



HAL
open science

On the Optimal Rate for the Convergence Problem in Mean Field Control

Samuel Daudin, François Delarue, Joe Jackson

► **To cite this version:**

Samuel Daudin, François Delarue, Joe Jackson. On the Optimal Rate for the Convergence Problem in Mean Field Control. 2024. hal-04395179

HAL Id: hal-04395179

<https://hal.science/hal-04395179>

Preprint submitted on 15 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ON THE OPTIMAL RATE FOR THE CONVERGENCE PROBLEM IN MEAN FIELD CONTROL

SAMUEL DAUDIN, FRANÇOIS DELARUE, AND JOE JACKSON

ABSTRACT. The goal of this work is to obtain optimal rates for the convergence problem in mean field control. Our analysis covers cases where the solutions to the limiting problem may not be unique nor stable. Equivalently the value function of the limiting problem might not be differentiable on the entire space. Our main result is then to derive sharp rates of convergence in two distinct regimes. When the data is sufficiently regular, we obtain rates proportional to $N^{-1/2}$, with N being the number of particles. When the data is merely Lipschitz and semi-concave with respect to the first Wasserstein distance, we obtain rates proportional to $N^{-2/(3d+6)}$. Noticeably, the exponent $2/(3d+6)$ is close to $1/d$, which is the optimal rate of convergence for uncontrolled particle systems driven by data with a similar regularity. The key argument in our approach consists in mollifying the value function of the limiting problem in order to produce functions that are almost classical sub-solutions to the limiting Hamilton-Jacobi equation (which is a PDE set on the space of probability measures). These sub-solutions can be projected onto finite dimensional spaces and then compared with the value functions associated with the particle systems. In the end, this comparison is used to prove the most demanding bound in the estimates. The key challenge therein is thus to exhibit an appropriate form of mollification. We do so by employing sup-convolution within a convenient functional Hilbert space. To make the whole easier, we limit ourselves to the periodic setting. We also provide some examples to show that our results are sharp up to some extent.

KEYWORDS: Mean Field Control; Convergence; Hamilton-Jacobi equation; Viscosity Solutions; Sup-convolution.

AMS CLASSIFICATION (2020): Primary: 49N80, 65C35; Secondary: 49L35.

CONTENTS

1. Introduction	2
2. Preliminaries and main results	10
3. Properties of the value function	17
4. Three regularization procedures	23
5. Proofs of the “hard inequalities”	30
6. Proofs of the “easy inequalities”	46
7. Proofs for the examples and the convex case	52
Appendix A. Some auxiliary estimates for finite-dimensional PDEs	57
Appendix B. A mollification argument	61
References	63

S. Daudin and F. Delarue acknowledge the financial support of the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (ELISA project, Grant agreement No. 101054746). Part of the work was achieved during J. Jackson’s visit to Université Côte d’Azur between January and March 2023. His visit was also supported by the ERC Grant No. 101054746. J. Jackson is supported by the NSF under Grant No. DGE1610403. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

1. INTRODUCTION

1.1. A short review of mean field control and games. Mean field control theory and its twin, mean field game theory, aim at the asymptotic study of equilibria within large populations of weakly interacting agents. Typically, each agent controls a d -dimensional state process which is impacted by a Brownian noise. In mean field control, equilibria are understood in a cooperative sense, while in mean field games, they are understood in a competitive sense. The limiting formulations, which arise as the number of players increases to infinity, are distinct: in the cooperative case, we arrive at an optimal control problem set on the Wasserstein space, while the competitive case leads to a well-known fixed point problem. We refer to [HPM03, HMC06, LL06a, LL06b, LL07, Lioa] for earlier contributions and [CP21, CD18b, CD18a, GS14] for surveys or monographs.

For almost twenty years, both theories have made parallel and profound advances. In particular, great progress has been made in understanding the infinite-dimensional partial differential equations which describe the relevant value functions - the value of the optimization problem in the case of control, and the value of the equilibrium in the case of games. We refer to [BFY15, CDLL19, GS15, Liob] for some key contributions in this direction. In the case of mean field control (with periodic data), the value function is a map

$$U = U(t, m) : [0, T] \times \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R},$$

where $\mathcal{P}(\mathbb{T}^d)$ denotes the set of probability measures on the d -dimensional torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. Roughly speaking, $U(t, m)$ denotes the value of the limiting optimization problem when the continuum of agents is distributed according to m at the initial time t . We postpone a definition of U as a value function to Subsection 2.2, but we mention already that U is expected to solve a first-order Hamilton-Jacobi equation on the space of probability measures, of the form

$$\begin{cases} -\partial_t U(t, m) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta U}{\delta m}(t, m, x) dm(x) \\ \quad + \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta U}{\delta m}(t, m, x)\right) dm(x) = \mathcal{F}(m), & (t, m) \in (0, T) \times \mathcal{P}(\mathbb{T}^d), \\ U(T, m) = \mathcal{G}(m), & m \in \mathcal{P}(\mathbb{T}^d), \end{cases} \quad (\text{HJB}(\infty))$$

for a Hamiltonian $H = H(x, p) : \mathbb{T}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ which is typically regular and convex in the second variable, and costs $\mathcal{F}, \mathcal{G} : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$. We refer to Section 2 for more details including the definition of the linear derivative $\frac{\delta U}{\delta m}$. For games, the value of the equilibrium problem is instead expected to solve the master equation, which resembles an infinite-dimensional system of first-order hyperbolic equations on the space of probability measures. In both cases, the study of the solution is subtle, but several regimes are known under which the relevant infinite-dimensional PDE admits a classical solution. For control, this is the case if the coefficients are convex in the measure argument and regular, see [CDLL19, CCDar, GS15]. For games, the convexity condition has to be replaced by a monotonicity condition, see [Ber21, CDLL19, CD18a, CCDar, Liob].

The importance of the regularity of the value function was explained in the book [CDLL19] by Cardaliaguet, Delarue, Lasry and Lions: the existence of a regular value makes it possible to obtain an optimal rate for the convergence of the values of finite-player games towards the value of their mean field counterparts. In this approach, the bounds on the regularity of the value play an essential role, and the resulting convergence rate is linear in the number N of agents in the finite system. The approach used for games has subsequently been extended to mean field control when the solution of the Hamilton-Jacobi equation is regular, with a rate of the same order, see [GPW22]. In particular, in the case of control, bounds on the second-order ‘Lions derivative’ $D_{mm}^2 U$ allow one to conclude that

$$|U(t, m_x^N) - V^N(t, \mathbf{x})| \leq C/N, \quad (1.1)$$

where $V^N : [0, T] \times (\mathbb{T}^d)^N \rightarrow \mathbb{R}$ denotes the value function for the corresponding N -particle control problem, and is, under mild assumptions on the data, the unique classical solution of the Hamilton-Jacobi-Bellman equation

$$\begin{cases} -\partial_t V^N(t, \mathbf{x}) - \sum_{i=1}^N \Delta_{x^i} V^N(t, \mathbf{x}) + \frac{1}{N} \sum_{i=1}^N H(x^i, N D_{x^i} V^N(t, \mathbf{x})) = \mathcal{F}(m_{\mathbf{x}}^N), \\ V^N(T, \mathbf{x}) = \mathcal{G}(m_{\mathbf{x}}^N), \quad \mathbf{x} \in (\mathbb{T}^d)^N. \end{cases} \quad (t, \mathbf{x}) \in (0, T) \times (\mathbb{T}^d)^N, \quad (\text{HJB}(N))$$

We note that here and throughout the paper we use the notation $m_{\mathbf{x}}^N = \frac{1}{N} \sum_{i=1}^N \delta_{x^i}$ when $\mathbf{x} = (x^1, \dots, x^N) \in (\mathbb{T}^d)^N$. The argument leading to (1.1) is relatively simple - if U is smooth, then explicit computation shows that $[0, T] \times (\mathbb{T}^d) \ni (t, \mathbf{x}) \mapsto U(t, m_{\mathbf{x}}^N)$ is a solution of (HJB(N)) up to an error term which is of order $1/N$ provided that $D_{mm}U$ is bounded, which, by the comparison principle, gives (1.1). We note that here and throughout the paper we use the notation $m_{\mathbf{x}}^N = \frac{1}{N} \sum_{i=1}^N \delta_{x^i}$ when $\mathbf{x} = (x^1, \dots, x^N) \in (\mathbb{T}^d)^N$. The argument leading to (1.1) is relatively simple - if U is smooth, then explicit computation shows that $[0, T] \times (\mathbb{T}^d) \ni (t, \mathbf{x}) \mapsto U(t, m_{\mathbf{x}}^N)$ is a solution of (HJB(N)) up to an error term which is of order $1/N$ provided that $D_{mm}^2 U$ is bounded, which, by the comparison principle, gives (1.1).

Thanks to the contributions discussed above, the convergence problem is now well-understood when the relevant infinite-dimensional PDE has a smooth solution, and the existence of smooth solutions in turn is well-understood under certain (fairly restrictive) convexity or monotonicity assumptions. Answering similar questions in the absence of structural conditions like convexity and monotonicity is now one of the main objectives of the theory of mean field games and mean field control. This issue is somewhat easier to understand for control problems than games, simply because it is easier to identify and characterize an optimizer than a fixed point. In fact, several recent works have been published on the Hamilton-Jacobi equation (HJB(∞)) in the absence of classical solutions. Most of them aim to understand viscosity solutions and, in particular, to obtain a comparison principle allowing the identification of the value of the mean field control problem as the unique viscosity solution in a class and sense as broad as possible, see [BIRS20, CKT21, CGK⁺21, SY22, WZ20]. Typically, these results cover the setting where U is Lipschitz but may not be differentiable, which is expected to be the case when \mathcal{F} and \mathcal{G} are regular but not convex and thus optimizers may not be unique. In comparison, there are much fewer general results on the master equation of mean field games in the absence of uniqueness of the equilibria: the work of [CD22] gives a possible approach in the case of potential games, which are, by definition, derived from a control problem. At this stage, there are no general results on the convergence of the value functions of finite games to a possible value of the mean field game outside the analysis of [CDLL19]. The best that is known are compactness results, see for instance [Dje22, Fis17, Lac20]. They suffice to establish convergence of the value functions at measures where the equilibrium is unique, but questions of selection remain very challenging when uniqueness does not hold.

1.2. Our motivation. The goal of the present paper is to understand the rate of convergence of the value functions V^N to U in the non-convex setting. Of course, as already discussed above, a quantitative answer to this convergence problem is already known when \mathcal{F} and \mathcal{G} are convex and sufficiently regular, with the (optimal) rate $1/N$. The convergence of V^N to U in the non-convex setting has received significant attention in the literature in recent years, and qualitative results have been obtained in [Lac17, DPT22] (see also [CLOS22, FLOS19, GMS21] for deterministic dynamics or dynamics with a sole common noise). More recently, a first quantitative result outside the convex setting has been obtained, under ‘natural’ assumptions, in a work [CDJS23] by the first and last author with Cardaliaguet and Souganidis. The main result of [CDJS23] (when specialized

to the periodic setting) is the estimate

$$|U(t, m_{\mathbf{x}}^N) - V^N(t, \mathbf{x})| \leq CN^{-\gamma(d)}, \quad (1.2)$$

with C independent of N and $\gamma(d)$ depending only on the dimension d . The value of the exponent $\gamma(d)$ is not given explicitly, but the calculations can be followed step by step, and it is clear that γ decreases faster than (any multiple) of $1/d$.

The result of [CDJS23] shows that one can indeed have an algebraic convergence rate of convergence even when optimizers of the limiting problem are not unique. But it leaves open a very natural question, which we aim to investigate in this work:

What is the optimal rate of convergence of V^N to U in the non-convex setting?

As far as we can tell, before the present paper there was not a clear conjecture about what the optimal rate should be, and in particular whether (or under what circumstances) it should be possible to obtain a dimension-free rate. Let us first emphasize that without convexity we cannot hope to obtain the rate $1/N$, as the calculations leading to (1.1) in the smooth case clearly indicate that the rate $1/N$ is tied to second-order regularity of U in the measure variable, which we cannot expect without convexity of \mathcal{F} and \mathcal{G} . To gain some intuition, notice that because U is only Lipschitz, the rate should be compared to the one for the uncontrolled case with data which is only Lipschitz continuous, and this should in turn be related to the rate observed in the convergence of uncontrolled weakly interacting particle systems. Due to the underlying statistical averaging phenomena, the latter convergence rate is actually related to fundamental results in probability theory on the convergence of empirical measures of an N -sample. In this context, there are two rates of convergence which play an especially important role:

- (a) the rate $N^{-1/d}$ governs convergence of empirical measures for the Kantorovich-Rubinstein distance (also called the 1-Wasserstein distance and denoted d_1) when¹ $d \geq 3$, see [AKT84, DSS13, FG15], and also describes the typical minimal distance, in dimension d , between two particles within a cloud of uniformly drawn particles.
- (b) The rate $N^{-1/2}$ corresponds to the central limit theorem, whose transposition to the convergence of empirical measures nevertheless requires some precautions, see [FM97, JM98, Mel96, Szn85, TH81].

The difference between (a) and (b) lies in the class of test functions used to measure the convergence rate of the empirical measure: in (a), the test functions are Lipschitz continuous, while in (b), the test functions are much smoother. To illustrate point (b), it is worth observing that the main fluctuation results in the literature on particle systems are stated in ‘‘sufficiently negative’’ Sobolev (Hilbertian) spaces. For instance, in [Mel96, FM97], fluctuations are estimated in the dual of a space of functions admitting $s := 1 + 2\lfloor d/2 \rfloor$ generalized derivatives that are square integrable with respect to some heavy (polynomially) tailed measure ν . Convergence of the fluctuations is obtained in a similar but larger space, obtained by replacing $s = 1 + 2\lfloor d/2 \rfloor$ by $s = 4 + 2\lfloor d/2 \rfloor$ and by changing accordingly the polynomial decay of the underlying reference measure ν . As we are working on the torus, the description of ν does not really matter here. Still, it is worth stressing that the mollification procedures implemented in the present paper also rely on the properties of the Hilbert space H^{-s} for $s > d/2 + 1$, i.e., the dual of the Hilbert space of functions with s generalized derivatives in L^2 , and could be used to recover the fact (hence already proven in [Mel96, FM97]) that fluctuations are on average of order $N^{-1/2}$ when measured in the Hilbert space H^{-s} for $s > d/2 + 1$.

In other words, if $(\xi^i)_{i=1,\dots,N}$ are i.i.d random variables with common law m , then

¹ The exponent becomes $1/2$ when $d = 1, 2$, with an additional logarithmic correction in the rate of convergence when $d = 2$, but we feel better to stick to the reference value $1/d$ throughout the introduction as it makes the presentation easier.

- $d_1(m_{\xi}^N, m) := \sup_{g \text{ 1-Lip}} \int g d(m_{\xi}^N - m)$ is typically of order $N^{-1/d}$, while
- $\|m_{\xi}^N - m\|_{-s} := \sup_{\|g\|_s=1} \int g d(m_{\xi}^N - m)$ is typically of order $N^{-1/2}$ when $s > d/2 + 1$.

The heuristic discussion above suggests the following conjecture: if we work under conditions on the data (\mathcal{F} , \mathcal{G} , and H) which guarantee only that U is Lipschitz with respect to d_1 , then the optimal rate should be $N^{-1/d}$, the size of typical fluctuations of empirical measures as measured with respect to d_1 . If, on the other hand, the value function U is Lipschitz with respect to a much weaker metric, like the one generated by $\|\cdot\|_{-s}$ for s large enough, then it should be possible to obtain the rate $N^{-1/2}$, the size of typical fluctuations of empirical measures as measured with respect to this weaker metric. Our objective is to verify this conjecture as far as possible. In order to do so, we limit our analysis to the periodic setting: this avoids any technicalities about the decay at infinity of the various functions that we manipulate. We also emphasize that this question has already been solved for mean field control problems on a finite state space, see [Cec21, Kol12]: the convergence rate is shown to be $1/\sqrt{N}$ when cost coefficients are non-convex in the measure argument, which is consistent with case (b) right above, keeping in mind that the regularity of the test functions does not matter in this case since the state space is finite.

1.3. Our results. We work with two sets of conditions of the data H , \mathcal{F} , and \mathcal{G} . Assumption 2.1 gives minimal conditions under which we can establish that U is Lipschitz and semi-concave (defined below) with respect to d_1 , and when Assumption 2.1 is in force we say that we are in the “ d_1 -regular case”. Assumption 2.4 gives minimal conditions under which we can establish that U is Lipschitz and semi-concave (defined below) with respect to $\|\cdot\|_{-s}$ for some $s > d/2 + 2$, and when Assumption 2.4 is in force we say that we are in the “ H^{-s} -regular case”. The role of semi-concavity in both cases is outlined in Subsection 1.4. We also refer to Remark 2.3.

Our contributions in the two “ H^{-s} and d_1 -regular cases” can be summarized as follows.

Rates of convergence: In the H^{-s} -regular case, we obtain in Theorem 2.7 the estimate

$$V^N(t, \mathbf{x}) - C/N \leq U(t, m_{\mathbf{x}}^N) \leq V^N(t, \mathbf{x}) + C/\sqrt{N}, \quad (1.3)$$

and in particular $|V^N(t, \mathbf{x}) - U(t, m_{\mathbf{x}}^N)| \leq C/\sqrt{N}$. This exactly matches the conjectured rate in the H^{-s} -regular case.

In the d_1 -regular case, we find (again taking $d \geq 3$ for simplicity and recalling footnote 1 for the peculiar cases $d = 1, 2$) that, for each $\eta > 0$, there is a constant $C > 0$ such that

$$V^N(t, \mathbf{x}) - CN^{-1/d} \leq U(t, m_{\mathbf{x}}^N) \leq V^N(t, \mathbf{x}) + CN^{-\beta(d)+\eta}, \quad \beta(d) := \frac{2}{3d+6}. \quad (1.4)$$

In particular, we see that for d large, the estimate on $|V^N(t, \mathbf{x}) - U(t, m_{\mathbf{x}}^N)|$ is roughly of order $N^{-2/(3d)}$. This is obviously slightly worse than the conjectured rate of $N^{-1/d}$, but still represents a significant improvement on existing results.

Finally, in the case where the coefficients are convex and Lipschitz with respect d_1 , we establish the estimate (again taking $d \geq 3$ for simplicity)

$$0 \leq V^N(t, \mathbf{x}) - U(t, m_{\mathbf{x}}^N) \leq CN^{-1/d}. \quad (1.5)$$

Even though the convex case is much simpler to analyze, this result seems to be new. Indeed, since \mathcal{F} and \mathcal{G} are only assumed to be d_1 -Lipschitz, U is not expected to be smooth, and so the well-known argument based on “projecting” U fails. We instead use purely control-theoretic arguments to obtain (1.5). We note in particular that the observation that $V^N \leq U$ in the convex regime appears to be new. In view of Example 1 presented in Subsection 2.4, the rate in Proposition 1.5 is sharp.

Examples: Of course, to provide a complete picture we need also to verify through examples

that the conjectured rates $N^{-1/d}$ and $N^{-1/2}$ cannot be improved. In Subsection 2.4, we first demonstrate in Example 1 that $N^{-1/d}$ is indeed the best possible rate when \mathcal{F} and \mathcal{G} are just d_1 -Lipschitz. We note that it is easy to construct such an example if $H = 0$ (there is no control), but this does not rule out the possibility that strict convexity or coercivity of the Hamiltonian H somehow benefits the convergence rate. Our example uses the “model” Hamiltonian $\frac{1}{2}|p|^2$, and the idea is to use the Cole-Hopf transform to analyze the functions V^N - this ultimately leads to an interesting probabilistic analysis related to the “coupon-collector problem”, which is presented in Section 7. Noticeably, this counter-example works due to the infinite dimensional nature of the optimal control problem that is treated here; to the best of our understanding, similar constructions, but in the Euclidean setting, would not provide interesting examples in the study of vanishing viscosity for finite-dimensional Hamilton-Jacobi equations.

Indeed, for the H^{-s} -regular case, we show (in Example 2) that the convergence problem is in fact related to a question of vanishing viscosity for finite-dimensional Hamilton-Jacobi equations, with the viscosity being of order $1/N$. And, remarkably, there are two distinct situations for this latter problem: (1) when the costs are convex and smooth enough, the convergence rate of the value functions is linear in the viscosity, see [Fle71, FS86]; (2) in general (but under reasonable regularity assumptions), the convergence rate is linear in the square root of the viscosity, see [CL84, Eva10, Lio82], and this rate is claimed to be optimal. As for the latter point, it is however fair to say that the precise conditions under which the root of the viscosity is the optimal rate are rather unclear to us and, in particular, we do not know whether these conditions cover or not the type of Hamiltonians we use below. As we announced in the previous paragraph, the construction based on the same Cole-Hopf transformation as the one used below for proving that $N^{-1/d}$ is indeed the optimal rate in the “ d_1 -regular case” does not provide a relevant example in the vanishing viscosity problem for finite-dimensional Hamilton-Jacobi equations. Anyway, we believe that the rate obtained in (1.3) is the optimal rate in the H^{-s} -regular case.

Regularity: While the convergence problem is our main focus, we also provide some new regularity results for the value function U . In the d_1 -regular case, we establish in Proposition 3.4 that the value function U is Lipschitz and semi-concave with respect to the metric d_1 . This result is expected, except for the fact that we obtain the result without assuming that \mathcal{F} and \mathcal{G} are differentiable. In fact we only assume that \mathcal{F} and \mathcal{G} are Lipschitz and semi-concave with respect to d_1 , so our assumptions on \mathcal{F} and \mathcal{G} in this result are in some sense optimal (certainly we cannot have a global Lipschitz and semi-concavity estimate for U without assuming one for \mathcal{G}). Without differentiability of \mathcal{F} and \mathcal{G} we do not have access to the usual description of optimizers in terms of a forward-backward PDE system, and we must instead proceed by a mollification procedure. We also prove in Proposition 3.2 that in the H^{-s} -regular case, U is Lipschitz and semi-concave with respect to $\|\cdot\|_{-s}$. This result appears to be new, and relies on stability estimates for a certain Fokker-Planck equation in negative Sobolev spaces which are presented in the Appendix.

Regularization procedures on the Wasserstein space: In order to achieve our main convergence and regularity results, we implement several regularization techniques for functions on the space $\mathcal{P}(\mathbb{T}^d)$. All three of these methods have appeared in some way in the literature before, but here we study for the first time their interplay with the Hamilton-Jacobi equation (HJB(∞)). In particular, much of our analysis in Section 5 is focused on analyzing the degree to which the regularization procedures preserve subsolutions of (HJB(∞)). We believe the techniques we employ in this section could be useful, for example in studying the comparison principle for equations like (HJB(∞)), where a procedure for approximating subsolutions by more regular subsolutions would clearly be useful.

1.4. Our method. As discussed above, one of the examples presented in Subsection 2.4 shows that the convergence problem we are considering can be viewed as an infinite-dimensional analogue of an evanescent viscosity problem in which the intensity of the evanescent noise would be $1/\sqrt{N}$. Accordingly our proof, both in the d_1 -regular and H^{-s} -regular cases, takes up essential ideas from the analysis of finite-dimensional Hamilton-Jacobi equations with an evanescent viscosity, as written, for example, in the notes [Cal18].

The upper bounds in both (1.3) and (1.4) are more challenging, and we refer to them as the “hard inequalities”. The basic idea for proving the hard inequality in the H^{-s} -regular case (i.e. the upper bound in (1.3)) is as follows. Let us first suppose that $\Phi = \Phi(t, m) : [0, T] \times \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ is a function such that

- (i) $\Phi(T, m) \leq \mathcal{G}(m) + C_0$,
- (ii) Φ is a subsolution of (HJB(∞)), up to an error of order C_1 ,
- (iii) $\|\text{tr}(D_{mm}^2 \Phi)\|_{L^\infty} \leq C_2$.

Then, at least formally, the function $[0, T] \times (\mathbb{T}^d)^N \ni (t, \mathbf{x}) \mapsto \Phi(t, m_{\mathbf{x}}^N)$ is a subsolution of the equation describing V^N , up to an error of order $C_1 + \frac{C_2}{N}$, so that the comparison principle gives

$$\Phi(t, m_{\mathbf{x}}^N) \leq V^N(t, \mathbf{x}) + C_0 + T(C_1 + C_2/N). \quad (1.6)$$

This argument is made precise under appropriate regularity conditions on Φ in Proposition 5.10.

At first, the estimate (1.6) appears to be useless in the non-convex setting, since as discussed already we do not expect $D_{mm}^2 U$ to be bounded. But suppose that we manage to produce regularization $(U^\epsilon)_{\epsilon > 0}$ of U such that for some $C > 0$,

- (1) $\|U - U^\epsilon\|_{L^\infty} \leq C\epsilon$,
- (2) U^ϵ is a subsolution of (HJB(∞)) up to an error of order $C\epsilon$, and
- (3) $\|D_{mm}^2 U^\epsilon\|_{L^\infty} \leq \frac{C}{\epsilon}$

Then, the estimate (1.6), together with the triangle inequality, gives

$$U(t, m_{\mathbf{x}}^N) \leq V^N(t, \mathbf{x}) + C\left(\epsilon + \frac{1}{N\epsilon}\right),$$

for some (new) constant C independent of N . Choosing $\epsilon = \frac{1}{\sqrt{N}}$ gives the upper bound in (1.3).

Of course, the question is how to produce functions U^ϵ satisfying properties (1)-(3). In finite dimensions (and even in Hilbert spaces, see [LL86]), sup-convolution is known to be a convenient way to create $\mathcal{C}^{1,1}$ regularity while preserving sub-solution properties (see for instance [BCD97, Chapter II]). This motivates our choice to define U^ϵ by

$$U^\epsilon(t, m) = \sup_{\nu} \left\{ U(t, \nu) - \frac{1}{2\epsilon} \|m - \nu\|_{-s}^2 \right\}, \quad (1.7)$$

with the supremum taken over the space of probability measures on \mathbb{T}^d , and with $s > 0$ chosen appropriately. Proposition 2.12 explains why H^{-s} is a good Hilbert space to work with - roughly speaking, when $s > d/2 + 1$, bounds on $U^\epsilon(t, \cdot)$ in $\mathcal{C}^{1,1}(H^{-s})$ (which arise naturally from the sup-convolution procedure) imply bounds on the L^∞ -norm of $D_{mm}^2 U^\epsilon$ as a consequence of Sobolev embeddings. Together with arguments taken mostly from [LL86] and presented in Proposition 4.3, this makes it possible to verify that U^ϵ in fact satisfies the estimates appearing in (1) and (3) above, provided that the original value function U is Lipschitz and semi-concave with respect to $\|\cdot\|_{-s}$. This explains why, in the H^{-s} -regular case, we must verify in Proposition 3.2 that U is Lipschitz and semi-concave with respect to $\|\cdot\|_{-s}$ when Assumption 2.4 is in force. Verifying that the functions U^ϵ satisfy point (2), the subsolution property, turns out to be much more subtle, and this is handled in Subsection 5.1. This is in fact the main challenge in the H^{-s} regular case.

