 \qquad \text{in } \mathbb{R}^d \times (0, +\infty) \end{aligned}$$

where, by convexity of H, (15) and Proposition 2.4 on the one hand, and by (16) and Proposition 2.3 on the other hand, we have

$$\mathrm{Tr}\left(D_{pp}H(\lambda Du_\lambda, x)D^2 u_\lambda\right) \leqslant C \text{ and } \mathrm{Tr}\left(D_{px}H(\lambda Du_\lambda, x)\right) \leqslant C,$$

where C does not depend on λ. Therefore, by the maximum principle again, the L^∞ norm of m_λ has at most an exponential growth in time, uniform with respect to λ. ■

Proof of Theorem 2.1. By Proposition 2.5, (m_λ) is relatively compact in $C^0([0,T], \mathcal{P}_1(\mathbb{R}^d))$ and is bounded in $L^\infty(\mathbb{R}^d \times [0,T])$ for any $T > 0$. Let (m_{λ_n}) be a converging subsequence in $C^0([0,T], \mathcal{P}_1)$ for any $T > 0$. Then (m_{λ_n}) converges to m in $L^\infty-$weak-* on $\mathbb{R}^d \times [0,T]$ for any $T > 0$. In particular, by our continuity assumption on F in (12), the maps $(x,t) \to F(x, m_{\lambda_n}(t))$ and $(x,t) \to D_x F(x, m_{\lambda_n}(t))$ converge locally uniformly to the maps $(x,t) \to F(x, m(t))$ and $(x,t) \to D_x F(x, m(t))$ respectively.

As u_λ solves (9), $w_\lambda := \lambda u_\lambda$ solves

$$-\lambda^{-1}\partial_t w_\lambda - \lambda^{-1}\nu_\lambda \Delta w_\lambda + w_\lambda + \lambda^{-1} H(Dw_\lambda, x) = F(x, m_\lambda(t)) \qquad \text{in } \mathbb{R}^d \times [0, +\infty). \quad (19)$$

Hence the half-relaxed limits w^* and w_* of (w_λ) (see, e.g., [3] for their definitions and properties), are locally uniformly bounded in view of Proposition 2.2 and, respectively, viscosity sub- and super-solutions of the zero-th order equation

$$w = F(x, m(t)) \qquad \text{in } \mathbb{R}^d \times [0, +\infty). \quad (20)$$

By a comparison principle we get $w^* = w_*$, and this proves the locally uniform convergence of $(\lambda_n u_{\lambda_n})$ to $F(x,m)$ in $\mathbb{R}^d \times [0, +\infty)$.

Next we use Theorem 3.3.3 in [11]. By Proposition 2.2 (λu_λ) is uniformly locally bounded, and by Proposition 2.4 it is uniformly semi-concave in space (locally in time). Then any sequence $(\lambda_n u_{\lambda_n})$ has a subsequence such that $(\lambda_n Du_{\lambda_n})$ converges to $D_x F(x,m)$ a.e. and therefore also in $L^1_{loc}(\mathbb{R}^d \times [0, +\infty))$. One easily derives from this that m solves (10) in the sense of distribution. ■

Remark 2.4. The existence of a solution (u_λ, m_λ) of the system (9) can be proved by approximating with solutions of the following system with finite time-horizon

$$\begin{cases} -\partial_t u^T - \nu_\lambda \Delta u^T + \lambda u^T + \lambda^{-1} H(\lambda Du^T, x) = F(x, m^T(t)) & \text{in } \mathbb{R}^d \times (0,T) \\ \partial_t m^T - \nu_\lambda \Delta m^T - \mathrm{div}(m^T D_p H(\lambda Du^T, x)) = 0 \qquad \text{in } \mathbb{R}^d \times (0,T) \\ u^T(T) = 0, \qquad m^T(0) = m_0, \qquad \text{in } \mathbb{R}^d. \end{cases} \quad (21)$$

The existence of a classical solution (u^T, m^T) for fixed $\lambda > 0$ follows from standard argument (see for instance Lions' course [29]). The estimates of Propositions 2.2, 2.3, 2.4, and 2.5 hold for (u^T, m^T) with the same proof (using comparison principles for Cauchy problems with constant terminal data). They show in particular that $m^T(t) \in \mathcal{P}_2(\mathbb{R}^d)$ and it has a bounded density for all t. Then there is enough compactness to pass to the limit as $T \to +\infty$, as in the proof of Theorem 2.1, and see that the limit satisfies (9).

2.3 Examples

In this section we present several examples of coupling functions F of the form

$$F(x,m) = k * m(x,t) = \int_{\mathbb{R}^d} k(x-y)m(dy), \quad (22)$$

where the convolution kernel k can take different forms and is at least globally Lipschitz continuous.

$$H(p,x) = |p|^2/2 - v(x)\cdot p,$$

with the vector field v bounded together with its first and second derivatives. Then

$$D_pH(D_xF(x,m),x) = \int_{\mathbb{R}^d} Dk(x-y)m(dy) - v(x),$$

and the limit equation (10) becomes

$$\begin{cases} \partial_t m + \operatorname{div}(m(v - Q[m])) = 0 \text{ in } \mathbb{R}^d \times (0,+\infty), \; Q[m](x,t) = \displaystyle\int_{\mathbb{R}^d} Dk(x-y)m(y,t)dy, \\ m(0) = m_0, \qquad \text{in } \mathbb{R}^d. \end{cases} \tag{23}$$

Note that the condition (12) is satisfied. In addition, we suppose that Dk is bounded, which implies condition (13), and the semi-concavity of k, which ensure condition (14). Under these assumptions Theorem 2.1 holds. Next we review some special cases that arise in applications.

2.3.1 The Aggregation Equation

The special case of (23) with $v \equiv 0$ is often called the Aggregation Equation. For suitable choices of the kernel k it models the collective behaviour of groups of animals, see, e.g., [9, 34] and the references therein. Most kernels used in the aggregation models are of the form $k(x) = \phi(|x|)$ with ϕ smooth but $\phi'(0)$ not necessarily 0, so k can be not differentiable in the origin. However, most of them satisfy the assumptions above.

Example 2.1. The kernel

$$k(x) = \alpha e^{-a|x|}, \qquad a > 0, \tag{24}$$

considered, e.g., in [9, 34] is bounded, globally Lipschitz continuous, and semiconcave if $\alpha > 0$. Note that the case $\alpha > 0$ describes repulsion among individuals at all distances, because $k(x) = \phi(|x|)$ with $\phi(r) = \alpha e^{-ar}$ and hence $\phi'(r) < 0$ implies repulsion. The case $\alpha < 0$, describing attraction, does not fit into our theory because $k \sim |x|$ near 0, so it is not semiconcave, which is consistent with the fact that solutions of the Aggregation Equation (23) are known to blow up in finite time for suitable initial data (at least in dimension $d = 1$, see [9]).

Example 2.2. The kernel

$$k(x) = -|x|e^{-a|x|}, \qquad a > 0, \tag{25}$$

considered in [9] is also bounded, globally Lipschitz continuous and semiconcave because $k \sim -|x|$ near 0. Note that this kernel describes repulsion at small distance and attraction at distance $|x| > 1/a$. Our theory is consistent with the global existence of solutions of the Aggregation Equation (23) in this case, at least for $d = 1$, proved in [9].

Example 2.3. To model repulsion at short distance and attraction at medium range, decaying at infinite, a commonly used kernel is the so-called Morse potential

$$k(x) = e^{-|x|} - Ge^{-|x|/L}, \qquad 0 < G < 1, \quad L > 1, \tag{26}$$

see [7] and the references therein. It is again bounded and globally Lipschitz continuous. It is also semiconcave because $k \sim 1 - G + |x|(G/L - 1)$ near 0, and $G/L - 1 < 0$.

2.3.2 Models of crowd dynamics

There is a large and fast growing literature on models of the interactions among pedestrians, see the survey in the book [22]. They split into first order models, where the velocity of the pedestrian is a prescribed function of the density of individuals and position, and second order models, where the acceleration is prescribed. In the next Section 3 we study second order models, focusing on the celebrated Cucker-Smale model of flocking, see Remark 3.4 for more references on crowd dynamics.

A first order model fitting in the assumptions of the present section is the one proposed in [21], where the velocity of each agent at position x and time t is of the form $v(x)-Q[m(t)](x)$, v being the desired velocity of the pedestrian, and the other term Q accounting for the interaction with the other agents. If we assume that Q does not depend on the angular focus of the walker in position x, then the model in [21] can be written as

$$Q[m(t)](x) = \int_{\mathbb{R}^d} Dk(x-y)m(y,t)dy, \quad k(x) = \phi(|x|)$$

with $\phi \in Lip([0,+\infty))$, decreasing in $(0,r)$, increasing in (r,R), and constant in $[R,+\infty)$, so with a behaviour similar to the Morse kernel (26) and to (25). If we take $\phi \in C^2((0,+\infty))$ with ϕ'' bounded, then F given by (22) satifies the assumptions of Theorem 2.1.

2.3.3 On uniqueness of solutions

If we assume in addition that $k \in C^2(\mathbb{R}^d)$ with D^2k bounded, then the limit equation (23) has a unique solution with compact support in space, as observed in Remark 2.1. This occurs, for instance, in Section 2.3.2 if $\phi \in C^2([0,+\infty))$ and $\phi'(0) = \phi''(0) = 0$ (recall that $k(x) = \phi(|x|)$). Uniqueness is also known for the Aggregation Equation with kernels like those of Section 2.3.1: see [9, 20] and the references therein. However, we expect uniqueness of solutions to the Mean-Field Game system (3) with F given by (22) only for the exponential kernel (24), and not in all other models where there is attraction among individuals in some range of densities. In fact, the uniqueness of solutions in Mean Field Games is strongly connected with a property of mononicity of F discovered by Lasry and Lions [28]. For coupling functions of the form (22) such monotonicity is equivalent to the property that k is a positive semidefinite kernel, namely,

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} k(x-y)v(y)v(x)\,dydx \geqslant 0 \qquad \forall\, v.$$

This property is deeply studied and has several characterizations. If $k(x-y) = \psi(|x-y|^2)$ with $\psi \in C^\infty((0,+\infty))$ and continuous in 0, then it is known that k is a positive semidefinite kernel if and only if ψ is completely monotone, namely, $\psi' \leqslant 0$ and all other derivatives have alternating signs [26]. In all examples describing attraction it occurs that ϕ', and therefore ψ', is instead positive in some range. Then the MFG is not expected to have a unique solution and our result also says that the distance among the possibly multiple solutions of the MFG system tends to 0 as λ becomes large.