Of course, there are several technical issues to overcome when implementing this argument which we have ignored in the above outline, for example even after regularizing we do not in fact have access to $D_{mm}^2 U^\epsilon$, so bounds on $\|D_{mm}^2 U^\epsilon\|_{L^\infty}$ have to be understood as bounds on the

Lipschitz constant of $D_m U(t, m, x)$ in m (with respect to d_1), and we must verify that bounds on the Lipschitz constant of $D_m U^\epsilon$ are in fact enough to execute the sketch of proof outlined above.

In the d_1 -regular case, the general strategy for the upper bound in (1.4) is the same - we want to approximate the value function U by functions which are smoother, but are still close to being sub-solutions to (HJB(∞)). But this time, we cannot directly apply sup-convolution in H^{-s} , because U is only Lipschitz with respect to d_1 . Instead, we start with a linear mollification procedure, which transforms U (which is only regular with respect to d_1) into some function U^δ which is regular with respect to $\|\cdot\|_{-s}$ for any s . Then we choose an appropriate s and apply sup-convolution in H^{-s} to produce $U^{\delta,\epsilon}$, which has the required regularity. The idea, like in points (1)-(3) above, is then to estimate the distance $\|U - U^{\delta,\epsilon}\|_{L^\infty}$, the amount by which $U^{\delta,\epsilon}$ fails to be a sub-solution to (HJB(∞)), and the size of $\|D_{mm}^2 U^{\delta,\epsilon}\|_\infty$, all as functions of the parameters δ and ϵ , and then choose δ and ϵ as appropriate functions of N to conclude.

Unfortunately, there is a last difficulty, which is that, roughly speaking, we could only find an efficient estimate for the subsolution property of $U^{\delta,\epsilon}$ at measures m which were bounded from below by a constant depending (explicitly) on δ and ϵ , i.e. for measures m such that $m \geq c(\delta, \epsilon)\text{Leb}$ for some appropriate c (with Leb denoting the Lebesgue measure on the torus). We refer to Lemma 5.8 for a precise statement of this result. The constraint on m arises when proving that $U^{\delta,\epsilon}$ inherits the regularity properties of U^δ , which is in fact just possible where the supremum defining $U^{\delta,\epsilon}$ is achieved in the interior of $\mathcal{P}(\mathbb{T}^d)$, with the interior being here defined with respect to the L^∞ norm. This step is the heart of the analysis in the d_1 -regular case, and requires a somewhat demanding technical result, see Proposition 4.4. The fact that we can only obtain a good estimate at such measures necessitates a final transformation, in which we replace $U^{\delta,\epsilon}$ by $U^{\delta,\epsilon,\lambda}(t, m) = U^{\delta,\epsilon}(t, (1 - \lambda)m + \lambda\text{Leb})$, and it is the function $U^{\delta,\epsilon,\lambda}$ which, in the d_1 -regular case, ultimately plays the role that U^ϵ played in the H^{-s} -regular case. The analysis of U^δ , $U^{\delta,\epsilon}$, and $U^{\delta,\epsilon,\lambda}$ is carried out in subsection 5.2.

For the lower bounds in (1.3) and (1.4), called the ‘easy inequalities’, we use a different, more control-theoretic argument. Very schematically, the idea is that optimal strategies identified in the asymptotic mean field regime are admissible in the finite setting, i.e., they can be played by a finite number of players. The converse is false and explains why the bounds in (1.3) and (1.4) are not symmetric (see also [CDJS23], where the two inequalities are also treated separately).

1.5. Further prospects. In the end, our regularization argument provides a different proof from [CDJS23]: in the latter, more effort is spent on the particle system itself; here, we mostly work with the Hamilton-Jacobi equation. Compared to [CDJS23], we work in a more restrictive setting (in particular we work with periodic data and we do not address the important issue of common noise), but we obtain much sharper estimates. Despite this improvement, the obtained bounds are not all optimal. The most interesting possible improvement, in our view, would be to obtain the bound

$$|U(t, m_{\mathbf{x}}^N) - V(t, \mathbf{x})| \leq CN^{-1/d}$$

in the d_1 -regular case (with the same corrections as in footnote 1 when $d = 1, 2$), which amounts to improving the lower bound in (1.4) to match the upper bound. Such a result would confirm our conjecture about the optimal rate of convergence in the d_1 -regular case. The reason why we obtain a slightly worse rate in (1.4) is related to the difficulty of estimating the subsolution property of $U^{\delta,\epsilon}$. Roughly speaking, we could have a more efficient estimate of the subsolution property of $U^{\delta,\epsilon}$, which would circumvent the need for $U^{\delta,\epsilon,\lambda}$ and lead to a rate closer to the optimal $N^{-1/d}$, if we could prove that the sup-convolution $U^{\delta,\epsilon}$ preserves the initial regularity of U^δ with respect to metrics other than H^{-s} - in particular, it is not clear to us whether the sup-convolutions $U^{\delta,\epsilon}$ preserves the d_1 -Lipschitz constant of U^δ on the whole space (and not only on the set of measures satisfying $m \geq c(\delta, \epsilon)\text{Leb}$, as we just explained). This would be immediate if U^δ were defined over

the entire Sobolev space H^{-s} , but things are much more difficult here, since the original problem is set on the smaller space of probability measures, which is of empty interior for any negative Sobolev norm. As discussed above, we partially circumvent this issue through Proposition 4.4, which requires a somewhat involved technical analysis, and it is possible that the latter could be refined. In any case, it seems clear to us that some new ideas are required in order to obtain the optimal rate $N^{-1/d}$ using our techniques.

A related possibility to obtain the optimal $N^{-1/d}$ rate in the d_1 -regular case would be to find another way to regularize the value function with similar features: the convergence rate should be explicit, the regularized value function should be regular enough (with explicit bounds) and should be a subsolution of the Hamilton-Jacobi equation up to an explicit residual term. In this regard, the sup-convolution is very convenient, but passing through a Hilbertian structure may seem somewhat unintuitive. At this stage, we have no intuition about the possible existence of a ‘better’ regularization technique. What is certain is that the same Sobolev spaces have been used in different contexts related to ours: not only in the analysis of fluctuations for non-controlled particle systems, as already mentioned above, but also in the study of a comparison principle for viscosity solutions of the Hamilton-Jacobi equation in the very recent work [SY22]. We are also convinced that the regularization technique we use could allow us to reprove a similar comparison principle. This leads us to believe that obtaining regularization methods (by sup-convolution or otherwise), specifically adapted to the Hamilton-Jacobi equation (HJB(∞)), is of broader interest than that of this work.

Another related question is to understand what our results say about the convergence of the optimal trajectories, i.e., the convergence of the optimal trajectories for the N -particle problem towards the optimal trajectories for the limiting. This problem is quite subtle, since optimal trajectories may not be unique when \mathcal{F} and \mathcal{G} are not convex. However, a recent result by Cardaliaguet and Souganidis [CS22] identifies an open and dense set $\mathcal{O} \subset [0, T] \times \mathcal{P}(\mathbb{T}^d)$ where U is locally of class $\mathcal{C}^{1,1}$, and shows that optimal trajectories which start in \mathcal{O} remain there, and in particular, U is $\mathcal{C}^{1,1}$ in a tube around an optimal trajectory which starts in \mathcal{O} . As shown in [CS22], this leads to a propagation of chaos result for initial conditions in \mathcal{O} , provided that one already has a rate of convergence of V^N to U . Lemma 3.3 of [CS22] explains clearly how the convergence rate of V^N to U impacts the rate of propagation of chaos, since R_N in that statement is just $\sup_{(t, \mathbf{x})} |V^N(t, \mathbf{x}) - U(t, m_{\mathbf{x}}^N)|$. In particular, our estimates can be immediately “plugged into” Lemma 3.1 of [CS22] to improve the main results of that paper (at least in the case of periodic data). The only unsatisfying point is that in the H^{-s} -regular case, the resulting rate of propagation of chaos will be dimension-free (in fact, the rate is $N^{-1/4}$), but only when measured up to stopping times $\tilde{\tau}^N$ which satisfies

$$\mathbb{P}[\tilde{\tau}^N < T] \leq CN^{-1/(d+8)}. \quad (1.8)$$

So, dimension still enters the propagation of chaos result through the asymptotic behavior of the stopping times. In order to get a truly dimension-independent propagation of chaos statement, one would need to replace the stopping times appearing in Lemma 3.3 of [CS22] with stopping times satisfying a dimension-free analogue of (1.8). To do this, we believe it would suffice to estimate the radius of the aforementioned tube as measured with respect to $\|\cdot\|_{-s}$ rather than d_1 , but we do not pursue this analysis here.

Of course, it would also be interesting to extend the analysis to the Euclidean (non-periodic) case. We believe that the weighted Sobolev spaces used in [FM97, M6196] in the analysis of the CLT for uncontrolled particle systems may also be useful here. The case with a common noise (which, as mentioned above, was treated in [CDJS23]) would also deserve some attention. Finally, it is certainly worth noting that, to our knowledge, the case of (non-cooperative) games remains beyond the reach of the methods developed here.

1.6. Organization of the paper. The article is organized as follows. Section 2 begins with a discussion of relevant notation and function spaces in Subsection 2.1, and then Subsections 2.2 and 2.3 contain the problem statement and main results. We also discuss the aforementioned examples in Subsection 2.4, and discuss the connection between analysis in $\mathcal{P}(\mathbb{T}^d)$ and analysis in H^{-s} in Subsection 2.5. Section 3 is devoted to establishing several fundamental properties of the value function in both the d_1 -regular and H^{-s} -regular cases. In Section 4, we propose and study several regularization methods for functions defined on the space of probability measures, including the sup-convolution method as well as the convolution method introduced in [CD22]. Section 5 is the true heart of the paper, with the establishment of the “hard inequalities”, i.e. the upper bounds in (1.3) and (1.4). Section 6 contains the proofs of the corresponding “easy inequalities”, while Section 7 contains some arguments related to the examples and the estimate (1.5) in the convex case. Finally, the Appendix contains a number of auxiliary results for finite dimensional linear PDEs and Hamilton-Jacobi-Bellman equations as well as mollification argument for functions defined on the space of probability measures which is borrowed from [CD22].

2. PRELIMINARIES AND MAIN RESULTS

2.1. Function spaces and notation. Firstly, as mentioned above, $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ is the d -dimensional flat torus, whose general element is written $x = (x_1, \dots, x_d)$. We write $\mathbf{x} = (x^1, \dots, x^N)$ for the general element of $(\mathbb{T}^d)^N$, with each $x^i = (x_1^i, \dots, x_d^i) \in \mathbb{T}^d$. We denote by $\mathcal{P}(\mathbb{T}^d)$ the set of probability measures on \mathbb{T}^d . We endow this space with the Wasserstein metric d_1 , defined by

$$d_1(m_1, m_2) = \sup_{g \text{ 1-Lip}} \int g(x)(m_1 - m_2)(dx),$$

the supremum being taken over 1-Lipschitz functions $g : \mathbb{T}^d \rightarrow \mathbb{R}$. Of course, by Kantorovich duality we have an equivalent definition in terms of couplings which we also use when convenient.

We are next going to describe several spaces of functions. First, we mention that we will use the notation $\text{Lip}(A)$ for the set of Lipschitz functions on a space A (the metric on A always being understood from context). Next, we fix some notation for multi-indices. We define multi-indices to be tuples $\mathbf{j} = (j_1, \dots, j_d)$ with each $j_i \in \{0, 1, 2, \dots\}$, and for $f : \mathbb{T}^d \rightarrow \mathbb{R}$ we interpret $D^{\mathbf{j}}f$ as

$$D^{\mathbf{j}}f = D_{x_1}^{j_1} f \dots D_{x_d}^{j_d} f.$$

We write $|\mathbf{j}| = \sum_{i=1}^d j_i$ for the order of the multi-index \mathbf{j} . We include the possibility that $\mathbf{j} = (0, \dots, 0)$, in which case $D^{\mathbf{j}}f = f$.

For $k \in \mathbb{N}$ (with $\mathbb{N} := \{1, 2, \dots\}$), we define $\mathcal{C}^k = \mathcal{C}^k(\mathbb{T}^d)$ to be the space of functions $\mathbb{T}^d \rightarrow \mathbb{R}$ with continuous derivatives up to order k . We endow \mathcal{C}^k with the norm

$$\|f\|_{\mathcal{C}^k} = \sum_{0 \leq |\mathbf{j}| \leq k} \|D^{\mathbf{j}}f\|_{L^\infty},$$

where $\|\cdot\|_{L^\infty}$ denotes the usual supremum norm.

For $f \in L^2(\mathbb{T}^d)$, we define the Fourier coefficients $(\widehat{f}^{\mathbf{k}})_{\mathbf{k} \in \mathbb{Z}^d}$ by

$$\widehat{f}^{\mathbf{k}} = \int_{\mathbb{T}^d} e^{i2\pi\mathbf{k}\cdot x} f(x) dx, \quad i^2 = -1.$$

Here we are writing $\mathbf{k} = (k_1, \dots, k_d)$ for an element of \mathbb{Z}^d . For $s \geq 0$, we define the Sobolev space H^s as the set of $f \in L^2(\mathbb{T}^d)$ such that

$$\|f\|_s^2 := \sum_{\mathbf{k} \in \mathbb{Z}^d} |\widehat{f}^{\mathbf{k}}|^2 (1 + \sum_{i=1}^d |k_i|^{2s}) < \infty.$$

The space H^s is a Hilbert space, with the norm $\|\cdot\|_s$ arising from the inner product

$$\langle f, g \rangle_s = \sum_{\mathbf{k} \in \mathbb{Z}^d} \widehat{f}^{\mathbf{k}} \overline{\widehat{g}^{\mathbf{k}}} \left(1 + \sum_{i=1}^d |k_i|^{2s}\right),$$

with the bar above denoting the usual conjugate of a complex number. We emphasize that this is not the usual definition of $\|\cdot\|_s$ – it would be more typical to replace $(1 + \sum_{i=1}^d |k_i|^{2s})$ with $(1 + |\mathbf{k}|^2)^s$. However, the norms generated by these two choices are equivalent, and in particular, this means that we have access to the usual Sobolev embeddings, e.g. for $s > d/2$ we have $\|f\|_{L^\infty} \leq C\|f\|_s$ for a constant C independent of f . The reason for using this particular choice of norm will become clear in the proof of Lemma 5.4.

Next, for $s \geq 0$ we define the space H^{-s} to be the dual of H^s , i.e. the set of bounded linear functionals $q : H^s \rightarrow \mathbb{R}$. We define the Fourier coefficients $(\widehat{q}^{\mathbf{k}})_{\mathbf{k} \in \mathbb{Z}^d}$ of $q \in H^{-s}$ by $\widehat{q}^{\mathbf{k}} = q(x \mapsto e^{i2\pi\mathbf{k}\cdot x})$. It is easy to check that the norm $\|\cdot\|_{-s}$ on H^{-s} inherited by duality is induced by the inner product

$$\langle p, q \rangle_{-s} := \sum_{\mathbf{k} \in \mathbb{Z}^d} \widehat{p}^{\mathbf{k}} \overline{\widehat{q}^{\mathbf{k}}} \left(1 + \sum_{i=1}^d |k_i|^{2s}\right)^{-1}. \quad (2.1)$$

For $f \in H^s$, we denote f^* for the dual element of H^{-s} , i.e. $\langle f^*, q \rangle_{-s} = \langle f, q \rangle_{s, -s} := q(f)$ for all $q \in H^{-s}$. Likewise, given $q \in H^{-s}$, we write q^* for the element of H^s such that $\langle q^*, f \rangle_s = q(f)$ for all $f \in H^s$. Using (2.1), it is easy to check that

$$\widehat{f^*}^{\mathbf{k}} = \left(1 + \sum_{i=1}^d |k_i|^{2s}\right) \widehat{f}^{\mathbf{k}}, \quad f \in H^s, \quad \text{and} \quad \widehat{q^*}^{\mathbf{k}} = \left(1 + \sum_{i=1}^d |k_i|^{2s}\right)^{-1} \widehat{q}^{\mathbf{k}}, \quad q \in H^{-s}. \quad (2.2)$$

We follow Chapter 5 of [CD18b] for the definition of the linear derivative of a function $\mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$. In particular, given a continuous function $\phi = \phi(m) : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$, we say that $\phi \in \mathcal{C}^1(\mathcal{P}(\mathbb{T}^d))$ if there is a continuous map $\frac{\delta\phi}{\delta m} = \frac{\delta\phi}{\delta m}(m, y) : \mathcal{P}(\mathbb{T}^d) \times \mathbb{T}^d \rightarrow \mathbb{R}$ such that

$$\phi(m_2) = \phi(m_1) + \int_0^1 \int_{\mathbb{T}^d} \frac{\delta\phi}{\delta m}(tm_2 + (1-t)m_1, y)(m_2 - m_1)(dy) dt. \quad (2.3)$$

The equation (2.3) determines $\frac{\delta\phi}{\delta m}$ only up to an m -dependent constant. We will make the additional normalization convention

$$\int_{\mathbb{T}^d} \frac{\delta\phi}{\delta m}(m, x) dx = 0, \quad (2.4)$$

under which $\frac{\delta\phi}{\delta m}$ is indeed unique. Thus if $\phi \in \mathcal{C}^1(\mathcal{P}(\mathbb{T}^d))$ then we can refer to the (unique) continuous function $\frac{\delta\phi}{\delta m}$ satisfying both (2.3) and (2.4) as the linear derivative of ϕ . Notice that our normalization convention differs to the usual one (which requires instead that the derivative at m has zero mean with respect to m). The reason is that, under our convention, the zeroeth Fourier coefficient of $\frac{\delta\phi}{\delta m}(m, \cdot)$ is zero, which simplifies an argument in the proof of Proposition 4.4. If ϕ has a linear derivative $\frac{\delta\phi}{\delta m}$ which is C^1 in y for each m , then we can define the L-derivative $D_m\phi : \mathcal{P}(\mathbb{T}^d) \times \mathbb{T}^d \rightarrow \mathbb{R}^d$ by

$$D_m\phi(m, y) = D_y \frac{\delta\phi}{\delta m}(m, y). \quad (2.5)$$

Unlike $\frac{\delta\phi}{\delta m}$, $D_m\phi$ is uniquely defined if it exists, see again [CD18b, Chapter 5].

2.2. Problem formulation. We fix a time horizon $T > 0$ and a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})$ satisfying the usual conditions, hosting independent d -dimensional Brownian motions W and $(W^i)_{i \in \mathbb{N}}$, and such that \mathcal{F}_0 is atomless. Our data consists of three functions

$$\mathcal{F}, \mathcal{G} : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}, \quad L = L(x, a) : \mathbb{T}^d \times \mathbb{R}^d \rightarrow \mathbb{R},$$

and the Lagrangian L determines a Hamiltonian H via the usual formula

$$H(x, p) = \sup_{a \in \mathbb{R}^d} \{ -L(x, a) - a \cdot p \}. \quad (2.6)$$

For $N \in \mathbb{N}$, V^N is defined by

$$V^N(t_0, \mathbf{x}_0) = \inf_{\alpha \in \mathcal{A}^N} \mathbb{E} \left[\int_{t_0}^T \left(\frac{1}{N} \sum_{i=1}^N L(X_t^i, \alpha_t^i) + \mathcal{F}(m_{\mathbf{X}_t}^N) \right) dt + \mathcal{G}(m_{\mathbf{X}_T}^N) \right], \quad (2.7)$$

for $(t_0, \mathbf{x}_0) \in [0, T] \times \mathbb{T}^d$, subject to

$$dX_t^i = \alpha_t^i dt + \sqrt{2} dW_t^i, \quad t_0 \leq t \leq T, \quad X_{t_0}^i = x_0^i, \quad (2.8)$$

where the infimum is taken over the set \mathcal{A}^N of all square-integrable and progressively measurable $(\mathbb{R}^d)^N$ -valued processes $\alpha = (\alpha^1, \dots, \alpha^N)$. We note that the Brownian motions $(W^i)_{i=1, \dots, N}$ and the processes $(X^i)_{i=1, \dots, N}$ in (2.8) are understood as taking values in \mathbb{R}^d , but they determine \mathbb{T}^d -valued processes in a canonical way (by composition with quotient map $\mathbb{R}^d \rightarrow \mathbb{R}^d/\mathbb{Z}^d$), which allows us to regard $m_{\mathbf{X}_t}^N$, for each $t \in [0, T]$, as a probability measure on \mathbb{T}^d . The function U , meanwhile, is given by

$$U(t_0, m_0) = \inf_{(m, \alpha)} \left\{ \int_{t_0}^T \left(\int_{\mathbb{R}^d} L(x, \alpha(t, x)) m_t(dx) + \mathcal{F}(m_t) \right) dt + \mathcal{G}(m_T) \right\}, \quad (2.9)$$

where the infimum is taken over all pairs (m, α) with $m = (m_t)_{t_0 \leq t \leq T} \in C([t_0, T], \mathcal{P}(\mathbb{T}^d))$ ($\mathcal{P}(\mathbb{T}^d)$ being equipped with the d_1 distance) and $\alpha : (t, x) \in [t_0, T] \times \mathbb{T}^d \mapsto \alpha(t, x) \in \mathbb{R}^d$ satisfying (in a weak distributional sense) the Fokker-Planck equation

$$\partial_t m_t = \Delta_x m_t - \operatorname{div}_x (m_t \alpha(t, \cdot)), \quad \text{in } [t_0, T] \times \mathbb{T}^d, \quad m_{t_0} = m_0,$$

and the integrability condition

$$\int_{t_0}^T \int_{\mathbb{T}^d} |\alpha(t, x)|^2 m_t(dx) dt < \infty.$$

We note that because there is no common noise, there is no subtlety in checking (under mild conditions on the data, and in particular under Assumption 2.1 below) that

$$U(t_0, m_0) = \inf_{\alpha \in \mathcal{A}} \mathbb{E} \left[\int_{t_0}^T \left(L(X_t, \alpha_t) + \mathcal{F}(\mathcal{L}(X_t)) \right) dt + \mathcal{G}(\mathcal{L}(X_T)) \right], \quad (2.10)$$

subject to

$$dX_t = \alpha_t dt + \sqrt{2} dW_t, \quad t_0 \leq t \leq T, \quad X_{t_0} = \xi \sim m_0, \quad (2.11)$$

where the infimum is taken over the set \mathcal{A} of square-integrable progressive processes \mathbb{R}^d -valued $\alpha = (\alpha_t)_{t_0 \leq t \leq T}$ and $\mathcal{L}(X_t) \in \mathcal{P}(\mathbb{T}^d)$ denotes, for any $t \in [t_0, T]$, the law of X_t (when viewed as a \mathbb{T}^d -valued random variable). Indeed, this is a consequence of a so-called mimicking argument, see e.g. the discussion in [LSZ20, Section 8]. We note that this alternative formulation is not crucial for any of our arguments, but at times is convenient, e.g. in the presentation of the examples in Subsection 2.4.

The function V^N is, under mild assumptions on the data (which are satisfied below under each of the two sets of assumptions spelled out in Subsection 2.3 below), the unique smooth solution of

(HJB(N)). The function U is expected to satisfy, in an appropriate viscosity sense, the equation (HJB(∞)).

2.3. Assumptions and main results. We have two main sets of assumptions, consistent with the d_1 -regular and the H^{-s} -regular cases already discussed at length in the introduction.

Assumption 2.1 (Assumptions for the d_1 -regular case). *We assume that there exists a constant $C \geq 1$ such that*

(1) $H \in C^2(\mathbb{T}^d \times \mathbb{R}^d)$, and for each $(x, p) \in \mathbb{T}^d \times \mathbb{R}^d$,

$$\frac{1}{C}I_{d \times d} \leq D_{pp}^2 H(x, p) \leq CI_{d \times d}.$$

(2) For all $(x, p) \in \mathbb{T}^d \times \mathbb{R}^d$,

$$|D_x H(x, p)| \leq C(1 + |p|). \quad (2.12)$$

(3) \mathcal{F} and \mathcal{G} are Lipschitz and semi-concave with respect to d_1 and to the constant C . Semi-concavity with respect to d_1 and to C means (explained only for \mathcal{F}) that

$$\mathcal{F}((1 - \lambda)m_0 + \lambda m_1) \geq (1 - \lambda)\mathcal{F}(m_0) + \lambda\mathcal{F}(m_1) - \lambda(1 - \lambda)\frac{C}{2}d_1^2(m_0, m_1), \quad (2.13)$$

for all $m_0, m_1 \in \mathcal{P}(\mathbb{T}^d)$ and all $\lambda \in (0, 1)$.

Remark 2.2. Let us remark that the conditions on H in Assumption 2.1 are fairly minimal. The strict convexity of H is standard, and ensures, among other things, that L inherits some regularity from H . Moreover, the bound $D_{pp}^2 H \leq CI_{d \times d}$ implies that L is coercive, in the sense that there is a constant C such that $L(x, a) \geq -C + \frac{1}{C}|a|^2$. The condition (2.12) is standard to find L^∞ estimates for the optimal controls in the problems defining U and V^N . The assumptions on H are satisfied for instance if $H(x, p) = \frac{1}{2}|p|^2 + H_1(x) \cdot p$ for some smooth vector field $H_1 : \mathbb{T}^d \rightarrow \mathbb{R}^d$.

Remark 2.3. The d_1 -Lipschitz continuity of \mathcal{F} and \mathcal{G} required in Assumption 2.1 is quite natural in this context, but the semi-concavity condition (2.13) may be more surprising. But a simple computation shows that (2.13) holds (for some constant C) if \mathcal{F} admits two linear derivatives $\frac{\delta \mathcal{F}}{\delta m}(m, y)$ and $\frac{\delta^2 \mathcal{F}}{\delta m^2}(m, y, z)$, the latter being defined analogously to the former, and

$$\forall m \in \mathcal{P}(\mathbb{T}^d), \forall y, z \in \mathbb{T}^d, \quad D_y D_z \frac{\delta^2 \mathcal{F}}{\delta m^2}(m, y, z) \leq CI_d,$$

with I_d denoting the d -dimensional identity matrix and the symbol \leq standing for the usual comparison between symmetric matrices. The reader will notice that the left-hand side of the inequality is nothing but the second-order Lions derivative $D_{mm}^2 \mathcal{F}(m, y, z)$. Thus the conditions on \mathcal{F} and \mathcal{G} in Assumption 2.1 are met, for instance, under the standing assumptions of [CDJS23]. A typical example of a function \mathcal{G} satisfying the conditions of Assumption 2.1 would be

$$\mathcal{G}(m) = G(m(\phi_1), \dots, m(\phi_k)),$$

where $k \in \mathbb{N}$, $G : \mathbb{R}^k \rightarrow \mathbb{R}$ is \mathcal{C}^2 with bounded first and second derivatives, and $\phi_1, \dots, \phi_k : \mathbb{T}^d \rightarrow \mathbb{R}$ are \mathcal{C}^1 with bounded derivatives.