3 Convergence for some MFGs of acceleration towards the Cucker-Smale model

For $\lambda > 0$ and $0 < T < +\infty$, we now consider the MFG systems of acceleration, which is written in a formal way as:

$$\begin{cases} -\partial_t u_\lambda + \lambda u_\lambda - v\cdot D_x u_\lambda + \dfrac{\lambda}{2}|D_v u_\lambda|^2 = F(x,v,m_\lambda(t)) & \text{in } \mathbb{R}^{2d}\times(0,T) \\ \partial_t m_\lambda + v\cdot D_x m_\lambda - \operatorname{div}_v(m_\lambda \lambda D_v u_\lambda) = 0 & \text{in } \mathbb{R}^{2d}\times(0,+\infty) \\ m_\lambda(0) = m_0,\ u_\lambda(x,v,T) = 0 & \text{in } \mathbb{R}^{2d}. \end{cases} \tag{27}$$

Here the space variables are denoted by (x, v), with $(x, v) \in \mathbb{R}^d \times \mathbb{R}^d$. System (27) models a Nash equilibrium of a game in which the (small) players, given the flow $(m_\lambda(t))$ of probability measures on $\mathbb{R}^{2d}$, try to minimize over γ the quantity

$$\int_0^T e^{-\lambda t}\left(\frac{1}{2\lambda}|\ddot{\gamma}(t)|^2 + F(\gamma(t), \dot{\gamma}(t), m_\lambda(t))\right) dt,$$

while the flow $(m_\lambda(t))$ is the evolution of the positions and the velocities of the players when they play in an optimal way.

We assume that the coupling function F is a cost associated to the Cucker-Smale model:

$$F(x, v, m(t)) = k * m(x, v, t) = \int_{\mathbb{R}^{2d}} k(x - y, v - v_*)m(y, v_*, t)dydv_*, \quad k(x, v) = \frac{|v|^2}{g(x)}, \quad (28)$$

where $g : \mathbb{R}^d \to \mathbb{R}$ is bounded below by a positive constant, is even, smooth and such that $|Dg|/g$ is globally bounded. For instance,

$$g(x) = (\alpha + |x|^2)^\beta, \quad \alpha > 0, \beta \geqslant 0. \quad (29)$$

In this case $D_v k(x, v) = \frac{2v}{g(x)}$ and so

$$D_v F(x, v, m(t)) = (D_v k) * m(x, v, t) = \int_{\mathbb{R}^{2d}} 2\frac{(v - v_*)}{g(x - y)}m(y, v_*, t)dydv_*.$$

The aim of this section is to show that $m_\lambda \to m$ as $\lambda \to +\infty$, where m solves the continuous version of the Cucker-Smale model:

$$\begin{cases} \partial_t m + v \cdot D_x m - \mathrm{div}_v(mD_v F(x, v, m)) = 0 & \text{in } \mathbb{R}^{2d} \times (0, +\infty), \\ m(0) = m_0, \quad \text{in } \mathbb{R}^{2d}. \end{cases} \quad (30)$$

3.1 The convergence result

Throughout this section, we assume that m_0 and F satisfy the following conditions:

$$m_0 \in \mathcal{P}(\mathbb{R}^{2d}) \text{ has a compact support,} \quad (31)$$

and

$$F \text{ is given by (28) where } g : \mathbb{R}^d \to \mathbb{R} \text{ is bounded below by a positive constant, is even, smooth, and } |Dg|/g \text{ is globally bounded.} \quad (32)$$

Let us start by describing what we mean by a weak (variational) solution of the MFG problem. Let $\Gamma = C^1([0, T], \mathbb{R}^d)$ endowed with usual C^1 norm and $\mathcal{P}(\Gamma)$ be the set of Borel probability measures on Γ. We consider, for $\eta \in \mathcal{P}(\Gamma)$,

$$\mathcal{J}_\lambda(\eta) = \int_\Gamma \int_0^T e^{-\lambda t}\frac{1}{2\lambda}|\ddot{\gamma}(t)|^2 dt\eta(d\gamma) + \int_0^T e^{-\lambda t}\mathcal{F}(m^\eta(t))dt,$$

where $m^\eta(t) = \tilde{e}_t \sharp \eta$ (with $\tilde{e}_t : \Gamma \to \mathbb{R}^{2d}$, $\tilde{e}_t(\gamma) = (\gamma(t), \dot{\gamma}(t))$) and

$$\mathcal{F}(m) = \frac{1}{2}\int_{\mathbb{R}^{4d}} k(x - x_*, v - v_*)m(dx, dv)m(dx_*, dv_*) \qquad \forall m \in \mathcal{P}(\mathbb{R}^{2d}).$$

Lemma 3.1. *For any $\lambda > 0$, there exists at least a minimizer $\bar{\eta}_\lambda$ of $\mathcal{J}_\lambda$ under the constraint $\tilde{e}_0 \sharp \bar{\eta}_\lambda = m_0$. It is a weak solution of the MFG problem of acceleration, in the sense that, for $\bar{\eta}_\lambda$−a.e. $\bar{\gamma} \in \Gamma$,*

$$\int_0^T e^{-\lambda t}(\frac{1}{2\lambda}|\ddot{\bar{\gamma}}(t)|^2 + F(\bar{\gamma}(t), \dot{\bar{\gamma}}(t), m^{\bar{\eta}_\lambda}(t)))dt \quad (33)$$

$$= \inf_{\gamma \in H^2,\ (\gamma(0), \dot{\gamma}(0)) = (\bar{\gamma}(0), \dot{\bar{\gamma}}(0))} \int_0^T e^{-\lambda t}(\frac{1}{2\lambda}|\ddot{\gamma}(t)|^2 + F(\gamma(t), \dot{\gamma}(t), m^{\bar{\eta}_\lambda}(t)))dt.$$

The link between the equilibrium condition (33) and the MFG system (27) is the following: if we set

$$u_\lambda(x,v,s) = \inf_{(\gamma(s),\dot\gamma(s))=(x,v)} \int_s^T e^{-\lambda(t-s)}(\frac{1}{2\lambda}|\ddot\gamma(t)|^2 + F(\gamma(t),\dot\gamma(t),m^{\bar\eta_\lambda}(t)))dt,$$

then the pair $(u_\lambda, m^{\bar\eta_\lambda})$ is (at least formally) a weak solution of (27), in the sense that u_λ is a viscosity solution to the first equation in (27) while $m^{\bar\eta_\lambda}$ is a solution in the sense of distribution of the second equation in (27). Existence of a solution to the equilibrium condition (33) for more general MFG systems is obtained in [12], however under a much more restrictive growth condition on F. In addition, [1, 12] show that there exists a weak solution to the MFG system of acceleration (27).

We postpone the (quite classical) proof of Lemma 3.1 to the next section and proceed with the notion of solution for the kinetic equation (30). Following [10], we say a map $m \in C^0([0,T],\mathcal{P}_2(\mathbb{R}^{2d}))$ is a measure-valued solution to (30) if $m(t) = P^{x,v}(t)\sharp m_0$ where $P^{x,v}(t) = (P_1^{x,v}(t), P_2^{x,v}(t)) \in \mathbb{R}^d \times \mathbb{R}^d$ solves the ODE

$$\begin{cases} \frac{d}{dt}P_1^{x,v}(t) = P_2^{x,v}(t), \\ \frac{d}{dt}P_2^{x,v}(t) = -D_vF(P_1^{x,v}(t), P_2^{x,v}(t), m(t)), \\ P^{x,v}(0) = (x,v). \end{cases} \tag{34}$$

In [10], the authors propose several conditions under which such a measure-valued solution exists and is unique. This include the case of the Cucker-Smale model studied here, under the assumption that m_0 has a compact support.

Our main result is the following:

Theorem 3.2. *Let $\bar\eta_\lambda$ be a minimizer of $\mathcal{J}_\lambda$ under the constraint $\tilde e_0\sharp\bar\eta_\lambda = m_0$. Then $(m^{\bar\eta_\lambda})$ converges as $\lambda \to +\infty$ to the unique measure-valued solution to* (30) *in $C^0_{loc}([0,T),\mathcal{P}_2(\mathbb{R}^{2d}))$.*

Remark 3.1. Note that we do not prove the convergence of all the equilibria $(\bar\eta_\lambda)$ of (33), but only of the minimizers of $\mathcal{J}_\lambda$. The reason is that we were not able to obtain enough estimates for the other equilibria.

3.2 Proof of the convergence result

Before starting the proof, let us note that, by our assumptions, there is a constant $C_0 > 0$ such that

$$g \geqslant C_0^{-1}, \qquad 0 \leqslant F \leqslant C_0(1+|v|^2+M_{2,v}(m)), \quad \text{where } M_{2,v}(m) := \int_{\mathbb{R}^{2d}} |v|^2 m(dx,dv). \tag{35}$$

$$|D_xF(x,v,m)| \leqslant C_0F(x,v,m), \qquad |D_vF(x,v,m)| \leqslant C_0F^{1/2}(x,v,m). \tag{36}$$

Indeed,

$$|D_xF(x,v,m)| \leqslant \int_{\mathbb{R}^{2d}} |Dg(x-x_*)|\frac{|v-v_*|^2}{(g(x-x_*))^2}m(dx_*,dv_*) \leqslant \|Dg/g\|_\infty F(x,v,m),$$

while, as $g \geqslant C_0^{-1}$,

$$\begin{aligned} |D_vF(x,v,m)| &\leqslant \int_{\mathbb{R}^{2d}} \frac{2|v-v_*|}{g(x-x_*)}m(dx_*,dv_*) \\ &\leqslant \Big(\int_{\mathbb{R}^{2d}} \frac{|v-v_*|^2}{g(x-x_*)}m(dx_*,dv_*)\Big)^{1/2}\Big(\int_{\mathbb{R}^{2d}} \frac{4}{g(x-x_*)}m(dx_*,dv_*)\Big)^{1/2} \leqslant 2C_0^{1/2}F^{1/2}(x,v,m). \end{aligned}$$

Throughout the proof (and unless specified otherwise), C denotes a constant which may vary from line to line and depends only on T, d, m_0 and the constant C_0 in (35) and (36).