Assumption 2.4 (Assumptions for the H^{-s} -regular case). *We assume that there is a number $s \in \mathbb{N}$ with $s > d/2 + 2$, such that*

(1) H belongs to $C^s(\mathbb{T}^d \times \mathbb{R}^d)$ and satisfies conditions (1) and (2) from Assumption (2.1).

(2) \mathcal{F} and \mathcal{G} belong $\mathcal{C}^1(\mathcal{P}(\mathbb{T}^d))$ with

$$\sup_{m \in \mathcal{P}} \left\| \frac{\delta \mathcal{F}}{\delta m}(m, \cdot) \right\|_{C^s} + \sup_{m \in \mathcal{P}} \left\| \frac{\delta \mathcal{G}}{\delta m}(m, \cdot) \right\|_{C^s} < +\infty. \quad (2.14)$$

Moreover \mathcal{F} and \mathcal{G} are semi-concave with respect to the norm $\|\cdot\|_{-s}$.

Remark 2.5. In Assumption 2.4, we keep the same basic structural conditions on the Hamiltonian, but require in addition “order s smoothness” of all the data. It is easy to check, using the definition of the linear derivatives that the condition (2.14) implies that \mathcal{F} and \mathcal{G} are Lipschitz with respect to $\|\cdot\|_{-s}$. A typical example of a function \mathcal{G} satisfying condition (2) in Assumption 2.4 would be

$$\mathcal{G}(m) = G(m(\phi_1), \dots, m(\phi_k))$$

where $k \in \mathbb{N}$, $G : \mathbb{R}^k \rightarrow \mathbb{R}$ is \mathcal{C}^2 with bounded first and second derivatives, and $\phi_1, \dots, \phi_k \in \mathcal{C}^s(\mathbb{T}^d)$.

Our main result in the d_1 -Lipschitz case is the following.

Theorem 2.6. *Let Assumption 2.1 hold. Then for each $\eta > 0$, there is a constant C such that for each $N \in \mathbb{N}$, we have, for all $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$,*

$$V^N(t, \mathbf{x}) - CR_{d,N} \leq U(t, m_{\mathbf{x}}^N) \leq V^N(t, \mathbf{x}) + CN^{-\beta(d)+\eta}, \quad \text{with } \beta(d) := \frac{2}{3d+6},$$

where

$$R_{d,N} := \begin{cases} N^{-1/2} & d = 1, \\ N^{-1/2} \log(1+N) & d = 2, \\ N^{-1/d} & d \geq 3. \end{cases} \quad (2.15)$$

And now we present our main result in H^{-s} -Lipschitz case.

Theorem 2.7. *Let Assumption 2.4 hold. Then there is a constant C such that for each $N \in \mathbb{N}$,*

$$V^N(t, \mathbf{x}) - \frac{C}{N} \leq U(t, m_{\mathbf{x}}^N) \leq V^N(t, \mathbf{x}) + \frac{C}{\sqrt{N}},$$

for all $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$.

Finally, we state a (much simpler) result in the convex setting.

Proposition 2.8. *Suppose that Assumption 2.1 is in force, and that in addition \mathcal{F} and \mathcal{G} are convex. Then there is a constant C such that*

$$0 \leq V^N(t, \mathbf{x}) - U(t, m_{\mathbf{x}}^N) \leq CR_{N,d},$$

for all $N \in \mathbb{N}$ and all $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$, where $R_{N,d}$ is as defined in (2.15).

2.4. Examples. In this section we give two examples which pinpoint the best convergence we can expect. In both cases, it is more convenient to state the examples in the Euclidean space \mathbb{R}^d rather than \mathbb{T}^d .

Example 1: In this example we define $V^N : [0, T] \times (\mathbb{R}^d)^N \rightarrow \mathbb{R}$ by

$$V^N(t_0, \mathbf{x}_0) = \inf_{\alpha \in \mathcal{A}^N} \mathbb{E} \left[\int_{t_0}^T \frac{1}{2N} \sum_{i=1}^N |\alpha_s^i|^2 ds + d_1(m_{\mathbf{X}_T}^N, \mathcal{N}_T) \right], \quad (2.16)$$

subject to dynamics

$$dX_t^i = \alpha_t^i dt + dW_t^i, \quad X_{t_0}^i = x_0^i, \quad (2.17)$$

with \mathcal{N}_T denoting the standard Gaussian measure on \mathbb{R}^d with mean 0 and variance $TI_{d \times d}$, and the infimum taken over all square-integrable and progressive processes $\alpha = (\alpha^1, \dots, \alpha^N)$. We note that the omission of the $\sqrt{2}$ the dynamics (2.17) was intentional, and based purely on notational convenience. Similarly, we define $U : (t, m) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^n) \mapsto U(t, m) \in \mathbb{R}$ by

$$U(t_0, m_0) = \inf_{\alpha \in \mathcal{A}} \mathbb{E} \left[\int_{t_0}^T \frac{1}{2} |\alpha_t|^2 dt + d_1(\mathcal{L}(X_T), \mathcal{N}_T) \right], \quad (2.18)$$

subject to the dynamics

$$dX_t = \alpha_t dt + dW_t, \quad X_{t_0} = \xi \sim m_0.$$

Then we have the following:

Proposition 2.9. *Let V^N and U be defined by (2.16) and (2.18). Then we have $U(0, \delta_0) = 0$, and if T is large enough then there is a constant $c > 0$ such that for infinitely many values of N ,*

$$V^N(t, 0) \geq cN^{-1/d}.$$

Of course, since $m \mapsto d_1(m, \mathcal{N}_T)$ is clearly d_1 -Lipschitz, Proposition 2.9 shows that we cannot have a rate better than $N^{-1/d}$ when the data is only d_1 -Lipschitz. The proof of Proposition 2.9 is given in Section 7. We note that the assumption that T is large enough is just to make the proof more straightforward, and it is also clear that the same argument should work for *all* N large enough, rather than for infinitely many N – we state things this way just to make the proof as transparent as possible.

Example 2: Given two costs functionals $F, G : \mathbb{R}^d \rightarrow \mathbb{R}$, consider the functions $V^N : [0, T] \times (\mathbb{R}^d)^N \rightarrow \mathbb{R}$ given by

$$V^N(t_0, \mathbf{x}_0) = \inf_{\alpha \in \mathcal{A}^N} \mathbb{E} \left[\int_{t_0}^T \left(\frac{1}{N} \sum_{i=1}^N L(\alpha^i(t, X_t^i)) + F\left(\frac{1}{N} \sum_{i=1}^N X_t^i\right) \right) dt + G\left(\frac{1}{N} \sum_{i=1}^N X_T^i\right) \right], \quad (2.19)$$

subject to the dynamics (2.8) as well as the function $U : [0, T] \times \mathcal{P}_2(\mathbb{R}^n) \mapsto U(t, m) \in \mathbb{R}$ given by

$$U(t, m) = \inf_{\alpha} \mathbb{E} \left[\int_t^T \left(L(\alpha(s, X_s)) + F(\mathbb{E}[X_s]) \right) ds + G(\mathbb{E}[X_T]) \right], \quad (2.20)$$

subject to the dynamics (2.11). In other words, we have defined V^N and U as in Subsection 2.2 but on the whole space and with

$$\mathcal{F}(m) = F(\bar{m}), \quad \mathcal{G}(m) = G(\bar{m}), \quad \bar{m} := \int_{\mathbb{R}} x dm(x).$$

Then we have the following result, whose elementary proof is presented in Section 7.

Proposition 2.10. *Suppose that F and G are Lipschitz, and $L = L(a)$ is C^2 with $\frac{1}{C}I_{d \times d} \leq D^2L \leq CI_{d \times d}$ for some constant $C \geq 1$. Let V^N and U be defined as in (2.19) and (2.20). Then we have*

$$U(t, m) = u(t, \bar{m}), \quad V^N(t, \mathbf{x}) = v^N(t, \bar{m}_{\mathbf{x}}^N) = v^N\left(t, \frac{1}{N} \sum_{i=1}^N x^i\right), \quad (2.21)$$

where u and v^N are the unique viscosity solutions of the finite-dimensional PDEs

$$-\partial_t u(t, x) + H(D_x u(t, x)) = F(x), \quad \text{for } (t, x) \in [0, T] \times \mathbb{R}^d, \quad u(T, x) = G(x), \quad \text{for } x \in \mathbb{R}^d.$$

and

$$\begin{aligned} -\partial_t v^N(t, x) - \frac{1}{N} \Delta_x v^N(t, x) + H(D_x v^N(t, x)) &= F(x), \quad \text{for } (t, x) \in [0, T] \times \mathbb{R}^d, \\ v^N(T, x) &= G(x), \quad \text{for } x \in \mathbb{R}^d, \end{aligned}$$

with H denoting the Hamiltonian defined by (2.6).

From Proposition 2.10 it is clear that

$$\sup_{(t, \mathbf{x}) \in [0, T] \times (\mathbb{R}^d)^N} |V^N(t, \mathbf{x}) - U(t, m_{\mathbf{x}}^N)| = \sup_{(t, x) \in [0, T] \times \mathbb{R}^d} |v^N(t, x) - u(t, x)|,$$

i.e. the convergence rate of V^N to U is completely governed by that of v^N to u . As discussed in the introduction, this is a well-studied problem, and classical results show that the rate is

$$\sup_{(t,x) \in [0,T] \times \mathbb{R}^d} |v^N(t,x) - u(t,x)| \leq \frac{C}{\sqrt{N}}.$$

2.5. Relationship between analysis on $\mathcal{P}(\mathbb{T}^d)$ and analysis on H^{-s} . In this paper we work both with the calculus for functions defined on $\mathcal{P}(\mathbb{T}^d)$ which is commonly used in the setting of mean field control and mean field games, and with the calculus on the space H^{-s} provided by the Hilbertian structure. In this Subsection we discuss both of these notions and their relationship with each other. Much of the analysis in this Subsection will be based on the fact that, as discussed in Subsection 2.1 we have access to the usual Sobolev embedding, and in particular we will use crucially the fact that

$$H^s \hookrightarrow L^\infty \text{ if } s > d/2, \quad H^s \hookrightarrow C^1 \text{ if } s > d/2 + 1. \quad (2.22)$$

Let us mention that we identify a probability measure m with an element of H^{-s} in the usual distributional way: for $f \in H^s$, $s > d/2$, we define $m(f) = \int f dm$, which makes sense thanks to (2.22). In this way we regard $\mathcal{P}(\mathbb{T}^d)$ as a subspace of H^{-s} , and it is easy to check that if $s > d/2$, then $\mathcal{P}(\mathbb{T}^d)$ is in fact a compact subset of H^{-s} . Given $s > 0$ and a function $\phi = \phi(q) : H^{-s} \rightarrow \mathbb{R}$ which has a Fréchet derivative at a point $q \in H^{-s}$, we use $D_{-s}\phi(q)$ and $\nabla_{-s}\phi(q) : H^{-s} \rightarrow H^{-s}$ to denote the derivative and the gradient of ϕ at q , respectively. That is, $D_{-s}\phi(q)$ is an element of H^s satisfying

$$\phi(p) = \phi(q) + \langle D_{-s}\phi(q), p - q \rangle_{s,-s} + o(\|p - q\|_{-s}), \quad p \in H^{-s}, \quad (2.23)$$

while $\nabla_{-s}\phi(q) = (D_{-s}\phi(q))^*$. We say that $\phi \in C^1(H^{-s})$ if ϕ is continuously differentiable.

We would like to understand the relationship between the derivatives $D_{-s}\phi$, $\nabla_{-s}\phi$ of ϕ on H^{-s} and the derivatives $\frac{\delta\phi}{\delta m}$, $D_m\phi$ of the restriction of ϕ to $\mathcal{P}(\mathbb{T}^d)$, which we again denote by ϕ . This is explained in the following Lemma.

Lemma 2.11. *Suppose that $s > d/2 + 1$, and that $\phi \in C^1(H^{-s})$. Then $\phi \in C^1(\mathcal{P}(\mathbb{T}^d))$ and*

$$\frac{\delta\phi}{\delta m}(m, y) = D_{-s}\phi(m)(y) - \int_{\mathbb{T}^d} D_{-s}\phi(m)(x) dx, \quad (m, y) \in \mathcal{P}(\mathbb{T}^d) \times \mathbb{T}^d.$$

Moreover, $\frac{\delta\phi}{\delta m}(m, \cdot) \in C^1(\mathbb{T}^d)$ for each m , and

$$D_m\phi(m, y) = D(D_{-s}\phi(m))(y).$$

In addition, $D_m\phi$ is jointly continuous on $\mathcal{P}(\mathbb{T}^d) \times \mathbb{T}^d$.

Proof. If $\phi \in C^1(H^{-s})$ and $m_1, m_2 \in \mathcal{P}(\mathbb{T}^d)$, then

$$\begin{aligned} \phi(m_2) - \phi(m_1) &= \int_0^1 \langle D_{-s}\phi(tm_2 + (1-t)m_1), m_2 - m_1 \rangle_{s,-s} dt \\ &= \int_0^1 \int_{\mathbb{T}^d} D_{-s}\phi(tm_2 + (1-t)m_1)(y)(m_2 - m_1)(dy) dt. \end{aligned}$$

To prove the first claim of the lemma, it remains only to check that $\mathcal{P}(\mathbb{T}^d) \times \mathbb{T}^d \ni (m, y) \mapsto D_{-s}\phi(m)(y)$ is continuous. Uniform Lipschitz continuity in y (uniformly in m) follows from the Sobolev embedding $H^s \hookrightarrow C^1(\mathbb{T}^d)$. For continuity in m , we use Sobolev embedding again, which says that if $m_k \xrightarrow{k \rightarrow \infty} m$ in d_1 , then $m_k \rightarrow m$ in H^{-s} (since $s > d/2 + 1$), so

$$|D_{-s}\phi(m_k)(y) - D_{-s}\phi(m)(y)| \leq \|D_{-s}\phi(m_k) - D_{-s}\phi(m)\|_{L^\infty} \leq \|D_{-s}\phi(m_k) - D_{-s}\phi(m)\|_s,$$

which tends to zero by continuity of $D_{-s}\phi$. The argument for the second claim is similar and so is omitted. \square

Next, we state a key result, which states how $\mathcal{C}^{1,1}$ regularity of a function $\phi : q \in H^{-s} \mapsto \phi(q) \in \mathbb{R}$ relates to Lipschitz bounds on the map $m \mapsto D_m \phi$. To state the lemma clearly, it is helpful to introduce the $\mathcal{C}^{1,1}$ seminorm of ϕ , which is defined by the formula

$$[\phi]_{\mathcal{C}^{1,1}(H^{-s})} = \sup_{q_1 \neq q_2} \frac{\|\nabla_{-s} \phi(q_1) - \nabla_{-s} \phi(q_2)\|_{-s}}{\|q_1 - q_2\|_{-s}}, \quad (2.24)$$

i.e. $[\phi]_{\mathcal{C}^{1,1}(H^{-s})}$ is the Lipschitz constant of the gradient $\nabla_{-s} \phi$.

Proposition 2.12. *Let $s > d/2 + 1$. Then there is a constant C depending only on d and s such that for any \mathcal{C}^1 map $\phi : q \in H^{-s} \mapsto \phi(q) \in \mathbb{R}$ and any $m_1, m_2 \in \mathcal{P}(\mathbb{T}^d)$, $y \in \mathbb{T}^d$, we have*

$$|D_m \phi(m_1, y) - D_m \phi(m_2, y)| \leq C [\phi]_{\mathcal{C}^{1,1}(H^{-s})} d_1(m_1, m_2).$$

Proof. On the one hand, using the Sobolev embedding $H^s \hookrightarrow \mathcal{C}^1(\mathbb{T}^d)$ together with Lemma 2.11, we find

$$\begin{aligned} |D_m \phi(m_1, y) - D_m \phi(m_2, y)| &= |D(D_{-s} \phi(m_1))(y) - D(D_{-s} \phi(m_2))(y)| \\ &\leq \|D_{-s} \phi(m_1) - D_{-s} \phi(m_2)\|_{\mathcal{C}^1} \\ &\leq C \|D_{-s} \phi(m_1) - D_{-s} \phi(m_2)\|_s \\ &\leq C [\phi]_{\mathcal{C}^{1,1}(H^{-s})} \|m_1 - m_2\|_{-s}. \end{aligned}$$

On the other hand, again using the Sobolev embedding $H^s \hookrightarrow \mathcal{C}^1$, we have $\|m_1 - m_2\|_{-s} \leq C d_1(m_1, m_2)$, and the result follows. \square

3. PROPERTIES OF THE VALUE FUNCTION

The goal of this section is to prove regularity properties of the value function U defined in (2.9) in the H^{-s} -regular case (this is Proposition 3.2) and in the d_1 -regular case (this is Proposition 3.4). In this latter case, we use an approximation argument to bypass the existence of linear functional derivatives for the mean field costs, usually required when one wants to use the system of optimality conditions. We also recall some viscosity solutions properties for the value function in Subsection 3.3.

3.1. Optimality conditions. We start with a standard lemma which characterizes the optimizers of the mean field control problem in terms of a forward-backward system of PDEs.

Lemma 3.1. *Let Assumption 2.1 hold, and assume in addition that $\mathcal{F}, \mathcal{G} \in \mathcal{C}^1(\mathcal{P}(\mathbb{T}^d))$ with*

$$\sup_{m \in \mathcal{P}(\mathbb{T}^d)} \left\| \frac{\delta \mathcal{F}}{\delta m}(m, \cdot) \right\|_{\mathcal{C}^2(\mathbb{T}^d)} + \sup_{m \in \mathcal{P}(\mathbb{T}^d)} \left\| \frac{\delta \mathcal{G}}{\delta m}(m, \cdot) \right\|_{\mathcal{C}^2(\mathbb{T}^d)} < \infty.$$

Then for any $(t_0, m_0) \in [0, T] \times \mathcal{P}(\mathbb{T}^d)$ there is at least one optimizer for the problem (2.9), and for any optimizer (m, α) we must have

(1) $\alpha(t, x) = -D_p H(x, Du(t, x))$ for some solution $(u, m) \in \mathcal{C}^{1,2}([t_0, T] \times \mathbb{T}^d) \times \mathcal{C}([t_0, T], \mathcal{P}(\mathbb{T}^d))$ of

$$\begin{cases} -\partial_t u(t, x) - \Delta_x u(t, x) + H(x, D_x u(t, x)) = \frac{\delta \mathcal{F}}{\delta m}(m_t, x), & (t, x) \in [t_0, T] \times \mathbb{T}^d, \\ \partial_t m(t, x) = \Delta_x m(t, x) + \operatorname{div}_x \left(m D_p H(x, D_x u(t, x)) \right), & (t, x) \in [t_0, T] \times \mathbb{T}^d, \\ m_{t_0} = m_0, \quad u(T, x) = \frac{\delta \mathcal{G}}{\delta m}(m_T, x), \end{cases} \quad (3.1)$$

with the second equation being understood in the weak sense.

(2) α satisfies

$$\|\alpha\|_{L^\infty} + \sup_{t \in [t_0, T]} \sqrt{T-t} \|D\alpha(t, \cdot)\|_{L^\infty} \leq C,$$

for some $C > 0$ depending on \mathcal{F} and \mathcal{G} only through their Lipschitz constants and independent of t_0 .

(3) If in addition Assumption 2.4 is in force, then $u(t, \cdot) : x \in \mathbb{T}^d \mapsto u(t, x) \in \mathbb{R}^d$ belongs to $\mathcal{C}^s(\mathbb{T}^d)$ for all $t \in [t_0, T]$ and to $\mathcal{C}^{s+1}(\mathbb{T}^d)$ for all $t \in [t_0, T]$ and we have

$$\sup_{t \in [t_0, T]} \|\alpha(t, \cdot)\|_{\mathcal{C}^{s-1}} + \sup_{t \in [t_0, T]} \sqrt{T-t} \|\alpha(t, \cdot)\|_{\mathcal{C}^s} \leq C.$$

Proof. Claim (1) follows easily from standard arguments (see e.g. [BC18] Lemma 3.1). Claims (2) and (3) follows easily from Claim (1), thanks to Lemma A.2 from the Appendix. \square

3.2. Regularity of the value function.

3.2.1. *The H^{-s} -regular case.* The properties of U in the H^{-s} -regular case are summarized in the following proposition.

Proposition 3.2. *Under Assumption 2.4, there is a constant C , depending only on the various data appearing in the assumption, such that the following holds:*

(1) *The function U is globally C -Lipschitz with respect to H^{-s} . More precisely, for all $t_1, t_2 \in [0, T]$ and $m_1, m_2 \in \mathcal{P}(\mathbb{T}^d)$, we have*

$$|U(t_1, m_1) - U(t_2, m_2)| \leq C(|t_1 - t_2| + \|m_1 - m_2\|_{-s}).$$

(2) *For each $t \in [0, T]$, the map $\mathcal{P}(\mathbb{T}^d) \ni m \mapsto U(t, m)$ is C -semi-concave with respect to the norm $\|\cdot\|_{-s}$. More precisely, for each $t \in [0, T]$, $m_1, m_2 \in \mathcal{P}(\mathbb{T}^d)$, and $\lambda \in (0, 1)$, we have*

$$U(t, (1-\lambda)m_1 + \lambda m_2) \geq (1-\lambda)U(t, m_1) + \lambda U(t, m_2) - \frac{C}{2}\lambda(1-\lambda)\|m_1 - m_2\|_{-s}^2.$$

To prove Proposition 3.2 we will need the following stability result.

Lemma 3.3. *Assume that $s > d/2$ and $\alpha : (t, x) \in [t_0, T] \times \mathbb{T}^d \mapsto \alpha(t, x) \in \mathbb{R}^d$ satisfies*

$$\sup_{t_0 \leq t \leq T} \|\alpha(t, \cdot)\|_{s-1} \leq C.$$

Then there is a constant C' depending only on C such that if $(m^i)_{i=1,2} \in \mathcal{C}([t_0, T], \mathcal{P}(\mathbb{T}^d))$ satisfy in the weak sense

$$\partial_t m^i(t, x) = \Delta_x m^i(t, x) - \operatorname{div}_x(m^i(t, x)\alpha(t, x)),$$

for $(t, x) \in [t_0, T] \times \mathbb{T}^d$, then we have

$$\sup_{t_0 \leq t \leq T} \|m_t^1 - m_t^2\|_{-s} \leq C' \|m_{t_0}^1 - m_{t_0}^2\|_{-s}. \quad (3.2)$$

The proof of Lemma 3.3 is given in Appendix A.1.

Proof of Proposition 3.2. Claim 1. For regularity in m , we fix $t_0 \in [0, T]$ and $m_1, m_2 \in \mathcal{P}(\mathbb{T}^d)$. Let α be an optimal control for (t_0, m_1) , and let $(m^i)_{t_0 \leq t \leq T}$, for $i = 1, 2$, denote the solutions on $[t_0, T]$ to

$$\partial_t m^i(t, x) = \Delta_x m^i(t, x) - \operatorname{div}_x(m^i(t, x)\alpha(t, x)), \quad m_{t_0}^i = m_i.$$

Then we have

$$\begin{aligned}
U(t_0, m_1) &= \int_{t_0}^T \left(\int_{\mathbb{R}^d} L(x, \alpha(t, x)) m_t^1(dx) + \mathcal{F}(m_t^1) \right) dt + \mathcal{G}(m_T^1) \\
&= \int_{t_0}^T \left(\int_{\mathbb{R}^d} L(x, \alpha(t, x)) (m_t^2 + m_t^1 - m_t^2)(dx) + \mathcal{F}(m_t^2) + \mathcal{F}(m_t^1) - \mathcal{F}(m_t^2) \right) dt \\
&\quad + \mathcal{G}(m_T^2) + \mathcal{G}(m_T^1) - \mathcal{G}(m_T^2) \\
&\geq U(t_0, m_2) - C \left(1 + \int_{t_0}^T \|L(\cdot, \alpha(t, \cdot))\|_{C^s} dt \right) \sup_{t_0 \leq t \leq T} \|m_t^1 - m_t^2\|_{-s} \\
&\geq U(t_0, m_2) - C \left(1 + \int_{t_0}^T \frac{1}{\sqrt{T-t}} dt \right) \sup_{t_0 \leq t \leq T} \|m_t^1 - m_t^2\|_{-s} \\
&\geq U(t_0, m_2) - C \|m_1 - m_2\|_{-s}.
\end{aligned}$$

The first inequality uses the fact that, by Lemma 3.1, we have

$$L(x, \alpha(t, x)) = L\left(x, -D_p H(x, D_x u(t, x))\right) = -H(x, D_x u(t, x)) - \alpha(t, x) \cdot D_x u(t, x),$$

where (m^1, u) is a solution of (3.1), and so by Lemma 3.1 and the regularity of H ,

$$\|L(\cdot, \alpha(t, \cdot))\|_{C^s} \leq \frac{C}{\sqrt{T-t}}.$$

For the final inequality, we used the bound $\sup_{t_0 \leq t \leq T} \|m_t^1 - m_t^2\|_{-s} \leq C \|m_1 - m_2\|_{-s}$ from Lemma 3.3. Up to inverting the roles of m_0 and m_1 , this completes the proof of the Lipschitz regularity in m .

For the time regularity we proceed as follows. Given $m_0 \in \mathcal{P}(\mathbb{T}^d)$, $t_0 \in [0, T)$ and $h > 0$ such that $t_0 + h \in [0, T]$ we can expand

$$U(t_0 + h, m_0) - U(t_0, m_0) = U(t_0 + h, m_0) - U(t_0 + h, m_{t_0+h}^0) + U(t_0 + h, m_{t_0+h}^0) - U(t_0, m_0), \quad (3.3)$$

where $(m_t^0)_{t_0 \leq t \leq T}$ is an optimal trajectory for $U(t_0, m_0)$ with control α . On the one hand, by dynamic programming, using that α is bounded independently from (t_0, m_0) and the boundness of \mathcal{F} , we can argue that

$$|U(t_0 + h, m_{t_0+h}^0) - U(t_0, m_0)| \leq \left| \int_{t_0}^{t_0+h} \int_{\mathbb{T}^d} L(x, \alpha(t, x)) m_t^0(dx) dt + \int_{t_0}^{t_0+h} \mathcal{F}(m_t^0) dt \right| \leq Ch$$

for some C depending only the growth properties of L and \mathcal{F} . On the other hand, by Lipschitz regularity of $U(t, \cdot)$ over $\mathcal{P}(\mathbb{T}^d)$ with respect to $\|\cdot\|_{-s}$ we can argue that

$$|U(t_0, m_{t_0+h}^0) - U(t_0, m_0)| \leq C \|m_{t_0+h}^0 - m_0\|_{-s}.$$

Now, using the Sobolev embedding $H^s \hookrightarrow C^2$ (recall that $s > d/2 + 2$ here), we have

$$\int_{\mathbb{T}^d} \phi(x) (m_{t_0+h}^0 - m_0)(dx) = \int_{t_0}^{t_0+h} \int_{\mathbb{T}^d} [D\phi(x) \cdot \alpha(t, x) + \Delta\phi(x)] m_t^0(dx) dt \leq Ch \|\phi\|_s,$$

so that $\|m_{t_0+h}^0 - m_0\|_s \leq Ch$, which, in light of (3.3), completes the proof that U is Lipschitz in time.