Let us now explain the existence of a minimizer for $\mathcal{J}_\lambda$.

Proof of Lemma 3.1. Let $\varepsilon > 0$ and η_ε be $\varepsilon-$optimal for $\mathcal{J}_\lambda$ with constraint $\tilde{e}_0\sharp\eta_\varepsilon = m_0$. We define $\eta \in \mathcal{P}(\Gamma)$ by

$$\int_\Gamma \phi(\gamma)\eta(d\gamma) = \int_{\mathbb{R}^{2d}} \phi(t \to x + tv)m_0(dx, dv) \qquad \forall \phi \in C_b^0(\Gamma).$$

Let $\pi_2 : \mathbb{R}^{2d} \to \mathbb{R}^d$ defined by $\pi_2(x, v) = v$. Then $\pi_2\sharp m^\eta(t) = \pi_2\sharp m_0$ for any $t \in [0, T]$ because, for any $\phi \in C_b^0(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \phi(v)\pi_2\sharp m^\eta(dv, t) = \int_\Gamma \phi(\dot{\gamma}(t))\eta(d\gamma) = \int_{\mathbb{R}^{2d}} \phi(\frac{d}{dt}(t \to x + tv))m_0(dx, dv) = \int_{\mathbb{R}^d} \phi(v)\pi_2\sharp m_0(dv).$$

Hence, by $\varepsilon-$optimality of η_ε,

$$\mathcal{J}_\lambda(\eta_\varepsilon) \leqslant \varepsilon + \mathcal{J}_\lambda(\eta) = \varepsilon + \int_\Gamma \int_0^T e^{-\lambda t}\mathcal{F}(m^\eta(t))dt,$$

where, for any $t \geqslant 0$, and as $\pi_2\sharp m^\eta(t) = \pi_2\sharp m_0$,

$$\mathcal{F}(m^\eta(t)) \leqslant C_0 \int_{\mathbb{R}^{4d}} |v - v_*|^2 m^\eta(x, v, t)m^\eta(x_*, v_*, t) \leqslant 2C_0 \int_{\mathbb{R}^{2d}} |v|^2 m^\eta(x, v, t) = 2C_0 M_{2,v}(m_0).$$

This shows that

$$\mathcal{J}_\lambda(\eta_\varepsilon) \leqslant \varepsilon + 2\lambda^{-1}C_0 M_{2,v}(m_0).$$

As $\mathcal{F}$ is nonnegative, this implies that

$$\int_\Gamma \int_0^T e^{-\lambda t}\frac{1}{2\lambda}|\ddot{\gamma}(t)|^2 dt\eta_\varepsilon(d\gamma) \leqslant \mathcal{J}_\lambda(\eta_\varepsilon) \leqslant \varepsilon + 2\lambda^{-1}C_0 M_{2,v}(m_0).$$

As m_0 has a compact support (say contained in B_{R_0}) and the set

$$\{\gamma \in \Gamma,\ |(\gamma(0), \dot{\gamma}(0))| \leqslant R_0,\ \int_0^T e^{-\lambda t}|\ddot{\gamma}(t)|^2 dt \leqslant C\}$$

is compact in Γ for any C, we conclude that the family (η_ε) is tight. By lower semi-continuity of $\mathcal{J}_\lambda$ we can then conclude that there exists a minimizer $\bar{\eta}_\lambda$ of $\mathcal{J}_\lambda$ under the (closed) constraint $\tilde{e}_0\sharp\bar{\eta}_\lambda = m_0$. Note for later use that, in view of the above estimates,

$$\int_\Gamma \int_0^T e^{-\lambda t}\frac{1}{2\lambda}|\ddot{\gamma}(t)|^2 dt\bar{\eta}_\lambda(d\gamma) \leqslant 2\lambda^{-1}C_0 M_{2,v}(m_0),$$

so that, as m_0 has a compact support,

$$\sup_{t\in[0,T]} M_{2,v}(m^{\bar{\eta}_\lambda}(t)) \leqslant C_\lambda, \tag{37}$$

for some constant C_λ depending on m_0, C_0 and λ.

Next we show equality (33). Let γ_0 belong to the support of $\bar{\eta}_\lambda$ and set $(x_0, v_0) = (\gamma_0(0), \dot{\gamma}_0(0))$. Fix $\gamma_1 \in H^2([0, T], \mathbb{R}^d)$ with $(\gamma_1(0), \dot{\gamma}_1(0)) = (x_0, v_0)$. For $\varepsilon, \delta > 0$, let $E_\varepsilon = \{\gamma \in \Gamma,\ \|\gamma - \gamma_0\|_{C^1} \leqslant \varepsilon\}$, $\tilde{m}_\varepsilon = \tilde{e}_t\sharp(\bar{\eta}_\lambda\lfloor E_\varepsilon)$ and define the Borel measure $\eta_{\varepsilon,\delta}$ on Γ by

$$\int_\Gamma \phi(\gamma)\eta_{\varepsilon,\delta}(d\gamma) = \int_{E_\varepsilon^c} \phi(\gamma)\bar{\eta}_\lambda(d\gamma) + (1 - \delta)\int_{E_\varepsilon} \phi(\gamma)\bar{\eta}_\lambda(d\gamma)$$
$$+ \delta\int_{\mathbb{R}^{2d}} \phi\Big(t \to \gamma_1(t) + (x - x_0 + t(v - v_0))\Big)\tilde{m}_\varepsilon(dx, dv, 0)$$

for any $\phi \in C_b^0(\Gamma)$. Let $\hat{m}_\varepsilon(t)$ be the Borel measure on $\mathbb{R}^{2d}$ defined by

$$\int_{\mathbb{R}^{2d}} \phi(x, v)\hat{m}_\varepsilon(dx, dv, t) = \int_{\mathbb{R}^{2d}} \phi(\gamma_1(t) + (x - x_0 + t(v - v_0)))\tilde{m}_\varepsilon(dx, dv, 0), \qquad \forall \phi \in C_b^0(\mathbb{R}^{2d}).$$

We note that

$$m^{\eta_{\varepsilon,\delta}}(t) = m^{\bar\eta_\lambda}(t) + \delta(\hat m_\varepsilon(t) - \tilde m_\varepsilon(t)), \qquad m^{\eta_{\varepsilon,\delta}}(0) = m_0. \tag{38}$$

Hence, testing the optimality of $\bar\eta_\lambda$ for $\mathcal{J}_\lambda$ against $\eta_{\varepsilon,\delta}$ and using the definition of $\eta_{\varepsilon,\delta}$, we obtain after simplification

$$\begin{aligned}\delta\int_{E_\varepsilon}\int_0^T e^{-\lambda t}\frac{1}{2\lambda}|\ddot\gamma(t)|^2 dt\bar\eta_\lambda(d\gamma) + \int_0^T e^{-\lambda t}\mathcal{F}(m^{\bar\eta_\lambda}(t))dt \\ \leqslant \delta(\int_{\mathbb{R}^{2d}}\tilde m_\varepsilon(dx,dv,0))(\int_0^T e^{-\lambda t}\frac{1}{2\lambda}|\ddot\gamma_1(t)|^2 dt) + \int_0^T e^{-\lambda t}\mathcal{F}(m^{\eta_{\varepsilon,\delta}}(t))dt.\end{aligned}$$

By definition of $\mathcal{F}$ and the fact that k is even, (38) implies that

$$\begin{aligned}&\delta\int_{E_\varepsilon}\int_0^T e^{-\lambda t}\frac{1}{2\lambda}|\ddot\gamma(t)|^2 dt\bar\eta_\lambda(d\gamma) \leqslant \delta(\int_{\mathbb{R}^{2d}}\tilde m_\varepsilon(dx,dv,0))(\int_0^T e^{-\lambda t}\frac{1}{2\lambda}|\ddot\gamma_1(t)|^2 dt)\\ &+\delta\int_0^T e^{-\lambda t}\int_{\mathbb{R}^{4d}} k(x-x_*, v-v_*)m^{\bar\eta_\lambda}(dx_*,dv_*,t)(\hat m_\varepsilon - \tilde m_\varepsilon)(dx,dv,t)dt\\ &+\frac{\delta^2}{2}\int_0^T e^{-\lambda t}\int_{\mathbb{R}^{4d}} k(x-x_*,v-v_*)(\hat m_\varepsilon-\tilde m_\varepsilon)(dx_*,dv_*,t)(\hat m_\varepsilon-\tilde m_\varepsilon)(dx,dv,t)dt.\end{aligned}$$

We divide by $\delta > 0$ and let $\delta \to 0$ to obtain, using the definition of F:

$$\begin{aligned}\int_{E_\varepsilon}\int_0^T e^{-\lambda t}\frac{1}{2\lambda}|\ddot\gamma(t)|^2 dt\bar\eta_\lambda(d\gamma) \leqslant (\int_{\mathbb{R}^{2d}}\tilde m_\varepsilon(dx,dv,0))(\int_0^T e^{-\lambda t}\frac{1}{2\lambda}|\ddot\gamma_1(t)|^2 dt)\\ +\int_0^T e^{-\lambda t}\int_{\mathbb{R}^{2d}} F(x,v,m^{\bar\eta_\lambda}(t))(\hat m_\varepsilon(dx,dv,t) - \tilde m_\varepsilon(dx,dv,t))dt.\end{aligned}$$

Rearranging, we find by the definition of $\tilde m_\varepsilon$ and $\hat m_\varepsilon$:

$$\int_{E_\varepsilon}\int_0^T e^{-\lambda t}(\frac{1}{2\lambda}|\ddot\gamma(t)|^2 + F(\gamma(t),\dot\gamma(t),m^{\bar\eta_\lambda}(t)))dt\ \bar\eta_\lambda(d\gamma) \tag{39}$$

$$\leqslant \int_{\mathbb{R}^{2d}}\int_0^T e^{-\lambda t}(\frac{1}{2\lambda}|\ddot\gamma_1(t)|^2 + F(\gamma_1(t)+x-x_0,\dot\gamma_1(t)+v-v_0,m^{\bar\eta_\lambda}(t)))dt\ \tilde m_\varepsilon(dx,dv,0). \tag{40}$$

Fix $\kappa > 0$ small. By lower-semicontinuity on Γ of the functional

$$\gamma \to \int_0^T e^{-\lambda t}(\frac{1}{2\lambda}|\ddot\gamma(t)|^2 + F(\gamma(t),\dot\gamma(t),m^{\bar\eta_\lambda}(t)))dt,$$

we have, for any $\varepsilon > 0$ small enough, that, for any $\gamma \in E_\varepsilon$,

$$\begin{aligned}\int_0^T e^{-\lambda t}(\frac{1}{2\lambda}|\ddot\gamma_0(t)|^2 + F(\gamma_0(t),\dot\gamma_0(t),m^{\bar\eta_\lambda}(t)))dt \\ \leqslant \int_0^T e^{-\lambda t}(\frac{1}{2\lambda}|\ddot\gamma(t)|^2 + F(\gamma(t),\dot\gamma(t),m^{\bar\eta_\lambda}(t)))dt + \kappa.\end{aligned}$$

On the other hand, by the regularity of F in (36) and the bound on $M_{2,v}(m^{\bar\eta_\lambda}(t))$ in (37), we have, for $|(x,v)| \leqslant \varepsilon$ and $\varepsilon \in (0,1)$,

$$\begin{aligned}\int_0^T e^{-\lambda t}F(\gamma_1(t)+x-x_0,\dot\gamma_1(t)+v-v_0,m^{\bar\eta_\lambda}(t))dt \\ \leqslant \int_0^T e^{-\lambda t}F(\gamma_1(t),\dot\gamma_1(t),m^{\bar\eta_\lambda}(t))dt + C(\gamma_1,\lambda)\varepsilon.\end{aligned}$$

Plugging the inequalities above into (39) gives

$$\bar\eta_\lambda(E_\varepsilon)\Big(\int_0^T e^{-\lambda t}(\frac{1}{2\lambda}|\ddot\gamma_0(t)|^2+F(\gamma_0(t),\dot\gamma_0(t),m^{\bar\eta_\lambda}(t)))dt+\kappa\Big)$$
$$\leqslant (\int_{\mathbb{R}^{2d}}\tilde m_\varepsilon(dx,dv,0))\Big(\int_0^T e^{-\lambda t}(\frac{1}{2\lambda}|\ddot\gamma_1(t)|^2+F(\gamma_1(t),\dot\gamma_1(t),m^{\bar\eta_\lambda}(t)))dt+C(\gamma_1,\lambda)\varepsilon\Big).$$

As $\bar\eta_\lambda(E_\varepsilon)=(\int_{\mathbb{R}^{2d}}\tilde m_\varepsilon(dx,dv,0))$, we can divide the inequality above by this quantity (which is positive since γ_0 is in the support of $\bar\eta_\lambda$) and then let $\varepsilon\to 0$, $\kappa\to 0$ to obtain

$$\int_0^T e^{-\lambda t}(\frac{1}{2\lambda}|\ddot\gamma_0(t)|^2+F(\gamma_0(t),\dot\gamma_0(t),m^{\bar\eta_\lambda}(t)))dt$$
$$\leqslant \int_0^T e^{-\lambda t}(\frac{1}{2\lambda}|\ddot\gamma_1(t)|^2+F(\gamma_1(t),\dot\gamma_1(t),m^{\bar\eta_\lambda}(t)))dt,$$

which gives (33). ■

From now on we fix $\bar\eta_\lambda$ a minimizer of $\mathcal{J}_\lambda$ under the constraint $\tilde e_0\sharp\bar\eta_\lambda=m_0$ and set

$$u_\lambda(x,v,s)=\inf_{\gamma\in H^2,(\gamma(s),\dot\gamma(s))=(x,v)}\int_s^T e^{-\lambda(t-s)}(\frac{1}{2\lambda}|\ddot\gamma(t)|^2+F(\gamma(t),\dot\gamma(t),m^{\bar\eta_\lambda}(t)))dt.$$

We now note that this value function is bounded:

Lemma 3.3. *We have*

$$\mathcal{J}_\lambda(\bar\eta_\lambda)\leqslant 2C_0\lambda^{-1}M_{2,v}(m_0), \tag{41}$$

and, for any $0\leqslant s\leqslant t\leqslant T$,

$$M_{2,v}(m^{\bar\eta_\lambda}(t))=\int_{\mathbb{R}^{2d}}|v|^2 m^{\bar\eta_\lambda}(dx,dv,t)\leqslant 2(1+4C_0\lambda^{-1}e^{\lambda(t-s)})M_{2,v}(m^{\bar\eta_\lambda}(s)) \tag{42}$$

and

$$0\leqslant u_\lambda(x,v,s)\leqslant C\lambda^{-1}(1+|v|^2+M_{2,v}(m^{\bar\eta_\lambda}(s))).$$

Remark 3.2. We use here the fact that we work in a finite horizon problem to obtain the last inequality from (42): see the end of the proof.

Proof. The key point of the proof consists in refining the estimate (37) obtained in the proof of Lemma 3.1. For this we need to introduce a few notations. Given $s\in[0,T)$, let $\Gamma_s=C^1([s,T],\mathbb{R}^d)$ and, for $\eta\in\mathcal{P}(\Gamma_s)$,

$$\mathcal{J}_{\lambda,s}(\eta):=\int_{\Gamma_s}\int_s^T e^{-\lambda(t-s)}\frac{1}{2\lambda}|\ddot\gamma(t)|^2dt\eta(d\gamma)+\int_s^T e^{-\lambda(t-s)}\mathcal{F}(m^\eta(t))dt.$$

By dynamic programming principle (see Lemma 3.4 below), the restriction $\bar\eta_{\lambda,s}$ of $\bar\eta_\lambda$ defined by

$$\int_{\Gamma_s}\phi(\gamma)\bar\eta_{\lambda,s}(d\gamma):=\int_\Gamma\phi(\gamma_{|_{[s,T]}})\bar\eta_\lambda(d\gamma)\qquad\forall\phi\in C^0_b(\Gamma_s),$$

is a minimizer of $\eta\to\mathcal{J}_{\lambda,s}(\eta)$ on $\mathcal{P}(\Gamma_s)$ under the constraint $\tilde e_s\sharp\eta=\tilde e_s\sharp\bar\eta_\lambda$.

Defining $\eta\in\mathcal{P}(\Gamma_s)$ by

$$\int_{\Gamma_s}\phi(\gamma)\eta(d\gamma)=\int_{\mathbb{R}^{2d}}\phi(t\to x+tv)m^{\bar\eta_\lambda}(dx,dv,s)\qquad\forall\phi\in C^0_b(\Gamma_s),$$

we obtain

$$\mathcal{J}_{\lambda,s}(\bar\eta_{\lambda,s})\leqslant\mathcal{J}_\lambda(\eta)=\int_\Gamma\int_s^T e^{-\lambda(t-s)}\mathcal{F}(m^\eta(t))dt,$$

where, as in the proof of Lemma 3.1, for any $t \geqslant s$,

$$\mathcal{F}(m^\eta(t)) \leqslant C_0 \int_{\mathbb{R}^{2d}} |v - v_*|^2 m^\eta(x,v,t) m^\eta(x_*,v_*,t) \leqslant 2C_0 M_{2,v}(m^{\bar\eta_\lambda}(s)).$$

This shows that

$$\mathcal{J}_{\lambda,s}(\bar\eta_{\lambda,s}) \leqslant 2\lambda^{-1} C_0 M_{2,v}(m^{\bar\eta_\lambda}(s)) \tag{43}$$

and inequality (41) holds if we choose $s = 0$.

Next we note that $M_{2,v}(m^{\bar\eta_\lambda}(t))$ is finite: we have, for $\bar\eta_\lambda$−a.e. $\bar\gamma$, and any $0 \leqslant s \leqslant t \leqslant T$,

$$|\dot{\bar\gamma}(t) - \dot{\bar\gamma}(s)| \leqslant \Big(\int_s^t e^{-\lambda(\tau - s)} |\ddot{\bar\gamma}(\tau)|^2 d\tau\Big)^{1/2} \Big(\int_s^t e^{\lambda(\tau-s)} d\tau\Big)^{1/2},$$

so that (by the elementary inequality $a^2 - 2b^2 \leqslant 2|a-b|^2$),

$$|\dot{\bar\gamma}(t)|^2 \leqslant 2|\dot{\bar\gamma}(s)|^2 + 2\lambda^{-1} e^{\lambda(t-s)} \Big(\int_s^t e^{-\lambda(\tau-s)} |\ddot{\bar\gamma}(\tau)|^2 d\tau\Big).$$