Claim 2. We now turn to the semi-concavity of U . We fix $m_1, m_2 \in \mathcal{P}(\mathbb{T}^d)$, $t_0 \in [0, T]$ and $\lambda \in (0, 1)$ and we consider α to be an optimal control for $U(t_0, m_\lambda)$ with $m_\lambda = (1 - \lambda)m_1 + \lambda m_2$. By Lemma 3.1, we know that $\alpha(t, \cdot)$ is bounded in H^{s-1} , uniformly in $t \in [t_0, T]$. This time we define $(m_t^1)_{t_0 \leq t \leq T}$, $(m_t^2)_{t_0 \leq t \leq T}$ and $(m_t^\lambda)_{t_0 \leq t \leq T}$ to be the solutions to

$$\partial_t m(t, x) = \Delta_x m(t, x) - \operatorname{div}_x(m(t, x)\alpha(t, x)), \quad \text{in } [t_0, T) \times \mathbb{T}^d, \quad (3.4)$$

starting respectively from m_1 , m_2 and m_λ . By linearity of the Fokker-Planck equation (3.4) it holds for all $t \in [t_0, T]$,

$$m_t^\lambda = (1 - \lambda)m_t^1 + \lambda m_t^2. \quad (3.5)$$

By optimality of α for $U(t_0, m_{t_0}^\lambda)$ we have

$$\begin{aligned} & (1 - \lambda)U(t_0, m_1) + \lambda U(t_0, m_2) - U(t_0, m_\lambda) \\ & \leq (1 - \lambda) \int_{t_0}^T \int_{\mathbb{R}^d} L(x, \alpha(t, x)) m_t^1(dx) dt + \lambda \int_{t_0}^T \int_{\mathbb{R}^d} L(x, \alpha(t, x)) m_t^2(dx) dt \\ & \quad - \int_{t_0}^T \int_{\mathbb{R}^d} L(x, \alpha(t, x)) m_t^\lambda(dx) dt \\ & \quad + \int_{t_0}^T \left[(1 - \lambda)\mathcal{F}(m_t^1) + \lambda\mathcal{F}(m_t^2) - \mathcal{F}(m_t^\lambda) \right] dt \\ & \quad + (1 - \lambda)\mathcal{G}(m_T^1) + \lambda\mathcal{G}(m_T^2) - \mathcal{G}(m_T^\lambda). \end{aligned}$$

The terms involving the running cost L cancel out since m_t^λ satisfies (3.5). For the remaining terms, we use the semi-concavity of \mathcal{F} and \mathcal{G} as well as Lemma 3.3 to deduce that, for some constant $C > 0$ depending on the semi-concavity constants of \mathcal{F} and \mathcal{G} as well as $\sup_{t_0 \leq t \leq T} \|\alpha(t, \cdot)\|_{s-1}$ (and independent of t_0),

$$\begin{aligned} & (1 - \lambda)U(t_0, m_1) + \lambda U(t_0, m_2) - U(t_0, m_\lambda) \\ & \leq \frac{C}{2} \lambda (1 - \lambda) \sup_{t \in [t_0, T]} \|m_t^1 - m_t^2\|_{-s}^2 \leq \frac{C}{2} \lambda (1 - \lambda) \|m_1 - m_2\|_{-s}^2, \end{aligned}$$

which concludes the proof of the proposition. \square

3.2.2. The d_1 -regular case. We now investigate the regularity of U under Assumption 2.1. We will use the notation

$$\|m^2 - m^1\|_{-2, \infty} := \sup_{\|\phi\|_{C^2} \leq 1} \int_{\mathbb{R}^d} \phi(x) (m^2 - m^1)(dx) \quad (3.6)$$

for any two probability measures $m^1, m^2 \in \mathcal{P}(\mathbb{T}^d)$. The regularity properties of U can be summarized as follows.

Proposition 3.4. *Under Assumption 2.1, there is a constant C , depending only on the various data appearing in the assumption, such that the following holds:*

- (1) U is d_1 -Lipschitz with respect to m with constant C , i.e. for each $t \in [0, T]$ and each $m_1, m_2 \in \mathcal{P}(\mathbb{T}^d)$, we have

$$|U(t, m_1) - U(t, m_2)| \leq C d_1(m_1, m_2).$$

- (2) U is locally $\|\cdot\|_{-2, \infty}$ -Lipschitz: it satisfies, for each $t \in [0, T]$ and each $m_1, m_2 \in \mathcal{P}(\mathbb{T}^d)$, the estimate

$$|U(t, m_1) - U(t, m_2)| \leq \frac{C}{\sqrt{T-t}} \|m_2 - m_1\|_{-2, \infty}. \quad (3.7)$$

- (3) U is d_1 -semiconcave with respect to m with constant C , i.e. for each $t \in [0, T]$, $\lambda \in (0, 1)$, and each $m_1, m_2 \in \mathcal{P}(\mathbb{T}^d)$, we have

$$U(t, (1 - \lambda)m_1 + \lambda m_2) \geq (1 - \lambda)U(t, m_1) + \lambda U(t, m_2) - \frac{C}{2} \lambda (1 - \lambda) d_1^2(m_1, m_2).$$

(4) U is Hölder continuous and locally Lipschitz continuous in time: it satisfies

$$|U(t+h, m) - U(t, m)| \leq C \left(\sqrt{h} \wedge \frac{h}{\sqrt{T-h-t}} \right),$$

for each $t \in [0, T]$, $h \in (0, T-t)$, and $m \in \mathcal{P}(\mathbb{T}^d)$.

The rest of this subsection is devoted to proving Proposition (3.4). Assumption 2.1 is in force, and we also assume that $(\mathcal{F}^n)_{n \geq 0}$ and $(\mathcal{G}^n)_{n \geq 0}$ are sequences of smooth approximations of \mathcal{F} and \mathcal{G} such that,

- (n.1) for all $n \geq 0$, \mathcal{F}^n and \mathcal{G}^n are Lipschitz and semi-concave with Lipschitz and semi-concavity constants bounded independently from n .
- (n.2) $\mathcal{F}^n \rightarrow \mathcal{F}$ and $\mathcal{G}^n \rightarrow \mathcal{G}$ uniformly over $\mathcal{P}(\mathbb{T}^d)$ as $n \rightarrow +\infty$.
- (n.3) For all $n \geq 0$, $\mathcal{F}^n, \mathcal{G}^n \in \mathcal{C}^1(\mathcal{P}(\mathbb{T}^d))$ with $\frac{\delta \mathcal{F}^n}{\delta m}(m, \cdot), \frac{\delta \mathcal{G}^n}{\delta m}(m, \cdot) \in \mathcal{C}^2(\mathbb{T}^d)$ for all $m \in \mathbb{T}^d$ with jointly continuous derivatives.

The existence of such approximations is proved in Lemma 4.1 below. For all $n \geq 0$ we define U^n to be the value function in (2.9) with costs \mathcal{F}^n and \mathcal{G}^n instead of \mathcal{F} and \mathcal{G} . It is plain to check from the control formulation that U^n converges uniformly to U over $[0, T] \times \mathcal{P}(\mathbb{T}^d)$ as $n \rightarrow +\infty$ (thanks to property (n.2) right above).

The proof of Proposition 3.4 will make use of the following stability result,

Lemma 3.5. *Suppose that $\alpha : (t, x) \in [t_0, T] \times \mathbb{T}^d \mapsto \alpha(t, x) \in \mathbb{R}^d$ is differentiable in x on $[t_0, T] \times \mathbb{R}^d$ and satisfies*

$$\sup_{t_0 \leq t \leq T} \|\alpha(t, \cdot)\|_{L^\infty} + \sup_{t_0 \leq t < T} \sqrt{T-t} \|D\alpha(t, \cdot)\|_{L^\infty} \leq C, \quad (3.8)$$

for some $C > 0$, then, there is C' depending only on C such that if $(m^i)_{i=1,2} \in \mathcal{C}([t_0, T], \mathcal{P}(\mathbb{T}^d))$ satisfy in the weak sense

$$\partial_t m^i(t, x) = \Delta_x m^i(t, x) - \operatorname{div}_x(m^i(t, x)\alpha(t, x)),$$

for all $(t, x) \in [t_0, T] \times \mathbb{T}^d$, then we have

$$\sup_{t \in [t_0, T]} d_1(m_t^1, m_t^2) \leq C' d_1(m_{t_0}^1, m_{t_0}^2), \quad (3.9)$$

and, using notation (3.6),

$$\sup_{t \in [t_0, T]} \sqrt{t-t_0} d_1(m_t^1, m_t^2) \leq C' \|m_{t_0}^2 - m_{t_0}^1\|_{-2, \infty}. \quad (3.10)$$

The proof is given in Appendix A.1. We now prove the main result of the subsection, Proposition 3.4.

Proof of Proposition 3.4. Claim 1: We first show the analog result for U^n . Fix $t_0 \in [0, T]$ and let $(m_t^i)_{i=1,2}$ be solutions to

$$\partial_t m^i(t, x) - \Delta_x m^i(t, x) + \operatorname{div}_x(\alpha^n(t, x)m^i(t, x)) = 0, \quad m_{t_0}^i = m_i,$$

as in the statement of Lemma 3.1, where α^n is an optimal control for $U^n(t_0, m_1)$. Thanks to Lemma 3.1 and to Property (n.1), we know that α^n satisfies (3.8) for some C independent from n . By

optimality of α for $U^n(t_0, m_1)$ and Lipschitz regularity of \mathcal{F}^n and \mathcal{G}^n , inequality

$$\begin{aligned}
& U^n(t_0, m_1) - U^n(t_0, m_2) \\
& \geq \int_{t_0}^T \int_{\mathbb{T}^d} L(x, \alpha^n(t, x))(m_t^1 - m_t^2)(dx)dt + \int_{t_0}^T (\mathcal{F}^n(m_t^1) - \mathcal{F}^n(m_t^2))dt + \mathcal{G}^n(m_T^1) - \mathcal{G}^n(m_T^2) \\
& \geq -C \int_{t_0}^T (1 + \|D\alpha^n(t, \cdot)\|_{L^\infty}) d_1(m_t^1, m_t^2)dt - C \int_{t_0}^T d_1(m_t^1, m_t^2) - Cd_1(m_T^1, m_T^2) \quad (3.11) \\
& \geq -C \int_{t_0}^T (1 + \|D\alpha^n(t, \cdot)\|_{L^\infty}) d_1(m_t^1, m_t^2)dt - C \sup_{t_0 \leq t \leq T} d_1(m_t^1, m_t^2) \\
& \geq -C \left(1 + \int_{t_0}^T \|D\alpha^n(t, \cdot)\|_{L^\infty} dt\right) \sup_{t \in [t_0, T]} d_1(m_t^1, m_t^2)
\end{aligned}$$

holds for some C independent of n (and of t_0, m_1, m_2). Using (3.9) in Lemma 3.5 and then letting $n \rightarrow +\infty$, we conclude that

$$U(t_0, m_2) - U(t_0, m_1) \leq Cd_1(m_{t_0}^1, m_{t_0}^2) = Cd_1(m_1, m_2).$$

Reversing the roles of m_1 and m_2 , we conclude that U is Lipschitz continuous in the measure variable.

Claim 2: Using (3.10) from Lemma 3.5 instead of (3.9) in (3.11) leads to Claim 2.

Claim 3: To prove the semi-concavity, we argue as in the H^{-s} -regular case, using the stability of the Fokker-Planck equation given by Lemma 3.5. This allows one to show that there is some $C > 0$ independent of $n \geq 0$ such that, for all $m_1, m_2 \in \mathcal{P}(\mathbb{T}^d)$ and all $\lambda \in (0, 1)$, it holds

$$U^n(t, (1-\lambda)m_1 + \lambda m_2) \geq (1-\lambda)U^n(t_0, m_1) + \lambda U^n(t_0, m_2) - \frac{C}{2}\lambda(1-\lambda)d_1^2(m_1, m_2).$$

Using the convergence of U^n toward U , we conclude that U is semi-concave in the measure variable.

Claim 4: We start by the analog statement for U^n for $n \geq 0$. Let $m \in \mathcal{P}(\mathbb{T}^d)$, $t \in [0, T]$ and $h \in (0, T-t]$. Following the same argument as in the proof of the time Lipschitz regularity in Proposition 3.2 and using (3.7), we infer that

$$|U^n(t+h, m) - U^n(t, m)| \leq \frac{Ch}{\sqrt{T-t-h}}. \quad (3.12)$$

The global Hölder regularity in time is, on the other hand, standard. Once we have (3.12) it is straightforward, simply by noticing that the latter implies that, for all $m \in \mathcal{P}(\mathbb{T}^d)$, $t \mapsto U(t, m)$ is absolutely continuous over $[0, T]$ with

$$\left| \frac{d}{dt} U^n(t, m) \right| \leq \frac{C}{\sqrt{T-t}}, \quad \text{for almost all } t \in [0, T],$$

and therefore, integrating in time this leads, for all $m \in \mathcal{P}(\mathbb{T}^d)$, all $t \in [0, T]$ and all $h \in [0, T-t]$ to

$$|U^n(t+h, m) - U^n(t, m)| \leq C\sqrt{h}.$$

Passing to the limit when $n \rightarrow +\infty$ gives the result. \square

3.3. Viscosity (sub)-solution property. Here we record the fact that U is a viscosity solution of the equation (HJB(∞)). In fact, we only need the sub-solution property, and we only need to consider test functions $\phi \in C^1((0, T) \times H^{-s})$.

Lemma 3.6. *Suppose that $\phi \in C^1((0, T) \times H^{-s})$ touches U from above at $(t_0, m_0) \in (0, T) \times \mathcal{P}(\mathbb{T}^d)$ that is*

$$U(t_0, m_0) - \phi(t_0, m_0) = \sup_{(t, m) \in (0, T) \times \mathcal{P}(\mathbb{T}^d)} [U(t, m) - \phi(t, m)], \quad (3.13)$$

then, it holds that

$$-\partial_t \phi(t_0, m_0) - \int_{\mathbb{R}^d} \Delta_x \frac{\delta \phi}{\delta m}(t_0, m_0, x) m_0(dx) + \int_{\mathbb{R}^d} H\left(x, D_x \frac{\delta \phi}{\delta m}(t_0, m_0, x)\right) m_0(dx) \leq \mathcal{F}(m_0).$$

The proof is standard, up to using Lemma 2.11 to check that tests functions $\phi \in \mathcal{C}^1((0, T) \times H^{-s})$ are regular enough for the usual argument based on the chain rule for flows of probability measures to apply. We refer for instance to [CD18a] section 4.4.3 for a proof and the corresponding supersolution property.

4. THREE REGULARIZATION PROCEDURES

In this section, we present and analyze three methods for regularizing a function $\Phi : m \in \mathcal{P}(\mathbb{T}^d) \mapsto \Phi(m) \in \mathbb{R}$.

4.1. Regularization by mollification of the Fourier coefficients.

Lemma 4.1. *For two constants c_1 and c_2 , let $\Phi : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ be c_1 -Lipschitz continuous and c_2 -semi-concave with respect to the d_1 -Wasserstein distance. Then, there exists a sequence of functions $(\Phi^n : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R})_{n \geq 1}$ such that*

- (1) $(\Phi^n)_{n \geq 1}$ converges to Φ , uniformly on $\mathcal{P}(\mathbb{T}^d)$;
- (2) for any $n \geq 1$, Φ^n is c_1 -Lipschitz continuous and c_2 -semi-concave with respect to d_1 ;
- (3) for any $n \geq 1$, Φ^n is continuously differentiable with respect to m (for d_1) and its derivative $\delta \Phi^n / \delta m$ (defined on $\mathcal{P}(\mathbb{T}^d) \times \mathbb{T}^d$) is jointly continuous (with $\mathcal{P}(\mathbb{T}^d)$ being equipped with d_1) and has jointly continuous derivatives of any order in the x -variable.

The proof of the above result is a direct consequence of a mollification argument explained in Proposition 3.14 in [CD22]. For completeness, we present a sketch of the proof in Appendix B.

4.2. Regularization by mollification of the measure argument.

Lemma 4.2. *For two constants c_1 and c_2 , let $\Phi : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ be c_1 -Lipschitz continuous and c_2 -semi-concave with respect to the d_1 -Wasserstein distance. For a smooth symmetric density ρ on \mathbb{R}^d with compact support and for $\delta > 0$, let*

$$\rho_\delta(x) := \frac{1}{\delta^d} \rho\left(\frac{x}{\delta}\right), \quad x \in \mathbb{R}^d,$$

and then

$$\Phi^\delta(m) := \Phi(m * \rho_\delta), \quad m \in \mathcal{P}(\mathbb{T}^d).$$

Then, for any $s > 0$, there exists a constant $\Gamma(d, \rho, s) \geq 0$ such that

- (1) $\sup_{m \in \mathcal{P}(\mathbb{T}^d)} |\Phi^\delta(m) - \Phi(m)| \leq c_1 \Gamma(d, \rho, s) \delta$.
- (2) Φ^δ is $\Gamma(d, \rho, s) c_1 \delta^{-(s-1)}$ -Lipschitz continuous and $\Gamma(d, \rho, s) c_2 \delta^{-2(s-1)}$ -semi-concave with respect to $\|\cdot\|_{-s}$.

Lastly, if Φ is continuous differentiable with respect to m , so is Φ^δ and the derivative is given by

$$\frac{\delta \Phi^\delta}{\delta m}(m, x) = \left(\frac{\delta \Phi}{\delta m}(m * \rho_\delta, \cdot) * \rho_\delta \right)(x), \quad m \in \mathcal{P}(\mathbb{T}^d), \quad x \in \mathbb{T}^d.$$

Proof. Let us first recall that there is a constant $\Gamma(d, \rho, s)$ such that for $f \in L^2$,

$$\|f * \rho_\delta\|_s \leq \Gamma(d, \rho, s) \delta^{-(s-1)} \|f\|_1.$$

This fact can easily be proven via Young's convolution inequality when $s \in \mathbb{N}$, and can be proved in a straightforward way via Fourier analysis when $s \notin \mathbb{N}$, so we omit the details. We now proceed in several steps.

Step 1. The first claim in the proof follows from the Lipschitz property of Φ under d_1 together with the fact that, for any $m \in \mathcal{P}(\mathbb{T}^d)$,

$$d_1(m * \rho_\delta, m) \leq \sup_{f \text{ 1-Lip}} \int_{\mathbb{T}^d} [f * \rho_\delta(x) - f(x)] m(dx) \leq \delta.$$

Similarly, the second claim in the proof follows from the fact that, for any two $m, m' \in \mathcal{P}(\mathbb{T}^d)$,

$$\begin{aligned} d_1(m * \rho_\delta, m' * \rho_\delta) &\leq \sup_{f \text{ 1-Lip}} \int_{\mathbb{T}^d} f * \rho_\delta(x) (m - m')(dx) \\ &\leq \|m - m'\|_{-s} \sup_{f \text{ 1-Lip}: \widehat{f}^0=0} \|f * \rho_\delta\|_s \\ &\leq \|m - m'\|_{-s} \frac{\Gamma(d, \rho, s)}{\delta^{s-1}} \sup_{f \text{ 1-Lip}: \widehat{f}^0=0} \|f\|_1 \\ &\leq \|m - m'\|_{-s} \frac{\Gamma(d, \rho, s)}{\delta^{s-1}}. \end{aligned}$$

which completes the proof of the first part.

Step 2. The claim related with semi-concavity is shown in a somewhat similar manner. Indeed, we recall (once again) that the semi-concavity of Φ may be expressed as

$$\Phi(\lambda m' + (1 - \lambda)m) \geq \lambda \Phi(m') + (1 - \lambda)\Phi(m) - \frac{c_2}{2} \lambda(1 - \lambda) d_1(m, m')^2.$$

Replacing m by $m * \rho_\delta$ and then m' by $m' * \rho_\delta$, we see that the only difficulty is to upper bound $d_1(m * \rho_\delta, m' * \rho_\delta)^2$ by $\|m - m'\|_{-s, 2}^2$, up to a multiplicative constant, but this is exactly what is done in the first step.

Step 3. The formula for the derivative may be found in [CD18b, Chapter 5]. □

4.3. Regularization by sup-convolution in H^{-s} . We now turn to another regularization method that is key in our approach as it is shown to preserve the viscosity sub-solution property (see Proposition 5.1 below).

Proposition 4.3. *Let $\Phi : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ be a function which is C_L -Lipschitz and C_S -semiconcave with respect to H^{-s} . For $\epsilon > 0$, let $\Phi^\epsilon : H^{-s} \rightarrow \mathbb{R}$ be defined by*

$$\Phi^\epsilon(q) := \sup_{m \in \mathcal{P}(\mathbb{T}^d)} \left\{ \Phi(m) - \frac{1}{2\epsilon} \|q - m\|_{-s}^2 \right\}. \quad (4.1)$$

Then for all $\epsilon < \frac{1}{2C_S}$, we have

(1) For all $m \in \mathcal{P}(\mathbb{T}^d)$,

$$0 \leq \Phi^\epsilon(m) - \Phi(m) \leq 2C_L^2 \epsilon.$$

(2) $\Phi^\epsilon \in \mathcal{C}^1(H^{-s})$, and we have (using the same notation as in (2.24))

$$[\Phi^\epsilon]_{\mathcal{C}^{1,1}(H^{-s})} \leq \left(\frac{1}{\epsilon} \vee (2C_S) \right). \quad (4.2)$$

(3) For $m \in \mathcal{P}(\mathbb{T}^d)$, we have (using the same notation as in (2.23))

$$\nabla_{-s} \Phi^\epsilon(m) = \frac{1}{\epsilon} (m_\epsilon - m),$$

where m_ϵ is the unique element of $\mathcal{P}(\mathbb{T}^d)$ such that

$$\Phi^\epsilon(m) = \Phi(m_\epsilon) - \frac{1}{2\epsilon} \|m_\epsilon - m\|_{-s}^2.$$

Moreover, we have $\|m - m_\epsilon\|_{-s} \leq 2C_L \epsilon$.

(4) For all $m_1, m_2 \in \mathcal{P}(\mathbb{T}^d)$, we have

$$|\Phi^\epsilon(m_1) - \Phi^\epsilon(m_2)| \leq 2C_L \|m_1 - m_2\|_{-s}.$$

Proof. Claim (1). The fact that $\Phi^\epsilon \geq \Phi$ on $\mathcal{P}(\mathbb{T}^d)$ is clear from the definition. For the other inequality, let us fix $m \in \mathcal{P}(\mathbb{T}^d)$ and let $m^\epsilon \in \mathcal{P}(\mathbb{T}^d)$ be a maximizer in the definition of Φ^ϵ , i.e.

$$\Phi^\epsilon(m) = \Phi(m^\epsilon) - \frac{1}{2\epsilon} \|m - m^\epsilon\|_{-s}^2. \quad (4.3)$$

Notice that a maximizer does exist since $\mathcal{P}(\mathbb{T}^d)$ is compact for the weak topology and the function $\Phi(\cdot) - \|\cdot - m\|_{-s}^2/2$ is, for a fixed m , continuous for the weak topology (which is easy to check since the Fourier coefficients are continuous for the weak topology and s is here assumed to be greater than $d/2$).

Since $\Phi^\epsilon(m) \geq \Phi(m)$, rearranging (4.3) gives

$$\frac{1}{2\epsilon} \|m - m^\epsilon\|_{-s}^2 \leq \Phi(m^\epsilon) - \Phi(m) \leq C_L \|m - m^\epsilon\|_{-s}.$$

We deduce that

$$\|m - m^\epsilon\|_{-s} \leq 2C_L \epsilon.$$

Coming back to the definition (4.3) of m^ϵ we get

$$\begin{aligned} \Phi^\epsilon(m) - \Phi(m) &= \Phi(m^\epsilon) - \Phi(m) - \frac{1}{2\epsilon} \|m - m^\epsilon\|_{-s}^2 \\ &\leq \Phi(m^\epsilon) - \Phi(m) \leq C_L \|m - m^\epsilon\|_{-s} \leq 2C_L^2 \epsilon. \end{aligned}$$

Claim (2). We first notice that, for every $m \in \mathcal{P}(\mathbb{T}^d)$, the map

$$q \mapsto \Phi(m) - \frac{1}{2\epsilon} \|q - m\|_{-s}^2 + \frac{1}{2\epsilon} \|q\|_{-s}^2$$

is linear in q (which is precisely where the Hilbertian structure of H^{-s} comes in) and therefore

$$q \mapsto \Phi^\epsilon(q) + \frac{1}{2\epsilon} \|q\|_{-s}^2 = \sup_{m \in \mathcal{P}(\mathbb{T}^d)} \left\{ \Phi(m) - \frac{1}{2\epsilon} \|q - m\|_{-s}^2 + \frac{1}{2\epsilon} \|q\|_{-s}^2 \right\}$$

is convex. That is, Φ^ϵ is $\frac{1}{\epsilon}$ semi-convex over H^{-s} for any $\epsilon > 0$.

On the other hand, it is straightforward to check that

$$(m, q) \mapsto \Phi(m) - \frac{1}{2\epsilon} \|m - q\|_{-s}^2 - C_S \|q\|_{-s}^2$$

is concave over $\mathcal{P}(\mathbb{T}^d) \times H^{-s}$ as soon as $\epsilon \leq \frac{1}{2C_S}$. In this case, a small adaptation of a lemma in [LL86] is enough to deduce that $\Phi^\epsilon - C_S \|\cdot\|_{-s}^2$ is concave. That is, for $\epsilon \leq \frac{1}{2C_S}$, we have that Φ^ϵ is $2C_S$ semi-concave. As a consequence, and as explained in [LL86], (4.2) follows. In particular, Φ^ϵ is continuously differentiable.

Claim (3). The fact that $\nabla_{-s} \Phi^\epsilon(m) = \frac{1}{\epsilon}(m_\epsilon - m)$ is straightforward since, by definition of m^ϵ , m is a maximum of $H^{-s} \ni q \mapsto \Phi(m^\epsilon) - \frac{1}{2\epsilon} \|q - m^\epsilon\|_{-s}^2 - \Phi^\epsilon(q)$ (the function is non-positive and matches 0 at m). The estimate on $m_\epsilon - m$ was already obtained in the proof of Claim (1).

Claim (4). Lipschitz regularity for Φ^ϵ follows easily from Claims (2) and (3), since they show that Φ^ϵ is \mathcal{C}^1 on H^{-s} and its gradient satisfies $\|\nabla_{-s} \Phi^\epsilon\|_{-s} = \frac{1}{\epsilon} \|m_\epsilon - m\|_{-s} \leq 2C_L$ on the convex subset $\mathcal{P}(\mathbb{T}^d)$ of H^{-s} .

□

One specificity of the sup-convolution operation introduced in (4.1) is that the supremum is just taken on a tiny subset of the Hilbert space H^{-s} . Indeed, it is easy to check that $\mathcal{P}(\mathbb{T}^d)$ has an empty interior w.r.t. to the H^{-s} norm. Quite surprisingly, this creates substantial difficulty to study the regularity properties of Φ^ϵ that may be possibly inherited from those of Φ . When the maximization supporting the sup-convolution is taken over the whole space, there is no difficulty for, say, transferring any uniform continuity property from Φ (w.r.t. any arbitrary topology, possibly different from the H^{-s} -topology) to Φ^ϵ . Basically, a mere linear change of variable in the definition of the sup-convolution suffices. This fact is well reported in the paper [LL86]. However, things become more subtle when the argument in the sup-convolution is taken in a strict subset of the whole Hilbert space (here H^{-s}), as it is the case in (4.1). Typically, issues may arise if maximizers are located at the boundary (here the boundary of $\mathcal{P}(\mathbb{T}^d)$), which becomes very likely if the set over which the maximization is performed is tiny. This is the purpose of the next proposition to address this problem and to give sufficient conditions on the measure m such that m_ϵ is ‘inside’ the space of probability measures (for some topology) and to deduce subsequently further regularity properties on Φ^ϵ in the neighborhood of m . One additional subtlety is that we do so when Φ is in fact replaced by Φ^δ , as given by the previous Lemma 4.2.

Proposition 4.4. *For two constants c_1 and c_2 , let $\Phi : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ be c_1 -Lipschitz continuous and c_2 -semi-concave with respect to the d_1 -Wasserstein distance.*

For a given $s > 0$ and for any $\delta, \epsilon > 0$, let (with the same notation as in the statement of Lemma 4.2 and, in particular, with the same choice for ρ),

$$\Phi^{\delta, \epsilon}(m_0) := \sup_{m \in \mathcal{P}(\mathbb{T}^d)} \left\{ \Phi^\delta(m) - \frac{1}{2\epsilon} \|m - m_0\|_{-s}^2 \right\}, \quad m_0 \in \mathcal{P}(\mathbb{T}^d).$$

Then, for any $\eta > 0$, there exists a constant $\gamma(d, \rho, s, \eta) > 0$ such that, whenever $\delta^{2s-2} > 2c_2\Gamma(d, \rho, s)\epsilon$ (with $\Gamma(d, \rho, s)$ as in the statement of Lemma 4.2) and under the condition

$$m_0 \geq \frac{c_1\epsilon\gamma(d, \rho, s, \eta)}{\delta^{(2s+d/2+\eta-1)}} \text{Leb},$$

the (unique) maximizer $m_{\delta, \epsilon}$ in the definition of $\Phi^{\delta, \epsilon}$ satisfies

$$\|m_{\delta, \epsilon} - m_0\|_{L^\infty} \leq \frac{c_1\gamma(d, \rho, s, \eta)\epsilon}{\delta^{(2s+d/2+\eta-1)}}.$$

Moreover,

$$\left\| D_x \frac{\delta\Phi^{\delta, \epsilon}}{\delta m}(m_0, \cdot) \right\|_{L^\infty} \leq c_1.$$

Remark 4.5. We note that here and in the rest of the paper we use the notation $m \geq c\text{Leb}$ to mean that $m(A) \geq c\text{Leb}(A)$ for all Borel sets $A \subset \mathbb{T}^d$. Equivalently, the density of the absolutely continuous part of m is lower bounded by c .

Proof. The proof of Proposition 4.4 relies on a series of steps.

Step (1). The very preliminary one is to mollify the function Φ by means of the mollification procedure introduced in Lemma 4.1. In turn, we can define $\Phi^{n, \delta}$ and next $\Phi^{n, \delta, \epsilon}$ by replacing Φ by Φ^n in the definitions of the latter functions (with the same choice of ρ as in Lemma 4.1). By Lemmas 4.1 and 4.2, the function $\Phi^{n, \delta}$ is continuously differentiable (continuity w.r.t. m being understood for d_1) and, for any $m \in \mathcal{P}(\mathbb{T}^d)$,

$$\left\| \frac{\delta\Phi^{n, \delta}}{\delta m}(m, \cdot) \right\|_s \leq \frac{c_1\Gamma(d, \rho, s)}{\delta^{s-1}}. \quad (4.4)$$

The main idea is to show that, for any $n \geq 1$ and under the lower bound for m_0 prescribed in the statement, the equation

$$m - m_0 = \epsilon \left(\frac{\delta \Phi^{n,\delta}}{\delta m}(m, \cdot) \right)^* \quad (4.5)$$

has a solution m_n and that this solution identifies with the maximizer $m_{n,\delta,\epsilon}$ in the definition of $\Phi^{n,\delta,\epsilon}$. Using (2.2), we have that for any $\mathbf{k} \in \mathbb{Z}^d$,

$$\overline{\left(\frac{\delta \Phi^{n,\delta}}{\delta m}(m, \cdot) \right)^*}^{\mathbf{k}} = \left(1 + \sum_{i=1}^d |k^i|^{2s} \right) \overline{\frac{\delta \Phi^{n,\delta}}{\delta m}(m, \cdot)}^{\mathbf{k}}. \quad (4.6)$$

When $\mathbf{k} = 0$, the left-hand side is equal to 0, courtesy of our choice of a normalization for the flat derivative. In particular, the dual element belongs to

$$H_0^{-s}(\mathbb{T}^d) := \left\{ q \in H^{-s}(\mathbb{T}^d) : \hat{q}^0 = 0 \right\}.$$

By (4.6) and then (4.4), we obtain, for any real $r \geq 0$,

$$\begin{aligned} \left\| \left(\frac{\delta \Phi^{n,\delta}}{\delta m}(m, \cdot) \right)^* \right\|_r^2 &= \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{0\}} \left(1 + \sum_{i=1}^d |k^i|^{2r} \right) \left(\left| \overline{\left(\frac{\delta \Phi^{n,\delta}}{\delta m}(m, \cdot) \right)^*}^{\mathbf{k}} \right|^2 \right) \\ &= \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{0\}} \left(\left(1 + \sum_{i=1}^d |k^i|^{2s} \right)^2 \left(1 + \sum_{i=1}^d |k^i|^{2r} \right) \left| \overline{\frac{\delta \Phi^{n,\delta}}{\delta m}(m, \cdot)}^{\mathbf{k}} \right|^2 \right) \\ &\leq C(d, r, s) \left\| \frac{\delta \Phi^{n,\delta}}{\delta m}(m, \cdot) \right\|_{2s+r}^2 \leq \frac{C(d, r, s) c_1^2 \Gamma^2(d, \rho, 2s+r)}{\delta^{2(2s+r-1)}}, \end{aligned} \quad (4.7)$$

where $C(d, r, s)$ is a constant only depending on d, r and s .

When $r = d/2 + \eta$, Sobolev's embedding implies

$$\left\| \left(\frac{\delta \Phi^{n,\delta}}{\delta m}(m, \cdot) \right)^* \right\|_{L^\infty} \leq \gamma(d, \eta) \left\| \left(\frac{\delta \Phi^{n,\delta}}{\delta m}(m, \cdot) \right)^* \right\|_{d/2+\eta} \leq \frac{c_1 \gamma(d, \rho, s, \eta)}{\delta^{2s+d/2+\eta-1}}, \quad (4.8)$$

for two (new) constants $\gamma(d, \eta)$ and $\gamma(d, \rho, s, \eta)$.

Step (2). We now address the solvability of the equation (4.5). The point is to apply Schauder's fixed point theorem in the subset

$$\mathcal{C} = \left\{ m \in \mathcal{P}(\mathbb{T}^d) : \|m - m_0\|_{L^\infty} \leq \frac{c_1 \epsilon \gamma(d, \rho, s, \eta)}{\delta^{2s+d/2+\eta-1}} \right\},$$

seen as a convex subset of $1 + H_0^{-s}(\mathbb{T}^d)$ (where 1 is seen as the constant function, equal to 1). In the above right-hand side, neither m nor m_0 is required to be in $L^\infty(\mathbb{T}^d)$ (i.e., to have a bounded density) but the difference has to be. The L^∞ norm of $m - m_0$ can be expressed as

$$\|m - m_0\|_{L^\infty} = \sup_{f \in H^s(\mathbb{T}^d) : \|f\|_{L^1} \leq 1} \int_{\mathbb{T}^d} f(x) d(m - m_0)(x). \quad (4.9)$$

Notably, \mathcal{C} is closed under $\|\cdot\|_{-s}$. Indeed, $\mathcal{P}(\mathbb{T}^d)$ is stable under $\|\cdot\|_{-s}$ and convergence of a sequence of probability measures under $\|\cdot\|_{-s}$ is equivalent to weak convergence. The latter is a direct consequence of Prokhorov's theorem: any sequence of probability measures converging in $\|\cdot\|_{-s}$ has a weakly converging subsequence and the limits should obviously coincide. Moreover, by a lower semi-continuity argument and by (4.9), the condition on $\|m - m_0\|_{L^\infty}$ in the definition of \mathcal{C} can also be shown to be stable under $\|\cdot\|_{-s}$ (and also under weak convergence).

The same reasoning shows that \mathcal{C} is compact under $\|\cdot\|_{-s}$.

Next, we observe that, whenever

$$m_0 \geq \frac{c_1 \epsilon \gamma(d, \rho, s, \eta)}{\delta^{2s+d/2+\eta-1}} \text{Leb}, \quad (4.10)$$

the set \mathcal{C} can be merely described as

$$\mathcal{C} = \left\{ m \in 1 + H_0^{-s}(\mathbb{T}^d) : \|m - m_0\|_{L^\infty} \leq \frac{c_1 \epsilon \gamma(d, \rho, s, \eta)}{\delta^{2s+d/2+\eta-1}} \right\}. \quad (4.11)$$

It suffices to check that any m as in the right-hand side is a probability measure. Obviously, m is a finite signed measure with total mass one. The point is to prove that, for any non-negative function $f \in H^s(\mathbb{T}^d)$, $\int_{\mathbb{T}^d} f(x) dm(x) \geq 0$.

Clearly, by (4.10) and by definition of \mathcal{C} ,

$$\begin{aligned} \int_{\mathbb{T}^d} f(x) dm(x) &= \int_{\mathbb{T}^d} f(x) dm_0(x) + \int_{\mathbb{T}^d} f(x) d(m - m_0)(x) \\ &\geq \frac{c_1 \epsilon \gamma(d, \rho, s, \eta)}{\delta^{2s+d/2+\eta-1}} \|f\|_{L^1} - \|f\|_{L^1} \|m - m_0\|_{L^\infty} \\ &\geq \frac{c_1 \epsilon \gamma(d, \rho, s, \eta)}{\delta^{2s+d/2+\eta-1}} \|f\|_{L^1} - \frac{c_1 \epsilon \gamma(d, \rho, s, \eta)}{\delta^{2s+d/2+\eta-1}} \|f\|_{L^1} \geq 0. \end{aligned}$$

The next step is to consider the map

$$\phi : m \in \mathcal{C} \mapsto m_0 + \epsilon \left(\frac{\delta \Phi^{n,\delta}}{\delta m}(m, \cdot) \right)^*,$$

under the assumption (4.10). We claim that \mathcal{C} is stable by ϕ . Indeed, by (4.8), we get, for any $m \in \mathcal{P}(\mathbb{T}^d)$ (which contains \mathcal{C}),

$$\|\phi(m) - m_0\|_\infty \leq \epsilon \left\| \left(\frac{\delta \Phi^{n,\delta}}{\delta m}(m, \cdot) \right)^* \right\|_{L^\infty} \leq \frac{c_1 \epsilon \gamma(d, \rho, s, \eta)}{\delta^{(2s+d/2+\eta-1)}},$$

which, by (4.11), says that $\phi(m)$ indeed belongs to \mathcal{C} .

In order to apply Schauder's theorem, it remains to show that ϕ is continuous from \mathcal{C} into itself, when the latter is equipped with $\|\cdot\|_{-s}$. By (4.7) (with $r = d/2 + \eta$), the Fourier coefficients of $([\delta \Phi^{n,\delta}/\delta m](m, \cdot))^*$ are dominated by a summable sequence and, in turn, it suffices to prove that each Fourier coefficient

$$\widehat{\frac{\delta \Phi^{n,\delta}}{\delta m}(m, \cdot)}^k$$

is continuous with respect to m for $\|\cdot\|_{-s}$. We already know that it is continuous with respect to m for d_1 (since $\Phi^{n,\delta}$ is continuously differentiable). We then recall that convergence in $\|\cdot\|_{-s}$ implies weak convergence and, in turn, convergence in d_1 (by compactness of \mathbb{T}^d).

Step (3). By the second step, we know that (4.5) has a fixed point m_n . We claim that this fixed point is in fact a strict maximizer of the function

$$\Psi : m \in \mathcal{P}(\mathbb{T}^d) \mapsto \Phi^{n,\delta}(m) - \frac{1}{2\epsilon} \|m - m_0\|_{-s}^2.$$

Indeed, we observe that, for any two $m, m' \in \mathcal{P}(\mathbb{T}^d)$,

$$\frac{d}{d\lambda} \Psi(\lambda m' + (1-\lambda)m) \Big|_{\lambda=0} = \int_{\mathbb{T}^d} \frac{\delta \Phi^{n,\delta}}{\delta m}(m, x) d(m' - m)(x) - \frac{1}{\epsilon} \langle m' - m, m - m_0 \rangle_{-s}.$$

By the definition of the dual element,

$$\frac{d}{d\lambda} \Psi(\lambda m' + (1-\lambda)m) \Big|_{\lambda=0} = \left\langle m' - m, \left(\frac{\delta \Phi^{n,\delta}}{\delta m}(m, \cdot) \right)^* \right\rangle_{-s} - \frac{1}{\epsilon} \langle m' - m, m - m_0 \rangle_{-s}.$$

In particular, choosing m as m_n (and then replacing m' by m), we get

$$\frac{d}{d\lambda}\Psi\left(\lambda m + (1-\lambda)m_n\right)|_{\lambda=0} = 0, \quad m \in \mathcal{P}(\mathbb{T}^d), \quad (4.12)$$

which shows that m_n is critical point of Ψ .

We then argue by concavity to prove that m_n is the unique maximizer of Ψ . Indeed, using the semi-concavity property of $\Phi^{n,\delta}$ (see Lemma 4.2), we get, for any $m, m' \in \mathcal{P}(\mathbb{T}^d)$,

$$\begin{aligned} \Psi\left(\lambda m' + (1-\lambda)m\right) &= \Phi^{n,\delta}\left(\lambda m' + (1-\lambda)m\right) - \frac{1}{2\epsilon}\|\lambda(m' - m_0) + (1-\lambda)(m - m_0)\|_{-s}^2 \\ &\geq \lambda\Phi^{n,\delta}(m') + (1-\lambda)\Phi^{n,\delta}(m) - \frac{\lambda}{2\epsilon}\|m'\|_{-s}^2 - \frac{1-\lambda}{2\epsilon}\|m\|_{-s}^2 \\ &\quad - \frac{c_2\Gamma(d, \rho, s)}{2\delta^{2(s-1)}}\lambda(1-\lambda)\|m - m'\|_{-s}^2 + \frac{1}{2\epsilon}\lambda(1-\lambda)\|m - m'\|_{-s}^2 \\ &\geq \lambda\Psi(m') + (1-\lambda)\Psi(m) + c\lambda(1-\lambda)\|m - m'\|_{-s}^2, \end{aligned}$$

with $c := 1/\epsilon - c_2\Gamma(d, \rho, s)/\delta^{2(s-1)} > 0$. And then, choosing m as m_n (and then replacing m' by m), we obtain, for any $m \in \mathcal{P}(\mathbb{T}^d)$,

$$\Psi(m) - \Psi(m_n) \leq \frac{1}{\lambda}\left[\Psi\left(\lambda m + (1-\lambda)m_n\right) - \Psi(m_n)\right] - c(1-\lambda)\|m - m_n\|_{-s}^2.$$

Letting λ tend to 0 and invoking (4.12), we deduce that m_n is indeed the unique maximizer of Ψ . Below, we thus write $m_{n,\delta,\epsilon}$ for m_n . The conclusion is that the maximizer of Ψ satisfies

$$\|m_{n,\delta,\epsilon} - m_0\|_{L^\infty} \leq \frac{c_1\epsilon\gamma(d, \rho, s, \eta)}{\delta^{2s+d/2+\eta-1}}, \quad (4.13)$$

under the lower bound (4.10). Also, from the fixed point identity

$$\left(\frac{\delta\Phi^{n,\delta}}{\delta m}(m_{n,\delta,\epsilon}, \cdot)\right)^* = \frac{m_{n,\delta,\epsilon} - m_0}{\epsilon},$$

we deduce that

$$\left(\frac{m_{n,\delta,\epsilon} - m_0}{\epsilon}\right)^* = \frac{\delta\Phi^{n,\delta}}{\delta m}(m_{n,\delta,\epsilon}, \cdot). \quad (4.14)$$

In particular, $[(m_{n,\delta,\epsilon} - m_0)/\epsilon]^*$ is a smooth function and (4.14) says that, as such, it is c_1 -Lipschitz continuous (on \mathbb{T}^d) (the x -Lipschitz property of $\frac{\delta\Phi^{n,\delta}}{\delta m}(m_{n,\delta,\epsilon}, \cdot)$ follows from the μ -Lipschitz property of $\Phi^{n,\delta}$, see for instance [CD18b, Propositions 5.33 and 5.36]). Importantly, $[(m_{n,\delta,\epsilon} - m_0)/\epsilon]^*$ has a zero mean (which can be shown by computing the zero Fourier mode) and is thus bounded by a constant that only depends on c_1 and d .

Step (4). We now let n tend to ∞ . We recall from Lemmas 4.1 and 4.2 that $(\Phi^{n,\delta})_{n \geq 1}$ converges uniformly to Φ^δ .

Obviously, the sequence $(m_{n,\delta,\epsilon})_{n \geq 1}$ is compact (for the weak convergence and also for d_1 and $\|\cdot\|_{-s}$) and passing to the limit in the inequality

$$\Phi^{n,\delta}(m_{n,\delta,\epsilon}) - \frac{1}{2\epsilon}\|m_{n,\delta,\epsilon} - m_0\|_{-s}^2 \geq \Phi^{n,\delta}(m) - \frac{1}{2\epsilon}\|m - m_0\|_{-s}^2, \quad m \in \mathcal{P}(\mathbb{T}^d),$$

we deduce that any weak limit point is a minimizer of the right-hand side. Recalling that $\delta^{2s-2} > 2c_2\Gamma(d, \rho, s)\epsilon$, using the semi-concavity of Φ^δ respect to $\|\cdot\|_{-s}$ given by Lemma (4.2) as well as Claim (3) in Proposition (4.3) we deduce that the only possible limit is $m_{\delta,\epsilon}$ the unique maximum in the definition of $\Phi^{\delta,\epsilon}(m_0)$. In particular the whole sequence $(m_{n,\delta,\epsilon})_{n \geq 0}$ converges to $m_{\delta,\epsilon}$.

Passing to the limit in (4.13) (which is possible by the same argument as the one used in the first step of the proof), we get that

$$\|m_{\delta,\epsilon} - m_0\|_{L^\infty} \leq \frac{c_1 \epsilon \gamma(d, \rho, s, \eta)}{\delta^{2s+d/2+\eta-1}}, \quad (4.15)$$

under the lower bound (4.10).

Lastly, we have that, for any $q \in H^{-s}(\mathbb{T}^d)$,

$$\left\langle \left(\frac{m_{n,\delta,\epsilon} - m_0}{\epsilon} \right)^*, q \right\rangle \xrightarrow{n \rightarrow \infty} \left\langle \left(\frac{m_{\delta,\epsilon} - m_0}{\epsilon} \right)^*, q \right\rangle, \quad (4.16)$$

as n tends to ∞ . As the functions $([(m_{n,\delta,\epsilon} - m_0)/\epsilon]^*)_{n \geq 1}$ in the left-hand side are uniformly bounded and c_1 -Lipschitz-continuous, they do converge in sup norm, which shows that $[(m_{\delta,\epsilon} - m_0)/\epsilon]^*$ is c_1 -Lipschitz continuous. It remains to recall from the analysis of the sup-convolution (see item (3) in Proposition 4.3) that

$$\left(\frac{\delta \Phi^{\delta,\epsilon}}{\delta m}(m_0, \cdot) \right)^* = \frac{m_{\delta,\epsilon} - m_0}{\epsilon},$$

which shows that

$$\left(\frac{m_{\delta,\epsilon} - m_0}{\epsilon} \right)^* = \frac{\delta \Phi^{\delta,\epsilon}}{\delta m}(m_0, \cdot).$$

And then the right-hand side is c_1 -Lipschitz continuous, which proves the last expected result. \square

5. PROOFS OF THE ‘‘HARD INEQUALITIES’’

In this section we prove the ‘‘hard inequalities’’, i.e. the second inequality in Theorem 2.6 and the second inequality in Theorem 2.7. Recall that our main difficulty is that U is not smooth, so simply plugging the projection $U^N(t, \mathbf{x}) = U(t, m_{\mathbf{x}}^N)$ into the PDE (HJB(N)) gives us no information. Our strategy is therefore to attempt to regularize U while keeping track of the subsolution property. That is, we want to approximate U by functions which are smooth enough and which are *almost* subsolutions of (HJB(∞)). In the H^{-s} -regular case, this is accomplished in just one step, since sup-convolution in H^{-s} (as presented in Subsection 4.3) creates enough regularity for our purposes. In the d_1 -regular case, the analysis is much more involved, since in this step we must start by ‘‘upgrading’’ d_1 -regularity to H^{-s} -regularity by means of the regularization by mollification presented in Subsection 4.2, and only then apply a sup-convolution in H^{-s} . In each step the regularity of the approximation must be tracked, and the amount by which it fails to be a subsolution must be estimated.

5.1. Analysis of U^ϵ . This subsection is concerned with the H^{-s} -regular case, and so throughout this subsection Assumption 2.4 holds, and in particular an integer $s > d/2 + 2$ is fixed. For $\epsilon > 0$, we follow (4.1) in Proposition 4.3 and define the sup-convolution $U^\epsilon : [0, T] \times H^{-s} \rightarrow \mathbb{R}$ by

$$U^\epsilon(t, q) := \sup_{m \in \mathcal{P}(\mathbb{T}^d)} \left\{ U(t, m) - \frac{1}{2\epsilon} \|q - m\|_{-s}^2 \right\}. \quad (5.1)$$

The goal of this subsection is to understand three questions:

- How close is U^ϵ to U ?
- What are the regularity properties of U^ϵ ?
- By how much does U^ϵ fail to be a sub-solution of the PDE (HJB(∞))?

Some properties of U^ϵ can be already be inferred directly from Propositions 3.2 and 4.3, and the other relevant properties are summarized in the following proposition.

Proposition 5.1. *Let Assumption 2.4 hold and let U^ϵ be the sup-convolution defined in (5.1). Then, there is a constant $C > 0$ such that the following holds for all ϵ small enough:*

(1) For all $t_1, t_2 \in [0, T]$ and $m_1, m_2 \in \mathcal{P}(\mathbb{T}^d)$, we have

$$|U^\epsilon(t_1, m_1) - U^\epsilon(t_2, m_2)| \leq C(|t_1 - t_2| + \|m_1 - m_2\|_{-s}).$$

(2) The map $[0, T] \times H^{-s} \ni (t, q) \mapsto \nabla_{-s} U^\epsilon(t, q) \in H^{-s}$ is jointly continuous.

(3) U^ϵ satisfies

$$-\partial_t U^\epsilon(t, m) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta U^\epsilon}{\delta m}(t, m, x) m(dx) + \int_{\mathbb{T}^d} H(x, D_x \frac{\delta U^\epsilon}{\delta m}(t, m, x)) m(dx) \leq \mathcal{F}(m) + C\epsilon, \quad (5.2)$$

in the viscosity sense with test functions $\phi \in \mathcal{C}^1((0, T) \times H^{-s})$.

Remark 5.2. When we say that U^ϵ satisfies (5.2) in the viscosity sense with test functions in $\mathcal{C}^1((0, T) \times H^{-s})$, we mean that for any $\phi \in \mathcal{C}^1((0, T) \times H^{-s})$ and any $(t_0, m_0) \in (0, T) \times \mathcal{P}(\mathbb{T}^d)$ such that

$$U^\epsilon(t_0, m_0) - \phi(t_0, m_0) = \sup_{(t, q) \in (0, T) \times H^{-s}} \left\{ U^\epsilon(t, q) - \phi(t, q) \right\},$$

we have

$$-\partial_t \phi(t_0, m_0) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta \phi}{\delta m}(t_0, m_0, x) m_0(dy) + \int_{\mathbb{T}^d} H(x, D_x \frac{\delta \phi}{\delta m}(t_0, m_0, x)) m_0(dx) \leq \mathcal{F}(m_0) + C\epsilon. \quad (5.3)$$

Remark 5.3. Combining the bound on $[U^\epsilon(t, \cdot)]_{\mathcal{C}^{1,1}(H^{-s})}$ obtained from Proposition 4.3 with Proposition 2.12, we deduce that $m \mapsto D_m U^\epsilon(t, \cdot, x)$ is Lipschitz continuous over $\mathcal{P}(\mathbb{T}^d)$, uniformly in $x \in \mathbb{T}^d$ (but the Lipschitz constant depends on ϵ).

The rest of this subsection is devoted to proving Proposition 5.1. We start with a preliminary observation, which is that the inner product $\langle \cdot, \cdot \rangle_s$ (see Subsection 2.1) can be re-written as

$$\langle f, g \rangle_s = \langle f, g \rangle_{L^2} + (2\pi)^{-2s} \sum_{i=1}^d \langle D_{x^i}^s f, D_{x^i}^s g \rangle_{L^2}, \quad (5.4)$$

with $\langle f, g \rangle_{L^2} = \int_{\mathbb{T}^d} f g dx$ the usual L^2 inner product. Indeed, recalling that s is an integer, (5.4) can be seen directly from the fact that for $\mathbf{k} \in \mathbb{Z}^d$, $\widehat{D_{x^i}^s f}^{\mathbf{k}} = (2\pi i k_i)^s \widehat{f}^{\mathbf{k}}$. We now give two key lemmas.