Integrating with respect to $\bar\eta_{\lambda,s}$ gives, using (43) in the last inequality,

$$\begin{aligned}
\int_{\mathbb{R}^{2d}} |v|^2 m^{\bar\eta_\lambda}(dx,dv,t) &= \int_\Gamma |\dot{\bar\gamma}(t)|^2 \bar\eta_\lambda(d\bar\gamma) = \int_{\Gamma_s} |\dot{\bar\gamma}(t)|^2 \bar\eta_{\lambda,s}(d\bar\gamma) \\
&\leqslant 2\int_{\Gamma_s} |\dot{\bar\gamma}(s)|^2 \bar\eta_{\lambda,s}(d\bar\gamma) + 2\lambda^{-1} e^{\lambda(t-s)} \int_{\Gamma_s}\int_s^T e^{-\lambda(\tau-s)} |\ddot{\bar\gamma}(\tau)|^2 d\tau \bar\eta_{\lambda,s}(d\bar\gamma) \\
&\leqslant 2\int_{\mathbb{R}^{2d}} |v|^2 m^{\bar\eta_\lambda}(dx,dv,s) + 4e^{\lambda(t-s)} \mathcal{J}_{\lambda,s}(\bar\eta_{\lambda,s}) \\
&\leqslant 2(1 + 4C_0\lambda^{-1} e^{\lambda(t-s)}) M_{2,v}(m^{\bar\eta_\lambda}(s)).
\end{aligned}$$

This proves (42). Finally, using $\gamma(t) = x + (t-s)v$ as a test function for $u_\lambda(x,v,s)$, we have:

$$\begin{aligned}
u_\lambda(x,v,s) &\leqslant \int_s^T e^{-\lambda(t-s)} F(x + (t-s)v, v, m^{\bar\eta_\lambda}(t)) dt \\
&\leqslant \int_s^T e^{-\lambda(t-s)} C_0 (1 + |v|^2 + M_{2,v}(m^{\bar\eta_\lambda}(t))) dt,
\end{aligned}$$

which gives the result thanks to (42). Note that if we were working with an infinite horizon problem, the right-hand side of the inequality above could be unbounded. ■

Lemma 3.4. *Under the notation of the proof of Lemma 3.3 and for any $s \in [0,T)$, $\bar\eta_{\lambda,s}$ is a minimizer of $\eta \to \mathcal{J}_{\lambda,s}(\eta)$ under the constraint $\tilde e_s \sharp \eta = \tilde e_s \sharp \bar\eta_\lambda$.*

Proof. Let us set, for $m \in \mathcal{P}(\mathbb{R}^{2d})$ and $s \in [0,T)$,

$$\mathcal{V}_\lambda(m,s) = \inf\{\mathcal{J}_{\lambda,s}(\eta),\ \eta \in \mathcal{P}(\Gamma_s),\ \tilde e_s \sharp \eta = m\}.$$

We claim that

$$\begin{aligned}
\mathcal{V}_\lambda(m_0,0) &= \inf_{\eta \in \mathcal{P}(\Gamma), \tilde e_0 \sharp \eta = m_0} \int_0^s e^{-\lambda\tau} \Big(\int_\Gamma \frac{1}{2\lambda} |\ddot\gamma(\tau)|^2 \eta(d\gamma) + \mathcal{F}(m^\eta(\tau))\Big) d\tau + e^{-\lambda s} \mathcal{V}_\lambda(m^\eta(s),s) \\
&= \int_0^s e^{-\lambda\tau} \Big(\int_\Gamma \frac{1}{2\lambda} |\ddot\gamma(\tau)|^2 \bar\eta_\lambda(d\gamma) + \mathcal{F}(m^{\bar\eta_\lambda}(\tau))\Big) d\tau + e^{-\lambda s} \mathcal{V}_\lambda(m^{\bar\eta_\lambda}(s),s),
\end{aligned} \tag{44}$$

which proves the lemma. The proof of (44) is a straightforward application of the usual techniques of dynamic programming, the only point being to be able to concatenate at time s two measures $\eta_1 \in \mathcal{P}(\Gamma)$ and $\eta_2 \in \mathcal{P}(\Gamma_s)$ such that $m := \tilde e_s \sharp \eta_1 = \tilde e_s \sharp \eta_2$. For this, let us denote by $\gamma_1 \wedge \gamma_2$ (for $\gamma_1 \in \Gamma$ and $\gamma_2 \in \Gamma_s$ such that $(\gamma_1(s), \dot\gamma_1(s)) = (\gamma_2(s), \dot\gamma_2(s))$) the map in Γ such that

$$\gamma_1 \wedge \gamma_2(t) = \begin{cases} \gamma_1(t) & \text{if } t \in [0, s], \\ \gamma_2(t) & \text{if } t \in [s, T]. \end{cases}$$

In order to define the concatenation $\eta_1 \wedge \eta_2$, we disintegrate η_1 (respectively η_2) with respect to the measure m. We have

$$\eta_1(d\gamma) = \int_{\mathbb{R}^{2d}} \eta_{1,x,v}(d\gamma) m(dx, dv) \qquad (\text{resp. } \eta_2(d\gamma) = \int_{\mathbb{R}^{2d}} \eta_{2,x,v}(d\gamma) m(dx, dv)),$$

where for $m-$a.e. (x, v) and for $(\eta_{1,x,v} + \eta_{2,x,v})-$a.e. γ, one has $(\gamma(s), \dot\gamma(s)) = (x, v)$. We then define $\eta_1 \wedge \eta_2 \in \mathcal{P}(\Gamma)$ by

$$\int_\Gamma \phi(\gamma)(\eta_1 \wedge \eta_2)(d\gamma) = \int_{\mathbb{R}^{2d}} \int_{\Gamma\times\Gamma_s} \phi(\gamma_1 \wedge \gamma_2) \eta_{1,x,v}(d\gamma_1) \eta_{2,x,v}(d\gamma_2) m(dx, dv) \qquad \forall \phi \in C^0_b(\Gamma).$$

By construction we have $m^{\eta_1 \wedge \eta_2}(t) = m^{\eta_1}(t)$ if $t \in [0, s]$, $m^{\eta_1 \wedge \eta_2}(t) = m^{\eta_2}(t)$ if $t \in [s, T]$ and

$$\begin{aligned}
&\int_0^T e^{-\lambda\tau}\Big(\int_\Gamma \frac{1}{2\lambda}|\ddot\gamma(\tau)|^2(\eta_1 \wedge \eta_2)(d\gamma) + \mathcal{F}(m^{\eta_1 \wedge \eta_2}(\tau))\Big)d\tau \\
&\quad = \int_0^s e^{-\lambda\tau}\Big(\int_\Gamma \frac{1}{2\lambda}|\ddot\gamma_1(\tau)|^2 \eta_1(d\gamma_1) + \mathcal{F}(m^{\eta_1}(\tau))\Big)d\tau \\
&\qquad + e^{-\lambda s}\int_s^T e^{-\lambda(\tau - s)}\Big(\int_{\Gamma_s} \frac{1}{2\lambda}|\ddot\gamma_2(\tau)|^2 \eta_2(d\gamma_2) + \mathcal{F}(m^{\eta_2}(\tau))\Big)d\tau.
\end{aligned}$$

The rest of the proof of (44) follows then the usual arguments of dynamic programming. ∎

As u_λ is the value function of an optimal control problem with smooth (in space) coefficients, it is locally Lipschitz continuous. We now evaluate its derivative with respect to v:

Lemma 3.5. *For any $\varepsilon > 0$, $\lambda \geqslant 32\varepsilon^{-2}$, we have*

$$|D_v u_\lambda(x, v, s)| \leqslant C_1(\lambda^{-1/2} u_\lambda^{1/2}(x, v, s) + \varepsilon u_\lambda(x, v, s)) \qquad \text{for a.e. } (x, v, s) \in \mathbb{R}^{2d} \times [0, T - \varepsilon],$$

where $C_1 = C_0 + 1$.

Proof. Let $\varepsilon > 0$, (x, v, s) be a point of differentiability of u_λ with $s \in [0, T - \varepsilon]$. Let $z^\varepsilon : [0, +\infty) \to \mathbb{R}$ be defined by $z^\varepsilon(t) = t - \frac{2t^2}{\varepsilon} + \frac{t^3}{\varepsilon^2}$ on $[0, \varepsilon]$ and $z^\varepsilon(t) = 0$ on $[\varepsilon, +\infty)$. Then $z^\varepsilon(0) = z^\varepsilon(\varepsilon) = \dot z^\varepsilon(\varepsilon) = 0$, $\dot z^\varepsilon(0) = 1$ and $z^\varepsilon \in H^2([0, +\infty))$. Therefore, if $\bar\gamma$ is optimal for $u_\lambda(x, v, s)$, we have, for any $h \in \mathbb{R}^d$ and using $t \to \bar\gamma(t) + z^\varepsilon(t - s)h$ as a competitor in $v_\lambda(x, v + h, s)$:

$$\begin{aligned}
&u_\lambda(x, v + h, s) \\
&\leqslant \int_s^T e^{-\lambda(t-s)}\Big(\frac{1}{2\lambda}|\ddot{\bar\gamma}(t) + \ddot z^\varepsilon(t - s)h|^2 + F(\bar\gamma(t) + z^\varepsilon(t - s)h, \dot{\bar\gamma}(t) + \dot z^\varepsilon(t - s)h, m^{\bar\eta_\lambda}(t))\Big)dt \\
&\leqslant u_\lambda(x, v) + \int_s^{s+\varepsilon} e^{-\lambda(t-s)}\Big(\frac{1}{\lambda}\ddot{\bar\gamma}(t)\cdot(\ddot z^\varepsilon(t - s)h) + \frac{1}{2\lambda}|\ddot z^\varepsilon(t - s)|^2|h|^2 \\
&\qquad + \int_0^1 (D_x F \cdot (z^\varepsilon(t - s)h) + D_v F \cdot (\dot z^\varepsilon(t - s)h))d\tau\Big)dt
\end{aligned}$$

where for simplicity we have omitted the argument $(\bar\gamma(t) + \tau z^\varepsilon(t - s)h, \dot{\bar\gamma}(t) + \tau \dot z^\varepsilon(t - s)h, m^{\bar\eta_\lambda}(t)))$ after $D_x F$ and $D_v F$. Dividing by $|h|$ and letting $h \to 0$ shows that

$$|D_v u_\lambda(x, v, s)| \leqslant \int_s^{s+\varepsilon} e^{-\lambda(t-s)}\Big(\frac{1}{\lambda}|\ddot{\bar\gamma}(t)|\ |\ddot z^\varepsilon(t - s)| + |D_x F|\ |z^\varepsilon(t - s)| + |D_v F|\ |\dot z^\varepsilon(t - s)|\Big)dt,$$

where, from now on, F, D_xF and D_vF have for argument $(\bar\gamma(t), \dot{\bar\gamma}(t), m^{\bar\eta_\lambda}(t))$. Recalling (36) and the expression of z^ε we get

$$\begin{aligned}|D_v u_\lambda(x,v,s)| &\leqslant \lambda^{-1}(\int_s^{s+\varepsilon} e^{-\lambda(t-s)}|\ddot{\bar\gamma}(t)|^2 dt)^{1/2}(\int_s^{s+\varepsilon} e^{-\lambda(t-s)}\ |\ddot z^\varepsilon(t-s)|^2 dt)^{1/2} \\ &\qquad + C_0\varepsilon\int_s^{s+\varepsilon} e^{-\lambda(t-s)} F dt + C_0\int_s^{s+\varepsilon} e^{-\lambda(t-s)}F^{1/2}dt \\ &\leqslant \left(\frac{2}{\lambda}\right)^{1/2}\left(\frac{16}{\varepsilon^2}\frac{1-e^{-\lambda\varepsilon}}{\lambda}\right)^{1/2} u_\lambda^{1/2}(x,v,s) + C_0\varepsilon u_\lambda(x,v,s) + C_0\lambda^{-1/2}u_\lambda^{1/2}(x,v,s).\end{aligned}$$