Lemma 5.4. *Let Assumption 2.4 hold, and let $f \in H^{s+1}$. Then there is a constant $C = C(\|f\|_s)$ depending on f only through the norm $\|f\|_s$ such that*

$$|\langle H(\cdot, Df), f \rangle_s| \leq C.$$

Proof. By (5.4), we have

$$|\langle H(\cdot, Df), f \rangle_s| \leq |\langle H(\cdot, Df), f \rangle_{L^2}| + (2\pi)^{-2s} \sum_{i=1}^d |\langle D_{x^i}^s [H(\cdot, Df)], D_{x^i}^s f \rangle_{L^2}|.$$

Since $s > d/2 + 2$, $\|Df\|_{L^\infty} \leq C\|f\|_s$, and so clearly

$$|\langle H(\cdot, Df), f \rangle_{L^2}| \leq \|H(\cdot, Df)\|_{L^\infty}^{1/2} \|f\|_{L^2}^{1/2} \leq C(\|f\|_s),$$

where $C(\|f\|_s)$ is a non-decreasing function of $\|f\|_s$. To complete the proof, it suffices to fix $i \in \{1, \dots, d\}$ and show that

$$\left| \int_{\mathbb{T}^d} D_{x^i}^s [H(\cdot, Df)] D_{x^i}^s f dx \right| \leq C, \quad C = C(\|f\|_s).$$

In order to do so, we can use (a generalization of) the Faà Di Bruno formula to write $D_{x^i}^s [H(\cdot, Df)]$ as a sum of terms of the form

$$D_{x^i}^{s-k} D_{p^{i_1} \dots p^{i_q}} H(\cdot, Df) D_{x^i}^{j_1} D_{x^{i_1}} f \dots D_{x^i}^{j_q} D_{x^{i_q}} f, \quad (5.5)$$

with $k, q \in \{0, \dots, s\}$, $i_l \in \{1, \dots, d\}$, $j_l \in \mathbb{N}$ satisfying $\sum_{l=1}^q j_l = k$. So, to prove the Lemma, it in fact suffices to prove the following claim:

Claim: For any k, q , $(i_l)_{l=1, \dots, q}$ and $(j_l)_{l=1, \dots, q}$ as above, we have

$$\left| \int_{\mathbb{T}^d} \left(D_{x^i}^{s-k} D_{p^{i_1} \dots p^{i_q}} H(\cdot, Df) D_{x^i}^{j_1} D_{x^{i_1}} f \dots D_{x^i}^{j_q} D_{x^{i_q}} f \right) D_{x^i}^s f dx \right| \leq C, \quad C = C(\|f\|_s). \quad (5.6)$$

We now prove the claim by considering several cases.

Case 1 ($q = 0$): In the case $q = 0$ (and thus $k = 0$), the left-hand side of (5.6) becomes

$$\left| \int_{\mathbb{T}^d} D_{x^i}^s H(\cdot, Df) D_{x^i}^s f dx \right|,$$

and the estimate is easily proved using Cauchy-Schwarz and then Sobolev embedding (here we use $s > d/2 + 1$), and the fact that $H \in C^s$. Indeed, by Sobolev embedding, the argument Df inside $D_{x^i}^s H$ is bounded by a constant $C(\|f\|_s)$ (depending on $\|f\|_s$), and by continuity, $D_{x^i}^s H$ is bounded on the compact subset $\{(x, p) \in \mathbb{T}^d \times \mathbb{R}^d : |p| \leq C(\|f\|_s)\}$.

Case 2 ($q = 1$ and $k = s$): If $q = 1$ and $k = s$, then the left-hand side of (5.6) becomes

$$\left| \int_{\mathbb{T}^d} D_{p^{i_1}} H(\cdot, Df) D_{x^i}^s D_{x^{i_1}} f D_{x^i}^s f dx \right|,$$

so we can use integration by parts to estimate

$$\begin{aligned} \left| \int_{\mathbb{T}^d} D_{p^{i_1}} H(\cdot, Df) D_{x^i}^s D_{x^{i_1}} f D_{x^i}^s f dx \right| &= \frac{1}{2} \left| \int_{\mathbb{T}^d} D_{p^{i_1}} H(\cdot, Df) D_{x^{i_1}} |D_{x^i}^s f|^2 dx \right| \\ &= \left| \frac{1}{2} \int_{\mathbb{T}^d} D_{x^{i_1}} [D_{p^{i_1}} H(\cdot, Df)] |D_{x^i}^s f|^2 dx \right| \leq \frac{1}{2} \|D_{x^{i_1}} [D_{p^{i_1}} H(\cdot, Df)]\|_{L^\infty} \|f\|_s^2. \end{aligned}$$

To conclude (5.6) (and thus complete the proof), we need only to proceed as in Case 1 and notice that because of the smoothness of H and the fact that $s > d/2 + 2$, we have $\|D_{x^{i_1}} [D_{p^{i_1}} H(\cdot, Df)]\|_{L^\infty} \leq C(\|f\|_s)$.

Case 3 ($q = 1$ and $k < s$): In this case, the left-hand side of (5.6) becomes

$$\left| \int_{\mathbb{T}^d} D_{x^i}^{s-k} D_{p^{i_1}} H(\cdot, Df) D_{x^i}^k D_{x^{i_1}} f D_{x^i}^s f dx \right|.$$

Using Sobolev embedding and the smoothness of H as in the Case 2, we see that the following estimate suffices:

$$\int_{\mathbb{T}^d} |D_{x^i}^k D_{x^{i_1}} f| |D_{x^i}^s f| dx \leq \|D_{x^i}^k D_{x^{i_1}} f\|_{L^2} \|D_{x^i}^s f\|_{L^2} \leq C \|f\|_s^2,$$

where we use Cauchy-Schwarz and the fact that $k < s$.

Case 4 ($q > 1$) In this case, we start again by noticing that by Sobolev embedding,

$$\|D_{x^i}^{s-k} D_{p^{i_1} \dots p^{i_q}} H(\cdot, Df)\|_{L^\infty} \leq C(\|f\|_s). \quad (5.7)$$

Moreover, if for some $l \in \{1, \dots, q\}$, we have

$$\frac{1}{2} + \frac{j_l + 1}{d} < \frac{s}{d}, \quad (5.8)$$

then by Sobolev embedding again, we have $\|D_{x^i}^{j_l} D_{x^{i_1}} f\|_{L^\infty} \leq C \|f\|_s$. Without loss of generality, we can arrange the terms in (5.6) so that for some $r \in \{0, 1, \dots, q\}$ we have that j_l satisfies (5.8) for $l > r$, and j_l fails to satisfy (5.8) for $l \leq r$. So, ignoring the terms in (5.6) which are bounded in L^∞ by constants depending on $\|f\|_s$, we find that in order to complete the proof in Case 4, it suffices to show that

$$\int_{\mathbb{T}^d} |D_{x^i}^{j_1} D_{x^{i_1}} f \dots D_{x^i}^{j_r} D_{x^{i_r}} f D_{x^i}^s f| dx \leq C, \quad C = C(\|f\|_s), \quad (5.9)$$

where $q, r, (j_l)_{l=1, \dots, q}, (i_l)_{l=1, \dots, q}$ are as above and, importantly,

$$\frac{1}{2} - \frac{s}{d} \geq -\frac{j_l + 1}{d}, \quad l = 1, \dots, r. \quad (5.10)$$

If $r = 0$, (5.9) obviously holds, while if $r = 1$, the same argument as in Case 3 applies (since $q > 1$ implies that $j_r < s$). Thus it suffices to prove (5.9) when $r > 1$. To this end, suppose that we can find a collection of positive reals $(p_l)_{l=1, \dots, r}$ such that

- (1) $\sum_{l=1}^r \frac{1}{p_l} = \frac{1}{2}$
- (2) $\frac{1}{2} - \frac{s}{d} < \frac{1}{p_l} - \frac{j_l + 1}{d}$, for all $l = 1, \dots, r$.

Then we have

$$\begin{aligned} \int_{\mathbb{T}^d} |D_{x^i}^{j_1} D_{x^{i_1}} f \dots D_{x^i}^{j_r} D_{x^{i_r}} f D_{x^i}^s f| dx &\leq \|D_{x^i}^{j_1} D_{x^{i_1}} f\|_{L^{p_1}} \dots \|D_{x^i}^{j_r} D_{x^{i_r}} f\|_{L^{p_r}} \|D_{x^i}^s f\|_{L^2} \\ &\leq \|f\|_{j_1+1, p_1} \dots \|f\|_{j_r+1, p_r} \|f\|_s \leq C \|f\|_s^{q+1}, \end{aligned}$$

where $\|\cdot\|_{j,p}$ denotes the usual Sobolev norm on the space $W^{j,p}$ of functions with j weak derivatives each lying in L^p , and where the first inequality comes from the generalized Hölder's inequality and the last one comes from the fact that if $\frac{1}{2} - \frac{s}{d} < \frac{1}{p_j} - \frac{j_l + 1}{d}$, then $H^s \hookrightarrow W^{j_l+1, p_l}$ by Sobolev embedding. So, clearly to prove Claim 2, it suffices to prove that when $r > 1$, we can find $(p_l)_{l=1, \dots, r}$ satisfying the two properties listed above. To prove this, it in turn suffices to establish that when $r > 1$, we have

$$\sum_{l=1}^r \left(\frac{1}{2} - \frac{s}{d} + \frac{j_l + 1}{d} \right) < \frac{1}{2}, \quad (5.11)$$

since then the $(p_l)_{l=1, \dots, r}$ defined by the equation

$$\frac{1}{p_l} = \frac{1}{2} - \frac{s}{d} + \frac{j_l + 1}{d} + \frac{1}{r} \left(\frac{1}{2} - \sum_{m=1}^r \left(\frac{1}{2} - \frac{s}{d} + \frac{j_m + 1}{d} \right) \right), \quad l = 1, \dots, r, \quad (5.12)$$

will do the job. Indeed, the condition (5.10) guarantees that the numbers p_l defined in (5.12) are positive reals, and the condition (1) and (2) listed above are satisfied by construction. To prove

(5.11), we use the fact that $1 < r \leq q$, $\sum_{l=1}^r j_l \leq k \leq s$, and $s > \frac{d}{2} + 2$ to estimate

$$\begin{aligned} \sum_{l=1}^r \left(\frac{1}{2} - \frac{s}{d} + \frac{j_l + 1}{d} \right) &= r \left(\frac{1}{2} + \frac{1}{d} - \frac{s}{d} \right) + \frac{1}{d} \sum_{l=1}^r j_l \\ &\leq q \left(\frac{1}{2} + \frac{1}{d} - \frac{s}{d} \right) + \frac{s}{d} = (q-1) \left(\frac{1}{2} + \frac{1}{d} - \frac{s}{d} \right) + \frac{1}{2} + \frac{1}{d} \\ &< (q-1) \left(\frac{1}{2} + \frac{1}{d} - \frac{2}{d} - \frac{1}{2} \right) + \frac{1}{2} + \frac{1}{d} \\ &\leq (q-1) \left(-\frac{1}{d} \right) + \frac{1}{2} + \frac{1}{d} \leq \frac{1}{2}. \end{aligned}$$

This completes the proof of the estimate (5.9) and hence of Case 4. \square

We next state a lemma which follows easily from the non-positivity of the Laplacian on L^2 , together with the representation (5.4).

Lemma 5.5. *The Laplacian is non-positive on H^s . More precisely, if $f \in H^{s+2}$, then*

$$\langle \Delta f, f \rangle_s \leq 0.$$

We now use Lemmas 5.4 and 5.5 to prove Lemma 5.1.

Proof of Proposition 5.1. For Claim (1), the regularity of U^ϵ in m follows directly from Proposition 4.3, and regularity time is inherited directly from the regularity of U in t , since for any $t_1, t_2 \in [0, T]$ and $q \in H^{-s}$,

$$U^\epsilon(t_2, q) - U^\epsilon(t_1, q) \leq \sup_{m \in \mathcal{P}(\mathbb{T}^d)} \left\{ U(t_1, m) - U(t_2, m) \right\} \leq C|t_1 - t_2|,$$

where we used the uniform in $m \in \mathcal{P}(\mathbb{T}^d)$ Lipschitz regularity of $U(\cdot, m)$.

For Claim (2), we first note by Proposition by Claim (3) in Proposition 4.3, for all ϵ small enough and all $(t, q) \in [0, T] \times H^{-s}$ it holds

$$\nabla_{-s} U^\epsilon(t, q) = \frac{1}{\epsilon} (m^\epsilon(t, q) - q).$$

Claim (2) of the lemma will be proved if we can prove the continuity of the map $[0, T] \times H^{-s} \ni (t, q) \mapsto m^\epsilon(t, q) \in H^{-s}$. But for each fixed $t, q \mapsto \nabla_{-s} U^\epsilon(t, q)$ is Lipschitz by Claim 2 in Proposition 4.3, so in fact it suffices to show that for each fixed $q, t \mapsto m^\epsilon(t, q)$ is continuous. This can be established using the same compactness argument as in the fourth step in the proof of Proposition 4.4, so we omit the details.

We now turn to Claim (3), i.e., the subsolution estimate. Let us fix a smooth function $\phi = \phi(t, q) : (0, T) \times H^{-s} \rightarrow \mathbb{R}$ such that $U^\epsilon - \phi$ attains a maximum at $(t_0, m_0) \in (0, T) \times \mathcal{P}(\mathbb{T}^d)$. Let $m_\epsilon \in \mathcal{P}(\mathbb{T}^d)$ be the unique point such that

$$U^\epsilon(t_0, m_0) = U(t_0, m_\epsilon) - \frac{1}{2\epsilon} \|m_0 - m_\epsilon\|_{-s}^2,$$

and define $\psi(t, m) = \phi(t, m_0 - m_\epsilon + m)$. By a standard argument from the theory of viscosity solutions which applies equally well in this infinite-dimensional setting (see e.g. the ‘‘alternative Proof of Theorem 5.8’’ in the notes [Cal18]), we have the following two facts :

- (1) $U - \psi|_{(0, T) \times \mathcal{P}(\mathbb{T}^d)}$ attains a maximum at (t_0, m_ϵ) .
- (2) We have

$$\nabla_{-s} \psi(t_0, m_\epsilon) = \nabla_{-s} \phi(t_0, m_0) = \frac{1}{\epsilon} (m_\epsilon - m_0), \quad (5.13)$$

and in particular

$$\frac{\delta\psi}{\delta m}(t_0, m_\epsilon) = \frac{\delta\phi}{\delta m}(t_0, m_0) = \frac{1}{\epsilon}(m_\epsilon - m_0)^*.$$

Since $U - \psi$ has a maximum at (t_0, m_ϵ) and U is a subsolution of (HJB(∞)) (by Lemma 3.6), we deduce from point (1) the inequality,

$$\begin{aligned} & -\partial_t\phi(t_0, m_0) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta\phi}{\delta m}(t_0, m_0, x) m_\epsilon(dx) + \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta\phi}{\delta m}(t_0, m_0, x)\right) m_\epsilon(dx) \\ & = -\partial_t\psi(t_0, m_\epsilon) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta\psi}{\delta m}(t_0, m_\epsilon, x) m_\epsilon(dx) + \int_{\mathbb{T}^d} H\left(y, D_x \frac{\delta\psi}{\delta m}(t_0, m_\epsilon, x)\right) m_\epsilon(dx) \leq \mathcal{F}(m_\epsilon). \end{aligned}$$

Consequently, we have

$$\begin{aligned} & -\partial_t\phi(t_0, m_0) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta\phi}{\delta m}(t_0, m_0, x) m_0(dx) + \int_{\mathbb{T}^d} H\left(y, D_x \frac{\delta\phi}{\delta m}(t_0, m_0, x)\right) m_0(dx) \\ & \leq \mathcal{F}(m_0) + E_1 + E_2 + E_3, \end{aligned} \tag{5.14}$$

where

$$\begin{aligned} E_1 & := \int_{\mathbb{T}^d} \Delta_x \frac{\delta\phi}{\delta m}(t_0, m_0, x) (m_\epsilon - m_0)(dx), \\ E_2 & := \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta\phi}{\delta m}(t_0, m_0, x)\right) (m_0 - m_\epsilon)(dx), \\ E_3 & := \mathcal{F}(m_\epsilon) - \mathcal{F}(m_0). \end{aligned}$$

Using point (2) above, we see that

$$\begin{aligned} E_1 & = \epsilon \left\langle \Delta_x \frac{\delta\phi}{\delta m}(t_0, m_0, \cdot), \left(\frac{\delta\phi}{\delta m}(t_0, m_0, \cdot) \right)^* \right\rangle_{s, -s} \\ & = \epsilon \left\langle \Delta_x \frac{\delta\phi}{\delta m}(t_0, m_0, \cdot), \frac{\delta\phi}{\delta m}(t_0, m_0, \cdot) \right\rangle_s \leq 0, \end{aligned} \tag{5.15}$$

by Lemma 5.5 (which applies here since $m_\epsilon - m \in H^{-s+2}$, hence $\frac{\delta\phi}{\delta m}(t_0, m_0) = \epsilon(m_\epsilon - m)^* \in (H^{-s+2})^* = H^{s+2}$, as can be easily checked using the formula (2.2)). For E_2 , we use the same trick to write

$$\begin{aligned} E_2 & = \epsilon \left\langle H\left(\cdot, D_x \frac{\delta\phi}{\delta m}(t_0, m_0, \cdot)\right), \left(\frac{\delta\phi}{\delta m}(t_0, m_0, \cdot) \right)^* \right\rangle_{s, -s} \\ & = \epsilon \left\langle H\left(\cdot, D_x \frac{\delta\phi}{\delta m}(t_0, m_0, \cdot)\right), \frac{\delta\phi}{\delta m}(t_0, m_0, \cdot) \right\rangle_s \leq C\epsilon, \end{aligned} \tag{5.16}$$

where the last inequality follows from Lemma 5.4 together with the fact that Claim (3) in Proposition 4.3 shows that there is a constant C independent of ϵ and ϕ such that

$$\left\| \frac{\delta\phi}{\delta m}(t_0, m_0, \cdot) \right\|_s = \frac{1}{\epsilon} \|m_\epsilon - m_0\|_{-s} \leq C.$$

Finally for the third error term we simply use Lipschitz continuity of \mathcal{F} (see Remark 2.5) to estimate

$$E_3 \leq C \|m_\epsilon - m_0\|_{-s} \leq C\epsilon. \tag{5.17}$$

Combining (5.14), (5.15), (5.16), (5.17) gives

$$-\partial_t\phi(t_0, m_0) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta\phi}{\delta m}(t_0, m_0, x) m_0(dx) + \int_{\mathbb{T}^d} H\left(y, D_x \frac{\delta\phi}{\delta m}(t_0, m_0, x)\right) m_0(dx) \leq \mathcal{F}(m_0) + C\epsilon,$$

which completes the proof. \square

We close this subsection with a useful lemma, which, among other things, is used to show that the viscosity subsolution property in Claim (3) of Proposition 5.1 can be understood in a pointwise sense.

Lemma 5.6. *Let $\Phi = \Phi(t, m) : [0, T] \times H^{-s} \rightarrow \mathbb{R}$ such that $\sup_{0 \leq t \leq T} [\Phi(t, \cdot)]_{\mathcal{C}^{1,1}(H^{-s})} \leq C$ for some constant C , and $\nabla_{-s}\Phi$ is jointly continuous. Then, for any $(t_0, m_0) \in [0, T] \times \mathcal{P}(\mathbb{T}^d)$ such that the derivative $\partial_t \Phi(t_0, m_0)$ exists, the function Φ is (fully) differentiable at (t_0, m_0) . As a consequence, if Φ satisfies*

$$-\partial_t \Phi(t, m) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta \Phi}{\delta m}(t, m, x) m(dx) + \int_{\mathbb{T}^d} H\left(y, D_x \frac{\delta \Phi}{\delta m}(t, m, x)\right) m(dx) \leq \mathcal{F}(m) + C, \quad (5.18)$$

in the viscosity sense on $[0, T] \times \mathcal{P}(\mathbb{T}^d)$, with test functions in $\mathcal{C}^1((0, T) \times H^{-s})$ (see Remark 5.2), then (5.18) is satisfied at any such point (t_0, m_0) .

Proof. We first prove that Φ is differentiable at (t_0, m_0) . Using the bound on $[\Phi(t, \cdot)]_{\mathcal{C}^{1,1}(H^{-s})}$, the existence of $\partial_t \Phi(t_0, m_0)$, and the joint continuity of $\nabla_{-s}\Phi$ (see Proposition 5.1), we can write

$$\begin{aligned} \Phi(t, q) - \Phi(t_0, m_0) &= \Phi(t, q) - \Phi(t, m_0) + \Phi(t, m_0) - \Phi(t_0, m_0) \\ &= \langle \nabla_{-s}\Phi(t, m_0), q - m_0 \rangle_{-s} + \partial_t \Phi(t_0, m_0)(t - t_0) + |t - t_0| e_1(t) + \|m - m_0\| e_2(t, q) \\ &= \langle \nabla_{-s}\Phi(t_0, m_0), q - m_0 \rangle_{-s} + \partial_t \Phi(t_0, m_0)(t - t_0) \\ &\quad + \langle \nabla_{-s}\Phi(t, m_0) - \nabla_{-s}\Phi(t_0, m_0), q - m_0 \rangle_{-s} + |t - t_0| e_1(t) + \|q - m_0\| e_2(t, m) \\ &= \langle \nabla_{-s}\Phi(t_0, m_0), m - m_0 \rangle_{-s} + \partial_t \Phi(t_0, m_0)(t - t_0) \\ &\quad + |t - t_0| e_1(t) + \|m - m_0\| e_2(t, q) + e_3(t) \|q - m_0\|, \end{aligned}$$

where $e_1, e_3 : [0, T] \rightarrow \mathbb{R}$ and $e_2 : [0, T] \times H^{-s} \rightarrow \mathbb{R}$ are functions such that

$$e_1(t_n) \rightarrow 0, \quad e_2(t_n, q_n) \rightarrow 0, \quad e_3(t_n) \rightarrow 0$$

as $(t_n, q_n) \rightarrow (t_0, m_0)$. We conclude that $\Phi(t_0, m_0)$ is differentiable at (t_0, m_0) .

By classical arguments, see e.g. [Bar94] p.18², this implies that we can find $\Psi \in \mathcal{C}^1((0, T) \times H^{-s})$ which touches Φ from above at (t_0, m_0) . In particular, $\partial_t \Psi(t_0, m_0) = \partial_t \Phi(t_0, m_0)$ and $D_{-s}\Psi(t_0, m_0) = D_{-s}\Phi(t_0, m_0)$. By Proposition 5.1, this implies that Inequality (5.2) is satisfied at (t_0, m_0) . \square

5.2. Analysis of U^δ , $U^{\delta, \epsilon}$, and $U^{\delta, \epsilon, \lambda}$. This section is concerned with the d_1 -regular case. We fix throughout this section a real number s with

$$s > d/2 + 1.$$

For each $\delta, \epsilon > 0$, and $\lambda \in (0, 1)$, we define functions $U^\delta : [0, T] \times \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$, $U^{\delta, \epsilon}, U^{\delta, \epsilon, \lambda} : [0, T] \times \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$, by

$$U^\delta(t, m) = U(t, m * \rho_\delta), \quad (5.19)$$

with ρ_δ being defined as in Lemma 4.2,

$$U^{\delta, \epsilon}(t, q) = \sup_{m \in \mathcal{P}(\mathbb{T}^d)} \left\{ U^\delta(t, m) - \frac{1}{2\epsilon} \|m - q\|_{-s}^2 \right\}, \quad (5.20)$$

and finally

$$U^{\delta, \epsilon, \lambda}(t, q) = U^{\delta, \epsilon}(t, (1 - \lambda)q + \lambda \text{Leb}). \quad (5.21)$$

Ultimately, we would like to answer the following questions:

²In fact this reference only treats the finite dimensional case but the proof extends readily to any Hilbert space

- How close is $U^{\delta, \epsilon, \lambda}$ to U ?
- What are the regularity properties of $U^{\delta, \epsilon, \lambda}$?
- By how much does $U^{\delta, \epsilon, \lambda}$ fail to be a sub-solution of the PDE (HJB(∞))?

Of course, a preliminary step will be to answer the same questions for U^δ and then $U^{\delta, \epsilon}$.

Lemma 5.7. *There is a constant $C > 0$ independent of $\delta > 0$ such that, for any $\phi \in \mathcal{C}^1((0, T) \times H^{-s})$ and any $(t_0, m_0) \in (0, T) \times \mathcal{P}(\mathbb{T}^d)$ such that $m_0 \geq c \text{Leb}$ for some $c > 0$ and*

$$U^\delta(t_0, m_0) - \phi(t_0, m_0) = \sup_{(t, q) \in (0, T) \times \mathcal{P}(\mathbb{T}^d)} \left\{ U^\delta(t, m) - \phi(t, m) \right\},$$

we have

$$-\partial_t \phi(t_0, m_0) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta \phi}{\delta m}(t_0, m_0, x) m_0(dx) + \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta \phi}{\delta m}(t_0, m_0, x)\right) m_0(dx) \leq \mathcal{F}(m_0) + C\delta. \quad (5.22)$$

Proof. First we take ϕ , t_0 and m_0 as in the statement of the lemma as well as some $h > 0$ such that $t_0 + h \leq T$. We let

$$\alpha(t, x) = -D_p H\left(x, D_x \frac{\delta \phi}{\delta m}(t_0, m_0, x)\right), \quad (t, x) \in [t_0, t_0 + h] \times \mathbb{T}^d, \quad (5.23)$$

and m be the associated trajectory starting from (t_0, m_0) , ie solution to

$$\begin{cases} \partial_t m(t, x) + \text{div}_x(\alpha(t, x)m(t, x)) - \Delta_x m(t, x) = 0 & \text{in } (t_0, t_0 + h) \times \mathbb{T}^d \\ m_{t_0} = m_0. \end{cases} \quad (5.24)$$

In particular, since $\alpha \in \mathcal{C}^1([t_0, t_0 + h] \times \mathbb{T}^d)$, m_t has a density for all $t \in (t_0, t_0 + h]$. Now if we take such a pair (α, m) , then $(m_t^\delta := \rho_\delta * m_t)_{t \in [t_0, t_0 + h]}$ solves

$$\begin{cases} \partial_t m^\delta(t, x) + \text{div}_x(\alpha^\delta(t, x)m^\delta(t, x)) - \Delta_x m^\delta(t, x) = 0 & \text{in } (t_0, t_0 + h) \times \mathbb{T}^d, \\ m_{t_0}^\delta = m_0^\delta, \end{cases} \quad (5.25)$$

with

$$\alpha^\delta(t, x) := \frac{\rho_\delta * (\alpha(t, \cdot)m(t, \cdot))(x)}{\rho_\delta * m(t, \cdot)(x)} \mathbb{1}_{\{\rho_\delta * m_t(x) > 0\}},$$

noticing that $\rho_\delta * (\alpha(t, \cdot)m(t, \cdot))(x) = 0$ if $\rho_\delta * m(t, \cdot)(x) = 0$ and $\alpha(t, \cdot) \in L^1(\mathbb{T}^d; m_t)$. By dynamic programming it holds

$$\begin{aligned} U^\delta(t_0, m_0) &= U(t_0, \rho_\delta * m_0) \\ &\leq \int_{t_0}^{t_0+h} \int_{\mathbb{T}^d} L(x, \alpha^\delta(t, x)) m_t^\delta(dx) dt + \int_{t_0}^{t_0+h} \mathcal{F}(m_t^\delta) dt + U(t_0 + h, m_{t_0+h}^\delta), \end{aligned}$$

and, using the definition of U^δ and the fact that ϕ touches U^δ from above at m_0 we have

$$\begin{aligned} \phi(t_0, m_0) &\leq \int_{t_0}^{t_0+h} \int_{\mathbb{T}^d} L(x, \alpha^\delta(t, x)) m_t^\delta(dx) dt + \int_{t_0}^{t_0+h} \mathcal{F}(m_t^\delta) dt + U^\delta(t_0 + h, m_{t_0+h}^\delta) \\ &\quad + \phi(t_0, m_0) - U^\delta(t_0, m_0) \\ &\leq \int_{t_0}^{t_0+h} \int_{\mathbb{T}^d} L(x, \alpha^\delta(t, x)) m_t^\delta(dx) dt + \int_{t_0}^{t_0+h} \mathcal{F}(m_t^\delta) dt + \phi(t_0 + h, m_{t_0+h}^\delta). \end{aligned}$$

Using the joint convexity of $\mathcal{L} : (a, b) \in \mathbb{R}^d \times \mathbb{R}^+ \mapsto L(x, \frac{a}{b})b \in \mathbb{R}$ (set to be $+\infty$ if $b=0$), we can apply Jensen's inequality and deduce that, for all $(t, x) \in (t_0, t_0 + h] \times \mathbb{T}^d$,

$$\begin{aligned} L(x, \alpha^\delta(t, x)) m_t^\delta(x) &= \mathcal{L}\left((\alpha(t, \cdot)m_t(\cdot), m_t(\cdot)) * \rho_\delta(x)\right) \\ &\leq \mathcal{L}(\alpha(t, \cdot)m_t(\cdot), m_t(\cdot)) * \rho_\delta(x) = \rho_\delta * [L(x, \alpha(t, \cdot))m_t(\cdot)](x). \end{aligned} \quad (5.26)$$

As a consequence, we have

$$\begin{aligned}
\int_{t_0}^{t_0+h} \int_{\mathbb{T}^d} L(x, \alpha^\delta(t, x)) m_t^\delta(x) dx dt &\leq \int_{t_0}^{t_0+h} \int_{\mathbb{T}^d} \rho_\delta * [L(x, \alpha(t, \cdot)) m_t(\cdot)](x) dx dt \\
&= \int_{t_0}^{t_0+h} \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} \rho_\delta(x-y) L(y, \alpha(t, y)) m_t(dy) dx dt \\
&\quad + \int_{t_0}^{t_0+h} \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} \rho_\delta(x-y) [L(x, \alpha(t, y)) - L(y, \alpha(t, y))] m_t(dy) dx dt \\
&\leq \int_{t_0}^{t_0+h} \int_{\mathbb{T}^d} L(x, \alpha(t, x)) m_t(dx) dt + C(\|\alpha\|_{L^\infty}) \int_{t_0}^{t_0+h} \int_{\mathbb{T}^d} \rho_\delta(x-y) |x-y| m_t(dy) dx dt \\
&\leq \int_{t_0}^{t_0+h} \int_{\mathbb{T}^d} L(x, \alpha(t, x)) m_t(dx) dt + C(\|\alpha\|_{L^\infty}) \delta h,
\end{aligned}$$

where $C(\|\alpha\|_{L^\infty})$ denotes the Lipschitz constant of L in the variable x over the set $\mathbb{T}^d \times B_{\|\alpha\|_{L^\infty}}$. On the other hand, \mathcal{F} being Lipschitz continuous with respect to d_1 , inequality $|\mathcal{F}(\rho_\delta * m) - \mathcal{F}(m)| \leq C\delta$ holds for all $m \in \mathcal{P}(\mathbb{T}^d)$ and some $C > 0$ depending only on the Lipschitz constant of \mathcal{F} . Therefore we have,

$$\begin{aligned}
\phi(t_0, m_0) - \phi(t_0 + h, m_{t_0+h}) &\leq \int_{t_0}^{t_0+h} \int_{\mathbb{T}^d} L(x, \alpha(t, x)) m_t(dx) dt + \int_{t_0}^{t_0+h} \mathcal{F}(m_t) dt \\
&\quad + C(\|\alpha\|_{L^\infty}) \delta h + C\delta h.
\end{aligned} \tag{5.27}$$

Since α is given by (5.23) it holds

$$L(x, \alpha(t, x)) = -\alpha(t, x) \cdot D_x \frac{\delta\phi}{\delta m}(t_0, m_0, x) - H\left(x, D_x \frac{\delta\phi}{\delta m}(t_0, m_0, x)\right), \quad \forall (t, x) \in [t_0, t_0 + h] \times \mathbb{T}^d,$$

and therefore, using the equation satisfied by m we easily deduce, after dividing (5.27) by h and letting h tends to 0 that

$$\begin{aligned}
-\partial_t \phi(t_0, m_0) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta\phi}{\delta m}(t_0, m_0, x) m_0(dx) + \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta\phi}{\delta m}(t_0, m_0, x)\right) m_0(dx) \\
\leq \mathcal{F}(m_0) + C(\|\alpha\|_{L^\infty}) \delta + C\delta.
\end{aligned} \tag{5.28}$$

To conclude, it remains to prove that $C(\|\alpha\|_{L^\infty})$ is bounded independently from $\delta > 0$. We are going to show that

$$\left\| D_x \frac{\delta\phi}{\delta m}(t_0, m_0, \cdot) \right\|_{L^\infty} \leq L_U, \tag{5.29}$$

where L_U is a Lipschitz constant for U in the argument m with respect to d_1 . By Lemma 4.2, L_U also provides a bound for the Lipschitz constant of U^δ (in m w.r.t. d_1). To see that indeed we have (5.29) we note that because ϕ touches U^δ from above at (t_0, m_0) , we must have

$$\begin{aligned}
\phi(t_0, m_0) - \phi(t_0, m_0 + \epsilon D_{\vec{e}} g \text{Leb}) &\leq U^\delta(t_0, m_0) - U^\delta(t_0, m_0 + \epsilon D_{\vec{e}} g \text{Leb}) \\
&\leq L_U |\epsilon| \sup_{\|Df\|_{L^\infty} \leq 1} \int_{\mathbb{T}^d} D_{\vec{e}} g f dx = L_U |\epsilon| \sup_{\|Df\|_{L^\infty} \leq 1} \left(- \int_{\mathbb{T}^d} g Df dx \right) = L_U |\epsilon| \|g\|_{L^1}, \tag{5.30}
\end{aligned}$$

for each smooth $g : \mathbb{T}^d \rightarrow \mathbb{R}$ — where $D_{\vec{e}} g := D_x g \cdot \vec{e}$ denotes the derivative of g in the direction of some unit vector $\vec{e} \in \mathbb{R}^d$ — and each $\epsilon \in (-1, 1)$ with $|\epsilon|$ small enough so that $m_0 + \epsilon D_{\vec{e}} g \text{Leb} \in \mathcal{P}(\mathbb{T}^d)$

(which is indeed possible since $m_0 \geq c\text{Leb}$). Being ϕ in $C^1([0, T] \times H^{-s})$, it holds

$$\begin{aligned} \phi(t_0, m_0 + \epsilon D_{\bar{e}} g \text{Leb}) - \phi(t_0, m_0) &= \epsilon \int_{\mathbb{T}^d} \frac{\delta \phi}{\delta m}(t_0, m_0, x) D_{\bar{e}} g(x) dx + o(\epsilon) \\ &= -\epsilon \int_{\mathbb{T}^d} \left(D_x \frac{\delta \phi}{\delta m}(t_0, m_0, x) \cdot \bar{e} \right) g(x) dx + o(\epsilon), \end{aligned} \quad (5.31)$$

and therefore, by arbitrariness of ϵ small enough, we deduce from (5.30) that

$$\left| \int_{\mathbb{T}^d} \left(D_x \frac{\delta \phi}{\delta m}(t_0, m_0, x) \cdot \bar{e} \right) g(x) dx \right| \leq L_U \|g\|_{L^1},$$

for all smooth g , which easily gives (5.29). \square

We now turn to the relevant properties of $U^{\delta, \epsilon}$, as defined in (5.19).

Lemma 5.8. *For any $\eta > 0$ there are constants $c, C > 0$ (which can depend on η), such that for each $\delta > 0$, each $0 < \epsilon < c\delta^{2(s-1)}$ and each $(t_0, m_0) \in (0, T) \times \mathcal{P}(\mathbb{T}^d)$ such that $m_0 \geq C\epsilon\delta^{-(2s+d/2+\eta-1)}\text{Leb}$ and $U^{\delta, \epsilon}$ is differentiable at (t_0, m_0) , we have*

$$\begin{aligned} -\partial_t U^{\delta, \epsilon}(t_0, m_0) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta U^{\delta, \epsilon}}{\delta m}(t_0, m_0, x) m_0(dx) + \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta U^{\delta, \epsilon}}{\delta m}(t_0, m_0, x)\right) m_0(dx) \\ \leq \mathcal{F}(m_0) + C(\delta + \epsilon\delta^{(2s+d/2+\eta-1)}). \end{aligned} \quad (5.32)$$

Moreover, $\|D_x \frac{\delta U^{\delta, \epsilon}}{\delta m}(t_0, m_0, \cdot)\|_{L^\infty} \leq C$.

Proof. Being $U^{\delta, \epsilon}$ differentiable at (t_0, m_0) , there exists $\phi \in C^1((0, T) \times H^{-s})$ which touches $U^{\delta, \epsilon}$ from above at (t_0, m_0) and which satisfies

$$\partial_t \phi(t_0, m_0) = \partial_t U^{\delta, \epsilon}(t_0, m_0), \quad \nabla_{-s} \phi(t_0, m_0) = \nabla_{-s} U^{\delta, \epsilon}(t_0, m_0).$$

Now we let $m_{\delta, \epsilon} \in \mathcal{P}(\mathbb{T}^d)$ be the unique point such that

$$U^{\delta, \epsilon}(t_0, m_0) = U^\delta(t_0, m_{\delta, \epsilon}) - \frac{1}{2\epsilon} \|m_0 - m_{\delta, \epsilon}\|_{-s}^2,$$

and we follow the reasoning in Proposition 5.1 to conclude that because ϕ touches $U^{\delta, \epsilon}$ from above at (t_0, m_0) , the function ψ defined by

$$\psi(t, m) = \phi(t, m_0 - m_{\delta, \epsilon} + m)$$

touches U^δ from above at $(t_0, m_{\delta, \epsilon})$. In particular assuming that the constant C appearing in the statement of the lemma satisfies $C > c_1$ where c_1 is a Lipschitz constant for U with respect to d_1 we can infer from Proposition 4.4 that $m^{\epsilon, \delta} \geq c\text{Leb}$ for some $c > 0$. As a consequence we can apply Lemma 5.7 and deduce that

$$\begin{aligned} -\partial_t U^{\delta, \epsilon}(t_0, m_0) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta U^{\delta, \epsilon}}{\delta m}(t_0, m_0, x) m_{\delta, \epsilon}(dx) + \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta U^{\delta, \epsilon}}{\delta m}(t_0, m_0, x)\right) m_{\delta, \epsilon}(dx) \\ = -\partial_t \psi(t_0, m_{\delta, \epsilon}) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta \psi}{\delta m}(t_0, m_{\delta, \epsilon}, x) m_{\delta, \epsilon}(dx) + \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta \psi}{\delta m}(t_0, m_{\delta, \epsilon}, x)\right) m_{\delta, \epsilon}(dx) \\ \leq \mathcal{F}(m_\epsilon) + C\delta. \end{aligned}$$

Thus

$$\begin{aligned} -\partial_t U^{\delta, \epsilon}(t_0, m_0) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta U^{\delta, \epsilon}}{\delta m}(t_0, m_0, x) m_0(dx) + \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta U^{\delta, \epsilon}}{\delta m}(t_0, m_0, x)\right) m_0(dx) \\ \leq \mathcal{F}(m_0) + C'\delta + E_1 + E_2 + E_3, \end{aligned} \quad (5.33)$$

where $C' > 0$ depends only on the Lipschitz constants of U and \mathcal{F} with respect to d_1 and

$$\begin{aligned} E_1 &:= \int_{\mathbb{T}^d} \Delta_x \frac{\delta U^{\delta, \epsilon}}{\delta m}(t_0, m_0, x) d(m_{\delta, \epsilon} - m_0)(x), \\ E_2 &:= \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta U^{\delta, \epsilon}}{\delta m}(t_0, m_0, x)\right) d(m_0 - m_{\delta, \epsilon})(x), \\ E_3 &:= \mathcal{F}(m_{\delta, \epsilon}) - \mathcal{F}(m_0). \end{aligned}$$

Lemma 5.5 (together with Claim (3) in Proposition 4.3) shows that $E_1 \leq 0$ exactly as in the proof of Proposition 5.1. To bound E_2 , we use Proposition 4.4 to see that because $m_0 \geq C\epsilon\delta^{-(2s+d/2+\eta-1)}\text{Leb}$ (see Remark 4.5), we have

$$\|m_0 - m_{\delta, \epsilon}\|_{L^\infty} \leq C\delta^{-(2s+d/2+\eta-1)}\epsilon,$$

and also, as already mentioned, we have (from Proposition 4.4) $\|D_x \frac{\delta U^{\delta, \epsilon}}{\delta m}(t_0, m_0, \cdot)\|_{L^\infty} \leq C$. Combining these two inequalities easily gives

$$E_2 \leq C\epsilon\delta^{-(2s+d/2+\eta-1)}.$$

To bound E_3 , we use d_1 -Lipschitz continuity of \mathcal{F} to get

$$E_3 \leq Cd_1(m_0, m_{\delta, \epsilon}) \leq C\|m_0 - m_{\delta, \epsilon}\|_{L^\infty} \leq C\epsilon\delta^{-(2s+d/2+\eta-1)}.$$

Combining the estimates of E_1 , E_2 , and E_3 with (5.33) completes the proof. \square

Finally, we have the properties of $U^{\delta, \epsilon, \lambda}$ defined in (5.21).

Lemma 5.9. *For any $\eta > 0$ there are constants $c, C > 0$ which can depend on η , such that for each $\delta > 0$, each $0 < \epsilon < c\delta^{2(s-1)}$, and each $\lambda \in (0, 1/2)$ such that*

$$\lambda \geq C\epsilon\delta^{-(2s+d/2+\eta-1)},$$

we have

(1) $U^{\delta, \epsilon, \lambda}$ is $C\delta^{-(s-1)}$ -Lipschitz over $\mathcal{P}(\mathbb{T}^d)$ with respect to $\|\cdot\|_{-s}$ and $C\delta^{-2(s-1)}$ -semi-concave (again with respect to $\|\cdot\|_{-s}$) over all of H^{-s} . Moreover, for each $S \in (0, T)$, the function $U^{\delta, \epsilon, \lambda}$ is Lipschitz continuous in time, uniformly in $(\delta, \epsilon, \lambda)$, on $[0, S] \times \mathcal{P}(\mathbb{T}^d)$.

(2) For each $(t, m) \in [0, T] \times \mathcal{P}(\mathbb{T}^d)$,

$$|U^{\delta, \epsilon, \lambda}(t, m) - U(t, m)| \leq C(\lambda + \delta + \epsilon\delta^{-2(s-1)}).$$

(3) The map $[0, T] \times H^{-s} \ni (t, q) \mapsto \nabla_{-s} U^{\delta, \epsilon, \lambda}(t, q) \in H^{-s}$ is jointly continuous.

(4) For each $t \in [0, T]$, $U^{\delta, \epsilon}(t, \cdot)$ lies in $C^{1,1}(H^{-s})$ and satisfies

$$\sup_{0 \leq t \leq T} [U^{\delta, \epsilon, \lambda}(t, \cdot)]_{C^{1,1}(H^{-s})} \leq \frac{C}{\epsilon}.$$

(5) If $(t_0, m_0) \in (0, T) \times \mathcal{P}(\mathbb{T}^d)$ is such that $t \mapsto U^{\delta, \epsilon, \lambda}(t, m_0)$ is differentiable at t_0 then $U^{\delta, \epsilon, \lambda}$ has a (full) derivative at (t_0, m_0) and we have

$$\begin{aligned} & -\partial_t U^{\delta, \epsilon, \lambda}(t_0, m_0) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta U^{\delta, \epsilon, \lambda}}{\delta m}(t_0, m_0, x) m_0(dx) + \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta U^{\delta, \epsilon, \lambda}}{\delta m}(t_0, m_0, x)\right) m_0(dx) \\ & \leq \mathcal{F}(m_0) + C(\delta + \lambda + \epsilon\delta^{(2s+d/2+\eta-1)}). \end{aligned}$$

Proof. Claims (1)–(4) are easily proved using the regularity of U from Proposition (3.4) as well as the properties of the various regularization in Lemma (4.2) and Proposition (4.3). For (5), we fix $\eta > 0$ and we choose $C, c > 0$ so that Lemma 5.8 holds with the corresponding constants.

We assume that $(t_0, m_0) \in (0, T) \times \mathcal{P}(\mathbb{T}^d)$ is such that $t \mapsto U^{\delta, \epsilon, \lambda}$ is differentiable at t_0 . Then, by Lemma (5.6), $U^{\delta, \epsilon, \lambda}$ has a full derivative at (t_0, m_0) . Set $m_\lambda = \lambda\text{Leb} + (1 - \lambda)m_0$. Then $U^{\delta, \epsilon}$

has a full derivative at (t_0, m_λ) . If $\lambda \geq C\epsilon\delta^{-(2s+d/2+\eta-1)}$, then we can apply Lemma (5.8) and deduce that (5.32) holds at (t_0, m_λ) (even though the absolutely continuous part of m_0 is not lower bounded). Since

$$\partial_t U^{\delta,\epsilon,\lambda}(t_0, m_0) = \partial_t U^{\delta,\epsilon}(t_0, m_\lambda), \quad \frac{\delta U^{\delta,\epsilon,\lambda}}{\delta m}(t_0, m_0, \cdot) = (1-\lambda) \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, \cdot),$$

we have

$$\begin{aligned} & -\partial_t U^{\delta,\epsilon,\lambda}(t_0, m_0) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta U^{\delta,\epsilon,\lambda}}{\delta m}(t_0, m_0, x) m_0(dx) + \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta U^{\delta,\epsilon,\lambda}}{\delta m}(t_0, m_0, x)\right) m_0(dx) \\ &= -\partial_t U^{\delta,\epsilon}(t_0, m_\lambda) - (1-\lambda) \int_{\mathbb{T}^d} \Delta_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x) m_0(dx) \\ & \quad + \int_{\mathbb{T}^d} H\left(x, (1-\lambda) D_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x)\right) m_0(dx) \\ &= -\partial_t U^{\delta,\epsilon}(t_0, m_\lambda) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x) m_\lambda(dx) - \lambda \int_{\mathbb{T}^d} \Delta_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x) \text{Leb}(dx) \\ & \quad + \frac{1}{1-\lambda} \int_{\mathbb{T}^d} H\left(x, (1-\lambda) D_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x)\right) m_\lambda(dx) \\ & \quad + \frac{\lambda}{1-\lambda} \int_{\mathbb{T}^d} H\left(x, (1-\lambda) D_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x)\right) \text{Leb}(dx). \end{aligned} \tag{5.34}$$

To go further, we first use integration by parts to conclude that

$$\int_{\mathbb{T}^d} \Delta_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x) \text{Leb}(dx) = 0. \tag{5.35}$$

Next, we use convexity of the Hamiltonian in the second argument to get

$$\begin{aligned} H\left(x, (1-\lambda) D_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x)\right) &\leq (1-\lambda) H\left(x, D_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x)\right) + \lambda H(x, 0) \\ &\leq (1-\lambda) H\left(x, D_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x)\right) + C\lambda, \end{aligned}$$

and hence

$$\begin{aligned} & \frac{1}{1-\lambda} \int_{\mathbb{T}^d} H\left(x, (1-\lambda) D_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x)\right) m_\lambda(dx) \\ & \leq \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x)\right) m_\lambda(dx) + C\lambda. \end{aligned} \tag{5.36}$$

Similarly, using the bound on $\|D_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t, m_\lambda, \cdot)\|_{L^\infty}$ from Lemma 5.8, we have

$$\frac{\lambda}{1-\lambda} \int_{\mathbb{T}^d} H\left(x, (1-\lambda) D_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x)\right) \text{Leb}(dx) \leq C\lambda. \tag{5.37}$$

Combining (5.35), (5.36), and (5.37) with (5.34) and using Lemma 5.8, we arrive at

$$\begin{aligned} & -\partial_t U^{\delta,\epsilon,\lambda}(t_0, m_0) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta U^{\delta,\epsilon,\lambda}}{\delta m}(t_0, m_0, x) m_0(dx) + \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta U^{\delta,\epsilon,\lambda}}{\delta m}(t_0, m_0, x)\right) m_0(dx) \\ & \leq -\partial_t U^{\delta,\epsilon}(t_0, m_\lambda) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x) m_\lambda(dx) + \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta U^{\delta,\epsilon}}{\delta m}(t_0, m_\lambda, x)\right) m_\lambda(dx) + C\lambda \\ & \leq \mathcal{F}(m_\lambda) + C\left(\lambda + \delta + \epsilon\delta^{(2s+d/2+\eta-1)}\right) \leq \mathcal{F}(m_0) + C\left(\lambda + \delta + \epsilon\delta^{(2s+d/2+\eta-1)}\right), \end{aligned} \tag{5.38}$$

where the last line uses the fact that \mathcal{F} is d_1 -Lipschitz. This completes the proof.

□

5.3. Projections and rates of convergence. The goal of this section is to use the properties of U^ϵ and $U^{\delta, \epsilon, \lambda}$ proved in the previous section to establish estimates on the distance between their projections and the finite-dimensional value functions Φ^N . The main technical tool we use to do this is the following proposition.

Proposition 5.10. *Suppose that $\Phi = \Phi(t, q) : [0, T] \times H^{-s} \rightarrow \mathbb{R}$ is a continuous map such that*

- Φ is uniformly continuous on $[0, T] \times \mathcal{P}(\mathbb{T}^d)$ ($\mathcal{P}(\mathbb{T}^d)$ being endowed with the metric induced by $\|\cdot\|_{-s}$) and Lipschitz (for the same metric) on each set of the form $[0, t] \times \mathcal{P}(\mathbb{T}^d)$ with $t < T$.
- $\sup_{0 \leq t \leq T} [\Phi(t, \cdot)]_{C^{1,1}(H^{-s})} < \infty$.
- The derivative $\nabla_{-s}\Phi$ is jointly continuous on $[0, T] \times \mathcal{P}(\mathbb{T}^d)$.
- The inequality

$$-\partial_t \Phi(t, m) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta \Phi}{\delta m}(t, m, x) m(dx) + \int_{\mathbb{T}^d} H\left(x, D_x \frac{\delta \Phi}{\delta m}(t, m, x)\right) m(dx) \leq \mathcal{F}(m) + c_1, \quad (5.39)$$

is satisfied at any $(t, m) \in [0, T] \times \mathcal{P}(\mathbb{T}^d)$ where Φ has a full derivative, and $\Phi(T, m) \leq \mathcal{G}(m) + c_2$, for two positive constants c_1 and c_2 .

Define for each $N \in \mathbb{N}$ a function $\Phi^N(t, \mathbf{x}) : [0, T] \times (\mathbb{T}^d)^N \rightarrow \mathbb{R}$ by

$$\Phi^N(t, \mathbf{x}) = \Phi(t, m_{\mathbf{x}}^N).$$

Then we have

$$\Phi^N(t, \mathbf{x}) \leq V^N(t, \mathbf{x}) + T \left(c_1 + \frac{C}{N} \sup_r [\Phi(r, \cdot)]_{C^{1,1}(H^{-s})} \right) + c_2,$$

for each $N \in \mathbb{N}$ and each $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$, and a constant C which depends only on d and s .

Before moving to the proof of Proposition 5.10, we explain how, together with the results of the previous section, it can be used to easily establish the “hard inequalities”, from Theorems 2.7 and 2.6.

Proposition 5.11 (Upper bound in Theorem 2.7). *Let Assumption 2.4 hold. Then there is a constant C such that for each $N \in \mathbb{N}$,*

$$U(t, m_{\mathbf{x}}^N) \leq V^N(t, m_{\mathbf{x}}^N) + \frac{C}{\sqrt{N}},$$

for all $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$.

Proof. For each ϵ and N , we combine Proposition 5.10, Proposition 5.1, and Proposition 4.3 to conclude that

$$\begin{aligned} U(t, m_{\mathbf{x}}^N) - V^N(t, \mathbf{x}) &= U(t, m_{\mathbf{x}}^N) - U^\epsilon(t, m_{\mathbf{x}}^N) + U^\epsilon(t, m_{\mathbf{x}}^N) - V^N(t, \mathbf{x}), \\ &\leq C \left(\epsilon + \frac{1}{N\epsilon} \right). \end{aligned}$$

We choose $\epsilon = \frac{1}{\sqrt{N}}$ to get the result. □

Proposition 5.12 (Upper bound in Theorem 2.6). *Let Assumption 2.4 hold. Then for each $\eta > 0$, there is a constant C such that for each $N \in \mathbb{N}$ and $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$,*

$$U(t, m_{\mathbf{x}}^N) \leq V^N(t, m_{\mathbf{x}}^N) + CN^{-\beta(d)+\eta}, \quad \beta(d) := \frac{2}{3d+6}.$$

Proof of Proposition 5.12. We use Proposition 5.10 and Lemma 5.9 to conclude that for any $\eta > 0$, there are constants c and C such that for any $\delta > 0$, $0 < \epsilon < c\delta^{2(s-1)}$, and any $\lambda \geq C\epsilon\delta^{-(2s+d/2+\eta-1)}$, we have

$$\begin{aligned} U(t, m_{\mathbf{x}}^N) - V^N(t, \mathbf{x}) &= U(t, m_{\mathbf{x}}^N) - U^{\delta, \epsilon, \lambda}(t, m_{\mathbf{x}}^N) + U^{\delta, \epsilon, \lambda}(t, m_{\mathbf{x}}^N) - V^N(t, \mathbf{x}) \\ &\leq C \left(\lambda + \delta + \epsilon(\delta^{-2(s-1)} + \delta^{-(2s+d/2+\eta-1)}) + \frac{1}{N\epsilon} \right) \\ &\leq C \left(\lambda + \delta + \epsilon\delta^{-(2s+d/2+\eta-1)} + \frac{1}{N\epsilon} \right). \end{aligned}$$

We choose $\lambda = C\epsilon\delta^{-(2s+d/2+\eta-1)}$ (with the C here greater than C in the statement of Lemma 5.9) to find

$$U(t, m_{\mathbf{x}}^N) - V^N(t, \mathbf{x}) \leq C \left(\delta + \epsilon\delta^{-(2s+d/2+\eta-1)} + \frac{1}{N\epsilon} \right),$$

and then choose

$$\delta = N^{-1/(2s+d/2+\eta+1)}, \quad \epsilon = N^{-1}\delta^{-1}.$$

This yields

$$U(t, m_{\mathbf{x}}^N) - V^N(t, \mathbf{x}) \leq CN^{-1/(2s+d/2+\eta+1)}.$$

Since s was chosen to be an arbitrary real number with $s > d/2 + 1$, we find that for any $\alpha > 0$, there is a constant C such that

$$U(t, m_{\mathbf{x}}^N) \leq V^N(t, \mathbf{x}) + CN^{-\frac{1}{(\frac{3d}{2}+3+\alpha)}},$$

for all $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$, as claimed. \square

The rest of this subsection is devoted to a proof of Proposition 5.10. We start with two lemmas.