So, if $\lambda \geqslant 32\varepsilon^{-2}$, we obtain

$$|D_v u_\lambda(x,v,s)| \leqslant (C_0+1)\lambda^{-1/2}u_\lambda^{1/2}(x,v,s) + C_0\varepsilon u_\lambda(x,v,s).$$

■

Lemma 3.6. *For any $\varepsilon > 0$, $\lambda \geqslant 32\varepsilon^{-2}$, $(x,v) \in \mathbb{R}^{2d}$ and $\bar\gamma$ optimal for $u_\lambda(x,v,0)$, we have, for any $t \in [0, T-\varepsilon]$,*

$$|\ddot{\bar\gamma}(t)| \leqslant 2C_1\Big(\lambda^{1/2}u_\lambda^{1/2}(\bar\gamma(t),\dot{\bar\gamma}(t),t) + \varepsilon\lambda u_\lambda(\bar\gamma(t),\dot{\bar\gamma}(t),t)\Big),$$

where C_1 is the constant in Lemma 3.5.

Remark 3.3. In fact we expect that $\ddot{\bar\gamma}(t) = -\lambda D_v u_\lambda(\bar\gamma(t),\dot{\bar\gamma}(t),t)$ for any $t \in (0,T]$, which would imply the lemma (without the "2" in the right-hand side) thanks to Lemma 3.5. This equality is known to hold in several frameworks [3, 11], but we are not aware of a reference for our precise setting. The estimate in Lemma 3.6, much simpler to prove, suffices however for our purpose.

Proof. As $\bar\gamma$ is a minimizer of a calculus of variation problem with smooth coefficients and with quadratic growth, it is known that $\bar\gamma$ satisfies the Euler-Lagrange equation

$$\frac{d^2}{dt^2}(\lambda^{-1}e^{-\lambda t}\ddot{\bar\gamma}_\lambda(t)) = \frac{d}{dt}\left(e^{-\lambda t}D_vF(\bar\gamma_\lambda(t),\dot{\bar\gamma}_\lambda(t),m^{\bar\eta_\lambda}(t))\right) - e^{-\lambda t}D_xF(\bar\gamma_\lambda(t),\dot{\bar\gamma}_\lambda(t),m^{\bar\eta_\lambda}(t)).$$

Therefore $\bar\gamma$ is actually of class H^4 and, in particular, C^3.

Fix $h > 0$ small and let $\gamma_h(s) = \bar\gamma(t) + (s-t)\dot{\bar\gamma}(t)$. By dynamic programming principle and the optimality of $\bar\gamma$ we have:

$$\begin{aligned}&u_\lambda(\bar\gamma(t),\dot{\bar\gamma}(t),t) \\ &\quad = \int_t^{t+h} e^{-\lambda(s-t)}(\frac{1}{2\lambda}|\ddot{\bar\gamma}(s)|^2 + F(\bar\gamma(s),\dot{\bar\gamma}(s),m^{\bar\eta_\lambda}(s)))ds + e^{-\lambda h}u_\lambda(\bar\gamma(t+h),\dot{\bar\gamma}(t+h),t+h) \\ &\quad \leqslant \int_t^{t+h} e^{-\lambda(s-t)}F(\gamma(s),\dot\gamma(s),m^{\bar\eta_\lambda}(s)))ds + e^{-\lambda h}u_\lambda(\gamma(t+h),\dot\gamma(t+h),t+h). \qquad (45)\end{aligned}$$

Note that, by C^3 regularity of $\bar\gamma$, $|\bar\gamma(t+h) - \gamma(t+h)| \leqslant C_\gamma h^2$ (where, here and below, C_γ depends here on γ and on λ). So, as u_λ is locally Lipschitz continuous and $\dot\gamma(t+h) = \dot{\bar\gamma}(t)$, we get

$$\begin{aligned}&u_\lambda(\gamma(t+h),\dot\gamma(t+h),t+h) - u_\lambda(\bar\gamma(t+h),\dot{\bar\gamma}(t+h),t+h) \\ &\qquad \leqslant u_\lambda(\bar\gamma(t+h),\dot{\bar\gamma}(t),t+h) - u_\lambda(\bar\gamma(t+h),\dot{\bar\gamma}(t+h),t+h) + C_\gamma h^2.\end{aligned}$$

Still by C^3 regularity of $\bar\gamma$ we also have $|\dot{\bar\gamma}(t+h) - \dot{\bar\gamma}(t) - \ddot{\bar\gamma}(t)h| \leqslant C_\gamma h^2$. Now the bound on $D_v u_\lambda$ of Lemma 3.5 yields (setting $(x,v) = (\bar\gamma(t),\dot{\bar\gamma}(t))$)

$$\begin{aligned}&u_\lambda(\gamma_h(t+h),\dot\gamma_h(t+h),t+h) - u_\lambda(\bar\gamma(t+h),\dot{\bar\gamma}(t+h),t+h) \\ &\qquad \leqslant C_1(\lambda^{-1/2}u_\lambda^{1/2}(x,v,t) + \varepsilon u_\lambda(x,v,t))|\ddot{\bar\gamma}(t)|h + C_\gamma h^2.\end{aligned}$$

Plugging this inequality into (45) gives, after dividing by h and letting $h \to 0^+$,

$$\frac{1}{2\lambda}|\ddot{\bar{\gamma}}(t)|^2 + F(\bar{\gamma}(t), \dot{\bar{\gamma}}(t), m^{\bar{\eta}_\lambda}(t)) \leqslant F(\gamma(t), \dot{\gamma}(t), m^{\bar{\eta}_\lambda}(t)) + C_1(\lambda^{-1/2}u_\lambda^{1/2}(x,v,t) + \varepsilon u_\lambda(x,v,t))|\ddot{\bar{\gamma}}(t)|.$$

Recalling that $(\gamma(t), \dot{\gamma}(t)) = (\bar{\gamma}(t), \dot{\bar{\gamma}}(t))$ gives the result. ■

In order to state the next Lemma, let us denote by $\bar{\eta}_{\lambda,\varepsilon}$ the restriction of $\bar{\eta}_\lambda$ to $[0, T-\varepsilon]$: more precisely $\bar{\eta}_{\lambda,\varepsilon} = \pi_\varepsilon \sharp \bar{\eta}_\lambda$ where $\pi_\varepsilon : C^1([0,T], \mathbb{R}^d) \to C^1([0, T-\varepsilon], \mathbb{R}^d)$ is the canonical restriction.

Lemma 3.7. *There exists $\varepsilon_0 > 0$ and a constant $C > 0$ such that, for any $\varepsilon \in (0, \varepsilon_0]$, any $\lambda \geqslant (32\varepsilon^{-2}) \vee 1$ and any $t \in [0, T-\varepsilon]$, the support of $m^{\bar{\eta}_\lambda}(t)$ is contained in B_C and*

$$\|\ddot{\bar{\gamma}}\|_{L^\infty([0,T-\varepsilon])} \leqslant C \qquad \text{for } \bar{\eta}_\lambda - \text{a.e. } \bar{\gamma}.$$

In particular, $(\bar{\eta}_{\lambda,\varepsilon})$ is tight and the family $(m^{\bar{\eta}_\lambda}(t))$ is relatively compact in $C^0([0,T], \mathcal{P}_2(\mathbb{R}^{2d}))$.

Proof. We have, by Lemmata 3.3 and 3.6, for any $\varepsilon > 0$ and $\lambda \geqslant 32\varepsilon^{-2}$, and for $\bar{\eta}_\lambda$–a.e. γ and a.e. $t \in [0, T-\varepsilon]$,

$$\begin{aligned}|\ddot{\bar{\gamma}}(t)| &\leqslant 2C_1(\lambda^{1/2}u_\lambda^{1/2}(\bar{\gamma}(t), \dot{\bar{\gamma}}(t), t) + \lambda\varepsilon u_\lambda(\bar{\gamma}(t), \dot{\bar{\gamma}}(t), t)) \\ &\leqslant C(1 + |\dot{\bar{\gamma}}(t)| + M_{2,v}^{1/2}(m^{\bar{\eta}_\lambda}(t))) + C\varepsilon(1 + |\dot{\bar{\gamma}}(t)|^2 + M_{2,v}(m^{\bar{\eta}_\lambda}(t))). \qquad (46)\end{aligned}$$

For $t \in [0, T-\varepsilon)$, let us set

$$R_\lambda(t) = \inf\{r > 0,\ \mathrm{Spt}(m^{\bar{\eta}_\lambda}(t)) \subset \mathbb{R}^d \times B_r\},$$

(with the convention $R_\lambda(t) = +\infty$ if there is no $r > 0$ such that $\mathrm{Spt}(m^{\bar{\eta}_\lambda}(t)) \subset \mathbb{R}^d \times B_r$). As a preliminary step, we first show that R_λ is finite on a maximal time interval $[0, \tau_\lambda)$, with $\tau_\lambda > 0$, with either $\tau_\lambda = T - \varepsilon$ or $\lim_{t\to\tau_\lambda^-} R_\lambda(t) = +\infty$. For the proof of this fact, $\lambda \geqslant 32\varepsilon^{-2}$ is fixed and all constants depend on λ unless specified otherwise. By (42) and (46), we have, for $0 \leqslant s \leqslant t \leqslant T-\varepsilon$ and $\bar{\eta}_\lambda$–a.e. $\bar{\gamma}$,

$$\begin{aligned}|\ddot{\bar{\gamma}}(t)| &\leqslant C(1 + |\dot{\bar{\gamma}}(t)| + \lambda^{-1/2}e^{\lambda(t-s)/2}M_{2,v}^{1/2}(m^{\bar{\eta}_\lambda}(s)))) \\ &\quad + C\varepsilon(1 + |\dot{\bar{\gamma}}(t)|^2 + \lambda^{-1}e^{\lambda(t-s)}M_{2,v}(m^{\bar{\eta}_\lambda}(s))).\end{aligned}$$

Then, as $M_{2,v}(m^{\bar{\eta}_\lambda}(s)) \leqslant CR_\lambda^2(s)$ for some constant C depending on dimension only,

$$\begin{aligned}|\ddot{\bar{\gamma}}(t)| &\leqslant C(1 + |\dot{\bar{\gamma}}(t)| + \lambda^{-1/2}e^{\lambda(t-s)/2}R_\lambda(s))) \\ &\quad + C\varepsilon(1 + |\dot{\bar{\gamma}}(t)|^2 + \lambda^{-1}e^{\lambda(t-s)}R_\lambda^2(s)). \qquad (47)\end{aligned}$$

So, if $R_\lambda(s)$ is finite for some s and $\lambda \geqslant 1$, $\varepsilon \leqslant 1$, one can find K depending only on $R_\lambda(s)$ and the constant C in (47) such that

$$|\dot{\bar{\gamma}}(t)| \leqslant |\dot{\bar{\gamma}}(s)| + K\int_0^t (1 + |\dot{\bar{\gamma}}(\tau)| + \varepsilon|\dot{\bar{\gamma}}(\tau)|^2)d\tau.$$

Then we can compare $|\dot{\bar{\gamma}}(t)|$ with the solution of the ODE

$$\dot{\phi} = K(1 + \phi + \varepsilon\phi^2), \qquad \phi(s) = |\dot{\bar{\gamma}}(s)|,$$

which is given by

$$\phi(t) = \Phi_\varepsilon^{-1}\Big(\Phi_\varepsilon(|\dot{\bar{\gamma}}(s)|) + K(t-s)\Big),$$

where
$$\Phi_\varepsilon(r) = \int_0^r \frac{1}{1+\tau+\varepsilon\tau^2}d\tau.$$
So one can find $\varepsilon_0, \sigma > 0$ depending only on K such that, for all $\varepsilon \in (0, \varepsilon_0]$,
$$|\dot{\bar\gamma}(t)| \leqslant \phi(t) \leqslant R_\lambda(s) + 1\,, \quad \forall t \in [s, s+\sigma],$$
for any $\bar\gamma \in H^2$ satisfying (47) and $|\dot{\bar\gamma}(s)| \leqslant R_\lambda(s)$. As, by definition of R_λ, $m^{\bar\eta_\lambda}(s)$ has a support contained in $\mathbb{R}^d \times B_{R_\lambda(s)}$, this shows that $m^{\bar\eta_\lambda}(t)$ has a support contained in $\mathbb{R}^d \times B_{R_\lambda(s)+1}$ for any $t \in [t, t+\sigma]$. In particular, as m_0 has a compact support, $R_\lambda(0)$ is finite and thus $R_\lambda(t)$ is finite at least on a small time interval $[0, \sigma]$ for some $\sigma > 0$. We denote by $[0, \tau_\lambda)$ the maximal time interval on which R_λ is finite. Let us assume that $\tau_\lambda < T - \varepsilon$. Let $t_n \to \tau_\lambda^-$. If $(R_\lambda(t_n))$ remains bounded by a constant M, then by the above argument R_λ is bounded by $M+1$ on $[\tau_\lambda, \tau_\lambda + \sigma]$ for some $\sigma > 0$ (depending on M), which contradicts the definition of τ_λ. Hence $\lim_{t\to\tau_\lambda^-} R_\lambda(t) = +\infty$. So we have proved that R_λ is finite on a maximal time interval $[0, \tau_\lambda)$, with $\tau_\lambda > 0$, with either $\tau_\lambda = T - \varepsilon$ or $\lim_{t\to\tau_\lambda^-} R_\lambda(t) = +\infty$.

We now prove the bound on $\ddot{\bar\gamma}$. By definition of $m^{\bar\eta_\lambda}(t)$, for any $\delta > 0$ and $t \in [0, \tau_\lambda)$ there exists $\bar\gamma \in \Gamma$ in the support of $\bar\eta_\lambda$ such that $|\dot{\bar\gamma}(t)| \geqslant R_\lambda(t) - \delta$. Thus
$$R_\lambda(t) - \delta \leqslant |\dot{\bar\gamma}(t)| \leqslant |\dot{\bar\gamma}(0)| + \int_0^t |\ddot{\bar\gamma}(s)|ds.$$
As $(\bar\gamma(t), \dot{\bar\gamma}(t))$ belongs to the support of $m^{\bar\eta_\lambda}(t)$ for any t, we get by (46) and the definition of R_λ:
$$\begin{aligned} R_\lambda(t) - \delta &\leqslant |\dot{\bar\gamma}(0)| + C\int_0^t (1 + |\dot{\bar\gamma}(s)| + M_{2,v}^{1/2}(m^{\bar\eta_\lambda}(s)))ds \\ &\quad + C\varepsilon \int_0^t (1 + |\dot{\bar\gamma}(s)|^2 + M_{2,v}(m^{\bar\eta_\lambda}(s)))ds \\ &\leqslant R_0 + C\int_0^t (1 + R_\lambda(s) + \varepsilon R_\lambda^2(s))ds. \end{aligned}$$
As δ is arbitrary, this proves that
$$R_\lambda(t) \leqslant R_0 + C\int_0^t (1 + R_\lambda(s) + \varepsilon R_\lambda^2(s))ds \qquad \forall t \in [0, \tau_\lambda).$$
Arguing as above we get
$$R_\lambda(t) \leqslant \Phi_\varepsilon^{-1}\Big(\Phi_\varepsilon(R_0) + Ct\Big).$$
For all $\varepsilon > 0$ small enough (but independent of λ) and $\lambda \geqslant (32\varepsilon^{-2}) \vee 1$, we have therefore that R_λ is bounded by a constant C independent of λ on $[0, \tau_\lambda)$. Thus $\tau_\lambda = T - \varepsilon$ and R_λ is bounded by C on $[0, T - \varepsilon]$.

This estimate gives immediately the bound on $|\dot{\bar\gamma}|$ and therefore, by (46), the bound on $|\ddot{\bar\gamma}|$ for $\bar\eta_\lambda$−a.e. $\bar\gamma$. As m_0 has a compact support, this also implies that the $m_\lambda(t)$ have a support contained in a ball B_C, where C is independent of λ and t. In addition the sequence $\bar\eta_{\lambda,\varepsilon}$ is tight.

Finally, we have, for any $0 \leqslant s \leqslant t \leqslant T - \varepsilon$,
$$\begin{aligned} \mathbf{d}_1(m^{\bar\eta_\lambda}(s), m^{\bar\eta_\lambda}(t)) &= \int_\Gamma (|\bar\gamma(t) - \bar\gamma(s)|^2 + |\dot{\bar\gamma}(t) - \dot{\bar\gamma}(s)|^2)^{1/2} \bar\eta_\lambda(d\bar\gamma) \\ &\leqslant C(t-s)^{1/2} \int_\Gamma \Big(\int_s^t |\ddot{\bar\gamma}(\tau)|^2 d\tau\Big)^{1/2} \bar\eta_\lambda(d\bar\gamma) \leqslant C(t-s)^{1/2}. \end{aligned}$$
As the $(m^{\bar\eta_\lambda}(t))$ have a support which is uniformly bounded, this shows that it is a relatively compact sequence in $C^0([0, T-\varepsilon], \mathcal{P}_2(\mathbb{R}^d))$. ■

We are now ready to prove the main result:

Proof of Theorem 3.2. In view of Lemma 3.7, the family $(m^{\bar\eta_\lambda})$ is relatively compact in $C^0_{loc}([0,T),\mathcal{P}_2(\mathbb{R}^d))$: let m be any limit, up to a subsequence, of $(m^{\bar\eta_\lambda})$. We have to prove that m is a measure valued solution to the kinetic equation (30). For this, still by Lemma 3.7 and a diagonal argument, we can find a subsequence $\lambda_n \to +\infty$ such that, for any $\varepsilon > 0$, $(\bar\eta_{\lambda_n,\varepsilon})$ converges weakly to the restriction to $[0,T-\varepsilon]$ of some η in $\mathcal{P}(C^1([0,T),\mathbb{R}^d))$ with $m(t) = \tilde e_t \sharp \eta$ for any $t \in [0,T)$.

We now identify the $\limsup(\mathrm{Spt}(\bar\eta_{\lambda_n}))$. Let us recall that, by Lemma 3.1, for $\bar\eta_\lambda$−a.e. $\bar\gamma_\lambda$, $\bar\gamma_\lambda$ minimizes problem (33). Hence by the Euler equation we have that $\bar\gamma_\lambda$ is of class H^4 and for a.e. $t \in [0,T]$,

$$\frac{d^2}{dt^2}(\lambda^{-1}e^{-\lambda t}\ddot{\bar\gamma}_\lambda(t)) = \frac{d}{dt}\left(e^{-\lambda t}D_vF(\bar\gamma_\lambda(t),\dot{\bar\gamma}_\lambda(t),m^{\bar\eta_\lambda}(t))\right) - e^{-\lambda t}D_xF(\bar\gamma_\lambda(t),\dot{\bar\gamma}_\lambda(t),m^{\bar\eta_\lambda}(t)).$$

We rewrite this equality as

$$\ddot{\bar\gamma}_\lambda(t) + D_vF(\bar\gamma_\lambda(t),\dot{\bar\gamma}_\lambda(t),m^{\bar\eta_\lambda}(t)) = \lambda^{-1}\Big(-\lambda^{-1}\bar\gamma_\lambda^{(iv)}(t) + 2\dddot{\bar\gamma}_\lambda(t) + \frac{d}{dt}D_vF(\bar\gamma_\lambda(t),\dot{\bar\gamma}_\lambda(t),m^{\bar\eta_\lambda}(t)) - D_xF(\bar\gamma_\lambda(t),\dot{\bar\gamma}_\lambda(t),m^{\bar\eta_\lambda}(t))\Big).$$

We integrate this equation by parts against a test function $z \in C_c^\infty((0,T),\mathbb{R}^d)$ to get

$$\int_0^T \Big(-\dot{\bar\gamma}_\lambda(t)\cdot\dot z(t) + D_vF(\bar\gamma_\lambda(t),\dot{\bar\gamma}_\lambda(t),m^{\bar\eta_\lambda}(t))\cdot z(t)\Big)dt$$
$$= \lambda^{-1}\int_0^T\Big(\lambda^{-1}\dot{\bar\gamma}_\lambda(t)\cdot\dddot z(t) + 2\dot{\bar\gamma}_\lambda(t)\cdot\ddot z(t) - D_vF(\bar\gamma_\lambda(t),\dot{\bar\gamma}_\lambda(t),m^{\bar\eta_\lambda}(t))\cdot\dot z(t) - D_xF(\bar\gamma_\lambda(t),\dot{\bar\gamma}_\lambda(t),m^{\bar\eta_\lambda}(t))\cdot z(t)\Big)dt.$$

By Lemma 3.7 $(\bar\gamma_\lambda)$ is relatively compact in $C^1_{loc}([0,T),\mathbb{R}^d)$, and for any sequence $\lambda_n \to +\infty$ we can extract a subsequence such that $\bar\gamma_{\lambda_n} \to \gamma \in C^1_{loc}([0,T),\mathbb{R}^d)$. As $m^{\bar\eta_{\lambda_n}} \to m \in C^0_{loc}([0,T),\mathcal{P}_2(\mathbb{R}^{2d}))$, we have therefore

$$\int_0^T\Big(-\dot\gamma(t)\cdot\dot z(t) + D_vF(\gamma(t),\dot\gamma(t),m(t))\cdot z(t)\Big)dt = 0, \qquad \forall z \in C_c^\infty((0,T),\mathbb{R}^d),$$

which means that it is a solution to

$$\ddot\gamma(t) = -D_vF(\gamma(t),\dot\gamma(t),m(t)).$$

In other words, $(\gamma(t),\dot\gamma(t)) = P^{x,v}(t)$, where P is defined by (34) and $(x,v) = (\gamma(0),\dot\gamma(0))$. So we have proved that the support of η consists of solutions to (34). As $\tilde e_0\sharp\eta = m_0$, we have

$$\eta = \int_{\mathbb{R}^{2d}} \delta_{P^{x,v}}m_0(dx,dv),$$

so that

$$m(t) = \tilde e_t\sharp\eta = P^{x,v}(t)\sharp m_0.$$

Hence m is the measure-valued solution to (30). Following [10] this solution is unique. We have proved therefore that any converging subsequence of the relatively compact familiy $(m^{\bar\eta_\lambda})$ has for limit the unique solution m to (30): the entire sequence converges. ■

Remark 3.4. The Cucker-Smale model is usually associated to the collective animal behaviour, such as flocking of birds or swarming of insects, see [23, 25, 19]. More recent models of consensus are reviewed in [18] and we believe that some of them can be treated by our methods. Similar models where the acceleration of the agents is prescribed have been proposed for describing the dynamics of crowds of pedestrians, and some of them fit in our results. We refer to the book [22], in particular the section on mesoscopic or kinetic models, and to the recent survey paper [32], where they are called social forces models.

References

[1] Y. Achdou, P. Mannucci, C. Marchi, N. Tchou: Deterministic mean field games with control on the acceleration. NoDEA Nonlinear Differential Equations Appl. 27 (2020), Paper No. 33, 32 pp.

[2] L. Ambrosio, N. Gigli, G. Savaré: Gradient flows: in metric spaces and in the space of probability measures. Second edition. Birkhäuser Verlag, Basel, 2008.

[3] M. Bardi, I. Capuzzo-Dolcetta: Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations. Birkhäuser Boston, Inc., Boston, MA, 1997.

[4] M. Barker: From mean field games to the best reply strategy in a stochastic framework. J. Dyn. Games 6 (2019), 291-314.

[5] M. Barker, P. Degond, M.T. Wolfram: Comparing the best reply strategy and mean field games: the stationary case (2019). arXiv preprint arXiv:1911.04300.

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

[7] A.J. Bernoff, C.M. Topaz: Nonlocal aggregation models: a primer of swarm equilibria. SIAM Rev. 55 (2013), 709–747

[8] C. Bertucci, P.-L. Lions, J.-M. Lasry: Some remarks on Mean Field Games, Comm. Partial Differential Equations 44 (2019), 205–227.

[9] M. Bodnar, J.J.L. Velazquez: An integro-differential equation arising as a limit of individual cell-based models. J. Differential Equations 222 (2006), 341–380.

[10] J.A. Canizo, J.A. Carrillo, J. Rosado: A well-posedness theory in measures for some kinetic models of collective motion. Mathematical Models and Methods in Applied Sciences, 21(2011), 515-539.

[11] P. Cannarsa, C. Sinestrari: Semiconcave functions, Hamilton-Jacobi equations, and optimal control (Vol. 58). Springer Science & Business Media, 2004.

[12] P. Cannarsa, C. Mendico: Mild and weak solutions of Mean Field Games problem for linear control systems. Minimax Theory Appl. 5 (2020), 221-250.

[13] P. Cardaliaguet, Notes on mean field games. Technical report.

[14] P. Cardaliaguet: Weak solutions for first order mean field games with local coupling. In Analysis and geometry in control theory and its applications (pp. 111-158). Springer, Cham, 2015.

[15] P. Cardaliaguet, P.J. Graber: Mean field games systems of first order. ESAIM: Control, Optimisation and Calculus of Variations, 21 (2015), 690-722.

[16] P. Cardaliaguet, S. Hadikhanloo: Learning in mean field games: the fictitious play. ESAIM: Control, Optimisation and Calculus of Variations, 23 (2017), 569-591.

[17] P. Cardaliaguet, A.R. Mészàros, F, Santambrogio: First order mean field games with density constraints: pressure equals price. SIAM Journal on Control and Optimization, 54 (2016), 2672-2709.

[18] J.A. Carrillo, Y.-P. Choi, S.P. Perez: A review on attractive-repulsive hydrodynamics for consensus in collective behavior. Active particles. Vol. 1. Advances in theory, models, and applications, 259–298, Birkhäuser/Springer, Cham, 2017.

[19] J.A. Carrillo, M. Fornasier, J. Rosado, G. Toscani: Asymptotic flocking dynamics for the kinetic Cucker-Smale model. SIAM J. Math. Anal. 42 (2010), 218–236.

[20] J.A. Carrillo, J. Rosado: Uniqueness of bounded solutions to aggregation equations by optimal transport methods. European Congress of Mathematics, 3-16, Eur. Math. Soc., Zürich, 2010.

[21] E. Cristiani, B. Piccoli, A. Tosin: Multiscale modeling of granular flows with application to crowd dynamics. Multiscale Model. Simul. 9 (2011), 155-182.

[22] E. Cristiani, B. Piccoli, A. Tosin: Multiscale modeling of pedestrian dynamics. Springer, Cham, 2014.

[23] F. Cucker, S. Smale: On the mathematics of emergence. Jpn. J. Math. 2 (2007), 197–227.

[24] P. Degond, M. Herty, J.G. Liu: Mean field games and model predictive control. Commun. Math. Sci. 15 (2017), 1403–1422.

[25] S.-Y. Ha, E. Tadmor: From particle to kinetic and hydrodynamic descriptions of flocking. Kinet. Relat. Models 1 (2008), 415–435.

[26] G.E. Fasshauer, Meshfree Approximations Methods with Matlab, World Scientific, Singapore, 2007.

[27] M. Huang, R.P. Malhamé, P. E. Caines: Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. Commun. Inf. Syst. 6 (2006), 221-252.

[28] J.-M. Lasry and P.-L. Lions, Mean field games, Jpn. J. Math. 2 (2007), 229–260.

[29] P.-L. Lions:. Cours au Collège de France (2010). Available at www.college-de-france.fr.

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

[31] B. Piccoli, F. Rossi, Transport equation with nonlocal velocity in Wasserstein spaces: convergence of numerical schemes. Acta Appl. Math. 124 (2013), 73-105.

[32] B. Piccoli, F. Rossi, Measure-theoretic models for crowd dynamics in "Crowd Dynamics Volume 1 - Theory, Models, and Safety Problems", N. Bellomo and L. Gibelli Eds, Birkhauser, 2018.

[33] R. Rossi, G. Savaré, A. Segatti, U. Stefanelli: Weighted Energy-Dissipation principle for gradient flows in metric spaces. J. Math. Pures Appl. 127 (2019), 1-66.

[34] C.M. Topaz, A.L. Bertozzi, M.A. Lewis: A nonlocal continuum model for biological aggregation. Bull. Math. Biol. 68 (2006), 1601–1623.