Lemma 5.13. *Let Φ satisfy the hypotheses of Proposition 5.10 and Φ^N be as in the statement of the same proposition. Then we have*

- (1) Φ^N is uniformly continuous on $[0, T] \times (\mathbb{T}^d)^N$ and $\Phi^N \in \text{Lip}([0, t] \times (\mathbb{T}^d)^N)$ for each $t < T$.
- (2) Φ^N is continuously differentiable in \mathbf{x} , with

$$D_{x^i} \Phi^N(t, \mathbf{x}) = \frac{1}{N} D_m \Phi(t, m_{\mathbf{x}}^N, x^i) = \frac{1}{N} D_y [D_{-s} \Phi(t, m_{\mathbf{x}}^N)](x^i), \quad i \in \{1, \dots, N\},$$

where we recall that x^i denotes the i^{th} d -dimensional entry of the N -tuple \mathbf{x} .

- (3) The derivative $D_{x^i} \Phi^N$ is Lipschitz continuous with respect to \mathbf{x} , uniformly in t .
- (4) The derivatives $\partial_t \Phi^N$, $D_{x^i} \Phi^N$, $D_{x^i x^j}^2 \Phi^N$ exist almost everywhere, and define versions of the corresponding weak derivatives of Φ^N . Moreover the spatial derivatives $D_{x^i} \Phi^N$ and $D_{x^i x^j}^2 \Phi^N$ are uniformly bounded on $[0, T] \times \mathbb{T}^d$ and the time derivative $\partial_t \Phi^N$ is uniformly bounded on $[0, t] \times \mathbb{T}^d$ for each $t < T$, so that in particular

$$\Phi^N \in W_{L^\infty}^{1,2}([0, t] \times \mathbb{T}^d),$$

for each $t < T$, with Φ^N , with $W_{L^\infty}^{1,2}([0, t] \times (\mathbb{T}^d)^N)$.

We note that in Claim (4) we use the notation $W_{L^\infty}^{1,2}([0, t] \times (\mathbb{T}^d)^N)$ for the set of functions $u : [0, t] \times (\mathbb{T}^d)^N \rightarrow \mathbb{R}$ with uniformly bounded weak derivatives $\partial_t u$, $D_{x^i} u$, $i = 1, \dots, N$, and $D_{x^i x^j}^2 u$, $i, j = 1, \dots, N$

Proof. Claim (1) follows from the fact that $\|m_{\mathbf{x}}^N - m_{\mathbf{y}}^N\|_{-s} \leq Cd_1(m_{\mathbf{x}}^N, m_{\mathbf{y}}^N) \leq \frac{C}{N} \sum_{i=1}^N |x^i - y^i|$.

Claim (2) follows easily from Lemma 2.11. For (3), we use (2), Proposition 2.12, the bound on $[\Phi(t, \cdot)]_{C^{1,1}(H^{-s})}$, and also the fact that $D_y D_m \Phi$ is bounded to conclude that

$$\begin{aligned} |D_{x^i} \Phi^N(t, \mathbf{x}) - D_{x^i} \Phi^N(t, \mathbf{y})| &= \frac{1}{N} |D_m \Phi(t, m_{\mathbf{x}}^N, x^i) - D_m \Phi(t, m_{\mathbf{y}}^N, y^i)| \\ &\leq C |D_m \Phi(t, m_{\mathbf{x}}^N, x^i) - D_m \Phi(t, m_{\mathbf{y}}^N, x^i)| + C |D_m \Phi(t, m_{\mathbf{y}}^N, x^i) - D_m \Phi(t, m_{\mathbf{y}}^N, y^i)| \\ &\leq C |\mathbf{x} - \mathbf{y}|. \end{aligned}$$

Finally claim (4) follows (1) and (3) together with Rademacher's Theorem. \square

Lemma 5.14. *Let Φ and Φ^N be as in the statement of Proposition 5.10. Then for almost every $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$, the function*

$$\mathbb{T}^d \ni x \mapsto D_m \Phi \left(t, \frac{1}{N} \sum_{j \neq i} \delta_{x^j} + \frac{1}{N} \delta_x, x^i \right) \quad (5.40)$$

is differentiable at x^i . At any point (t, \mathbf{x}) such that the map in (5.40) is differentiable at x^i , the second derivative $D_{x^i x^i}^2 \Phi^N$ exists (in the classical sense) and satisfies

$$D_{x^i x^i}^2 \Phi^N(t, \mathbf{x}) = \frac{1}{N} D_y D_m \Phi(t, m_{\mathbf{x}}^N, x^i) + \frac{1}{N^2} R^{N,i}(t, \mathbf{x}),$$

where $R^{N,i}$ is defined explicitly as

$$R^{N,i}(t, \mathbf{x}) = N D_y \left[\mathbb{T}^d \ni y \mapsto D_m \Phi \left(t, \frac{1}{N} \sum_{j \neq i} \delta_{x^j} + \frac{1}{N} \delta_y, x^i \right) \right] (x^i) \quad (5.41)$$

and satisfies $\|R^{N,i}\|_{L^\infty} \leq C \sup_{0 \leq t \leq T} [\Phi(t, \cdot)]_{C^{1,1}(H^{-s})}$, for C depending only on s and d .

Remark 5.15. Notice that if Φ in the above statement were twice differentiable (in the variable m), then $R^{N,i}(t, \mathbf{x})$ would be equal to

$$R^{N,i}(t, \mathbf{x}) = D_{mm}^2 \Phi(t, m_{\mathbf{x}}^N, x^i, x^i).$$

The very much spirit of Lemma 5.14 is that we are still able to provide an almost everywhere formula for $R^{N,i}(t, \mathbf{x})$ even though Φ just satisfies the assumptions of Proposition 5.10 (under which Φ may not be twice differentiable).

Proof. The first claim follows easily from the Lipschitz continuity of $m \mapsto D_m \Phi(t, m, y)$ (uniformly in t and y), see Remark 5.3. Recall from Lemma 5.13 that

$$D_{x^i} \Phi^N(t, \mathbf{x}) = \frac{1}{N} D_m \Phi(t, m_{\mathbf{x}}^N, x^i).$$

However, $y \mapsto D_m \Phi(t, m_{\mathbf{x}}^N, y) = D_y [D_{-s} \Phi(t, m_{\mathbf{x}}^N)](y)$ is continuously differentiable since $D_{-s} \Phi(t, m_{\mathbf{x}}^N)(\cdot) \in H^s \subset C^2(\mathbb{T}^d)$. And so it is clear that if $y \mapsto D_m \Phi(t, \frac{1}{N} \sum_{j \neq i} \delta_{x^j} + \frac{1}{N} \delta_y, x^i)$ is differentiable at x^i , then $D_{x^i x^i}^2 \Phi(t, \mathbf{x})$ exists and is given by

$$D_{x^i x^i}^2 \Phi(t, \mathbf{x}) = \frac{1}{N} D_y D_m \Phi(t, m_{\mathbf{x}}^N, x^i) + \frac{1}{N} D_y \left[y \mapsto D_m \Phi \left(t, \frac{1}{N} \sum_{j \neq i} \delta_{x^j} + \frac{1}{N} \delta_y, x^i \right) \right] (x^i),$$

as required. The L^∞ bound on $R^{N,i}$ follows easily from the fact that by Proposition 2.12, for all $\mathbf{x}, \bar{\mathbf{x}} \in (\mathbb{T}^d)^N$,

$$|D_m \Phi(t_0, m_{\mathbf{x}}^N, y) - D_m \Phi(t_0, m_{\bar{\mathbf{x}}}^N, y)| \leq C [\Phi(t_0, \cdot)]_{C^{1,1}} d_1(m_{\mathbf{x}}^N, m_{\bar{\mathbf{x}}}^N) \leq C [\Phi(t_0, \cdot)]_{C^{1,1}} \frac{1}{N} \sum_{i=1}^N |x^i - \bar{x}^i|.$$

\square

We are now ready to combine the previous two lemmas to prove Proposition 5.10

Proof of Proposition 5.10. We start by proving that for each N , we have, for almost every $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$,

$$\begin{aligned} & -\partial_t \Phi^N(t, \mathbf{x}) - \sum_{i=1}^N \Delta_{x^i} \Phi^N(t, \mathbf{x}) + \frac{1}{N} \sum_{i=1}^N H\left(x^i, ND_{x^i} \Phi^N(t, \mathbf{x})\right) \\ & \leq \mathcal{F}(m_{\mathbf{x}}^N) + \frac{C}{N} \sup_{r \in [0, T]} [\Phi(r, \cdot)]_{\mathcal{C}^{1,1}(H^{-s})}. \end{aligned} \quad (5.42)$$

To do this, we start by observing that for almost every (t, \mathbf{x}) , we have

- (1) the time derivative $\partial_t \Phi^N$ exists at (t, \mathbf{x}) ,
- (2) the map

$$\mathbb{T}^d \ni y \mapsto D_m \Phi\left(t, \frac{1}{N} \sum_{j \neq i} \delta_{x^j} + \frac{1}{N} \delta_y, x^i\right) \quad (5.43)$$

is differentiable at x^i .

But at any point (t, \mathbf{x}) satisfying both of these conditions, Lemma 5.6 guarantees that Φ is differentiable at (t, \mathbf{x}) , and so we can use Lemma 5.14 to get

$$\begin{aligned} & -\partial_t \Phi^N(t, \mathbf{x}) - \sum_{i=1}^N \Delta_{x^i} \Phi^N(t, \mathbf{x}) + \frac{1}{N} \sum_{i=1}^N H\left(x^i, ND_{x^i} \Phi^N(t, \mathbf{x})\right) \\ & = -\partial_t \Phi(t, m_{\mathbf{x}}^N) - \sum_{i=1}^N \left(\frac{1}{N} \Delta_x \frac{\delta \Phi}{\delta m}(t, m_{\mathbf{x}}^N, x^i) \right. \\ & \quad \left. + \frac{1}{N^2} R^{N,i}(t, \mathbf{x}) \right) + \frac{1}{N} \sum_{i=1}^N H\left(x^i, D_m \Phi(t, m_{\mathbf{x}}^N, x^i)\right) \\ & = -\partial_t \Phi(t, m_{\mathbf{x}}^N) - \int_{\mathbb{T}^d} \Delta_x \frac{\delta \Phi}{\delta m}(t, m_{\mathbf{x}}^N, y) m_{\mathbf{x}}^N(dy) \\ & \quad + \int_{\mathbb{T}^d} H(y, D_m \Phi(t, m_{\mathbf{x}}^N, y)) m_{\mathbf{x}}^N(dy) - \frac{1}{N^2} \sum_{i=1}^N R^{N,i}(t, \mathbf{x}) \\ & \leq \mathcal{F}(m_{\mathbf{x}}^N) + \frac{C}{N} \sup_{r \in [0, T]} [\Phi(r, \cdot)]_{\mathcal{C}^{1,1}(H^{-s})}, \end{aligned}$$

where we used the bound on $\|R^{N,i}\|_{L^\infty}$ from Lemma 5.14. This completes the proof of (5.42).

At this point, we know that for each $t < T$, Φ^N lies in the space $W_\infty^{1,2}([0, t] \times (\mathbb{T}^d)^N)$ of functions with L^∞ weak derivatives $\partial_t \Phi^N$, $D\Phi^N$, $D^2\Phi^N$, and also that Φ^N satisfies the inequality (5.42) at almost every (t, \mathbf{x}) . Moreover, we have by assumption

$$\Phi^N(T, \mathbf{x}) \leq V^N(T, \mathbf{x}) + c_2 = \mathcal{G}(m_{\mathbf{x}}^N) + c_2.$$

The rest of the proof is a standard ‘‘verification’’-type argument. Indeed, we fix (t_0, \mathbf{x}_0) and an admissible control $\alpha = (\alpha^1, \dots, \alpha^N)$ in feedback form, and we define $\mathbf{X} = (X^1, \dots, X^N)$ by

$$dX_t^i = \alpha^i(t, X_t^i) dt + dW_t^i, \quad t \in [t_0, T]; \quad X_{t_0}^i = x_0^i.$$

We use the Itô-Krylov formula, see [Kry80] Section 2.10, to see that

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}[\Phi^N(t, \mathbf{X}_t)] &= \mathbb{E}\left[\partial_t\Phi^N(t, \mathbf{X}_t) + \sum_{i=1}^N \Delta_{x^i}\Phi^N(t, \mathbf{X}_t) + \sum_{i=1}^N D_{x^i}\Phi^N(t, \mathbf{X}_t) \cdot \alpha^i(t, \mathbf{X}_t)\right] \\
&\geq \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^N \left(H(X_t^i, ND_{x^i}\Phi^N(t, \mathbf{X}_t)) + ND_{x^i}\Phi^N(t, \mathbf{X}_t) \cdot \alpha^i(t, \mathbf{X}_t)\right) \right. \\
&\quad \left. - \mathcal{F}(m_{\mathbf{X}_t}^N) - c_1 - C \sup_{r \in [0, T]} [\Phi(r, \cdot)]_{C^{1,1}(H^{-s})}\right] \\
&\geq -\mathbb{E}\left[\frac{1}{N} \sum_{i=1}^N L(X_t^i, \alpha_t^i) + \mathcal{F}(m_{\mathbf{X}_t}^N) - c_1 - C \sup_{r \in [0, T]} [\Phi(r, \cdot)]_{C^{1,1}(H^{-s})}\right].
\end{aligned}$$

Thus, integrating between t_0 and T the above inequality we get

$$\begin{aligned}
\Phi^N(t_0, \mathbf{x}_0) &\leq \mathbb{E}\left[\Phi^N(T, \mathbf{X}_T) + \int_{t_0}^T \left(\frac{1}{N} \sum_{i=1}^N L(X_t^i, \alpha_t^i) + \mathcal{F}(m_{\mathbf{X}_t}^N)\right) dt\right] + T\left(c_1 + C \sup_r [\Phi(r, \cdot)]_{C^{1,1}(H^{-s})}\right) \\
&\leq \mathbb{E}\left[\mathcal{G}(\mathbf{X}_T) + c_2 + \int_{t_0}^T \left(\frac{1}{N} \sum_{i=1}^N L(X_t^i, \alpha_t^i) + \mathcal{F}(m_{\mathbf{X}_t}^N)\right) dt\right] + T\left(c_1 + C \sup_r [\Phi(r, \cdot)]_{C^{1,1}(H^{-s})}\right).
\end{aligned}$$

Taking the infimum over α gives the result. \square

6. PROOFS OF THE “EASY INEQUALITIES”

6.1. The “easy inequality” in the d_1 -regular case. We start with a Lemma stating that V^N satisfies uniform in N Lipschitz estimates. The proof is essentially the same as Lemma 3.1 in [CDJS23], and so is omitted.

Lemma 6.1. *Let Assumption 2.1 hold, except for possibly d_1 -semiconcavity of \mathcal{F} and \mathcal{G} . Then there is a constant C such that for each $N \in \mathbb{N}$, $t \in [0, T]$ and $\mathbf{x}, \mathbf{y} \in (\mathbb{T}^d)^N$,*

$$|V^N(t, \mathbf{x}) - V^N(t, \mathbf{y})| \leq \frac{C}{N} \sum_{i=1}^N |x^i - y^i|.$$

Proposition 6.2. *Let Assumption 2.1 hold, except for possibly d_1 -semiconcavity of \mathcal{F} and \mathcal{G} . Then there is a constant C such that for each $N \in \mathbb{N}$,*

$$V^N(t, \mathbf{x}) \leq U(t, m_{\mathbf{x}}^N) + CR_{N,d},$$

for all $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$, where $R_{N,d}$ is as defined in (2.15).

Proof. We first define a lift $\widehat{V}^N = \widehat{V}^N(t, m) : [0, T] \times \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ by

$$\widehat{V}^N(t, m) = \int_{(\mathbb{T}^d)^N} V^N(t, \mathbf{x}) m^{\otimes N}(d\mathbf{x}).$$

Using the Markov structure of the control problem, it is standard to show that \widehat{V}^N satisfies

$$\widehat{V}^N(t, m) = \inf_{\alpha \in \mathcal{A}^N} \mathbb{E}\left[\int_t^T \left(\frac{1}{N} \sum_{i=1}^N L(X_s^i, \alpha_s^i) + \mathcal{F}(m_{\mathbf{X}_s}^N)\right) ds + \mathcal{G}(m_{\mathbf{X}_T}^N)\right] \quad (6.1)$$

with

$$dX_s^i = \alpha_s^i ds + dW_s^i, \quad s \in [t, T]; \quad X_t^i = \xi^i,$$

with $(\xi^i)_{i=1,\dots,N}$ i.i.d. with common law m . Now let (α, m) be a candidate for the optimization problem defining $U(t, m)$, and suppose that α is bounded so that $m_s = \mathcal{L}(Y_s)$, for $s \in [t, T]$, where Y is the unique strong solution of

$$dY_s = \alpha(s, Y_s)ds + dW_s, \quad s \in [t, T]; \quad X_t^i = \xi \sim m.$$

Next define $\mathbf{Y} = (Y^1, \dots, Y^N)$ by

$$dY_s^i = \alpha(s, Y_s^i)ds + dW_s^i, \quad s \in [t, T]; \quad X_t^i = \xi^i,$$

where $(\xi^i)_{i=1,\dots,N}$ are as above. Then using (6.1), we have

$$\begin{aligned} \widehat{V}^N(t, m) &\leq \mathbb{E} \left[\int_t^T \left(\frac{1}{N} \sum_{i=1}^N L(Y_s^i, \alpha(s, Y_s^i)) + \mathcal{F}(m_{\mathbf{Y}_s^N}^N) \right) ds + \mathcal{G}(m_{\mathbf{Y}_T^N}^N) \right] \\ &= \mathbb{E} \left[\int_t^T \left(L(Y_s, \alpha(s, Y_s)) + \mathcal{F}(m_{\mathbf{Y}_s^N}^N) \right) ds + \mathcal{G}(m_{\mathbf{Y}_T^N}^N) \right] \\ &\leq \mathbb{E} \left[\int_t^T \left(L(Y_s, \alpha(s, Y_s)) + \mathcal{F}(\mathcal{L}(Y_s)) \right) ds + \mathcal{G}(\mathcal{L}(Y_T)) \right] + CR_{N,d}, \end{aligned}$$

with the last inequality following from [FG15]. Taking an infimum in α , we conclude that

$$\widehat{V}^N(t, m) \leq U(t, m) + CR_{N,d}. \quad (6.2)$$

Next, we use the fact that V^N is symmetric in (x^1, \dots, x^N) (the entries of \mathbf{x}) and satisfies the estimate in Lemma 6.1 to get

$$\begin{aligned} \left| \widehat{V}^N(t, m_{\mathbf{x}^N}^N) - V^N(t, \mathbf{x}) \right| &= \left| \int_{(\mathbb{T}^d)^N} \left[V^N(t, \mathbf{y}) - V^N(t, \mathbf{x}) \right] m_{\mathbf{x}^N}^{\otimes N}(d\mathbf{y}) \right| \\ &\leq C \int_{(\mathbb{T}^d)^N} \left[\frac{1}{N} \inf_{s \in \mathcal{S}_N} \sum_{i=1}^N |y^{s(i)} - x^i| \right] m_{\mathbf{x}^N}^{\otimes N}(d\mathbf{y}) \\ &= C \int_{(\mathbb{T}^d)^N} d_1(m_{\mathbf{y}^N}^N, m_{\mathbf{x}^N}^N) m_{\mathbf{x}^N}^{\otimes N}(d\mathbf{y}) \leq CR_{N,d}, \end{aligned} \quad (6.3)$$

with the last inequality coming from [FG15]. Combining (6.3) with (6.2) completes the proof. \square

Proof of Theorem 2.6. Combine Propositions 6.2 and 5.12. \square

6.2. An auxiliary estimate for linear PDEs on the Wasserstein space. In order to prove the lower bound in Theorem 2.7, we need to first prove a similar bound for a class of linear PDEs on the Wasserstein space. We fix two functions

$$\alpha = \alpha(t, x) : [0, T] \times \mathbb{T}^d \rightarrow \mathbb{R}, \quad \phi = \phi(m) : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}.$$

The following Assumption will be in force throughout the Subsection.

Assumption 6.3. *We assume that s is a real number with $s > d/2 + 1$, and that $\alpha = \alpha(t, x) : [0, T] \times \mathbb{T}^d \rightarrow \mathbb{R}$ is jointly continuous and satisfies*

$$\sup_{0 \leq t \leq T} \|\alpha(t, \cdot)\|_{s-1} < \infty.$$

We will consider the maps

$$\Phi^N = \Phi^N(t, \mathbf{x}) : [0, T] \times (\mathbb{T}^d)^N \rightarrow \mathbb{R}, \quad \Phi^\infty = \Phi^\infty(t, \mu) : [0, T] \times \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}, \quad (6.4)$$

defined by

$$\Phi^N(t, \mathbf{x}) = \mathbb{E}[\phi(m_{\mathbf{X}_T^{t, \mathbf{x}}}^N)], \quad \text{and} \quad \Phi^\infty(t, \mu) = \phi(m_T^{t, \mu}), \quad (6.5)$$

where $(\mathbf{X}_s^{t,\mathbf{x}})_{t \leq s \leq T}$ denotes the N -tuple $(X_s^1, \dots, X_s^N)_{t \leq s \leq T}$ solving the system of SDEs

$$dX_s^i = \alpha(s, X_s^i)ds + \sqrt{2}dW_s^i, \quad s \in [t, T]; \quad X_t^i = x^i; \quad i = 1, \dots, N,$$

and $(m_s)_{t \leq s \leq T}$ denotes the solution to the Fokker Planck equation

$$\partial_s m_s = \Delta_x m_s - \operatorname{div}_x(m_s \alpha(s, \cdot)), \quad s \in [t, T]; \quad m_t = \mu.$$

Formally, we expect (by the method of characteristics) that Φ^N solves the PDE

$$\begin{cases} -\partial_t \Phi^N(t, \mathbf{x}) - \sum_{i=1}^N \Delta_{x^i} \Phi^N(t, \mathbf{x}) - \sum_{i=1}^N \alpha(s, x^i) \cdot D_{x^i} \Phi^N(t, \mathbf{x}) = 0, & (t, \mathbf{x}) \in [0, T) \times (\mathbb{T}^d)^N, \\ \Phi^N(T, \mathbf{x}) = \phi(m_{\mathbf{x}}^N), & \mathbf{x} \in (\mathbb{T}^d)^N, \end{cases} \quad (6.6)$$

and that Φ^∞ solves the PDE

$$\begin{cases} -\partial_t \Phi^\infty(t, m) - \int_{\mathbb{T}^d} \Delta_y \frac{\delta \Phi^\infty}{\delta m}(t, m, y) m(dy) - \int_{\mathbb{R}^d} \alpha(t, y) \cdot D_y \frac{\delta \Phi^\infty}{\delta m}(t, m, y) m(dy) = 0, \\ (t, m) \in [0, T) \times \mathcal{P}(\mathbb{T}^d), \\ \Phi^\infty(T, m) = \phi(m), \quad m \in \mathcal{P}(\mathbb{T}^d). \end{cases} \quad (6.7)$$

The existence and uniqueness of classical solutions to (6.7) under appropriate technical conditions is by now well-understood, and we refer to [BLPR17, CF22, Tse21] for some results in this direction. It is also well-known that when Φ^∞ is smooth enough, the functions Φ^N converge to Φ^∞ with the rate $\frac{1}{N}$, see e.g. [CST22, CF21, DT21]. In the proof of the lower bound in Theorem 2.7, we need a similar result, which is stated here.

Proposition 6.4. *Suppose that Assumption 6.3 holds, and that $\phi = \phi(m) : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ is Lipschitz and semi-concave with respect to $\|\cdot\|_{-s}$. Then we have*

$$\Phi^N(t, \mathbf{x}) \leq \Phi^\infty(t, m_{\mathbf{x}}^N) + \frac{C}{N},$$

for all $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$ and a constant C independent of N .

We emphasize that the estimate appearing in Proposition 6.4 is by now fairly standard. Compared to existing results, the main novelty is that we do not require ϕ to be twice differentiable, we only require a one-sided bound in the form of semi-concavity. This means that Φ^∞ may not in fact be a classical solution to the PDE (6.7), and some mollification arguments are required. Lemma 6.5 below is the key technical result which shows how semi-concavity with respect to $\|\cdot\|_{-s}$ can lead to a one-sided estimate between Φ^N and Φ^∞ .

Before proving Proposition 6.4 we need several preliminary results.

Lemma 6.5. *Suppose that Assumption 6.3 holds and that $\psi = \psi(q) : H^{-s} \rightarrow \mathbb{R}$ is \mathcal{C}^1 with $[\psi]_{\mathcal{C}^{1,1}(H^{-s})} < \infty$ and C_S semi-concave, i.e. the map*

$$q \mapsto \psi(q) - \frac{C_S}{2} \|q\|_{-s}^2$$

is concave. Then there is a constant C depending on ψ only through C_S such that the inequality

$$\sum_{j=1}^d \left(D_{y_j} \left[\mathbb{T}^d \ni y = (y_1, \dots, y_d) \mapsto (D_m \psi)^j \left(\frac{1}{N} \sum_{k \neq i} \delta_{x^k} + \frac{1}{N} \delta_y, x^i \right) \right] (x^i) \right) \leq \frac{C}{N} \quad (6.8)$$

holds at any point $\mathbf{x} = (x^1, \dots, x^N) \in (\mathbb{R}^d)^N$ and any index $i \in \{1, \dots, N\}$ such that the derivatives appearing in (6.8) exist.

Proof. The starting point is to notice that by semi-concavity, we have, for all $q_1, q_2 \in H^{-s}$,

$$\langle D_{-s}\psi(q^1) - D_{-s}\psi(q^2), q^1 - q^2 \rangle_{s,-s} \leq C_S \|q_1 - q_2\|_{-s}^2. \quad (6.9)$$

We fix $\mathbf{x} = (x^1, \dots, x^N) \in (\mathbb{T}^d)^N$, $\lambda > 0$, $i \in \{1, \dots, N\}$ and $j \in \{1, \dots, d\}$ and test the inequality (6.9) with

$$q^1 = m_{\mathbf{x}}^N, \quad q^2 = m_{\mathbf{x}}^N + \frac{\lambda}{N} D_{y_j}(\delta_{x^i}),$$

to find (using the fact that $s > d/2 + 1$ and that H^s embeds into $\mathcal{C}^2(\mathbb{T}^d)$) that

$$-\left\langle D_{-s}\psi(m_{\mathbf{x}}^N) - D_{-s}\psi\left(m_{\mathbf{x}}^N + \frac{\lambda}{N} D_{y_j}(\delta_{x^i})\right), \frac{\lambda}{N} D_{y_j}(\delta_{x^i}) \right\rangle_{s,-s} \leq C_S \frac{\lambda^2}{N^2} \|D_{y_j}(\delta_{x^i})\|_{-s}^2 \leq C \frac{\lambda^2}{N^2},$$

or equivalently

$$D_{y_j} [D_{-s}\psi(m_{\mathbf{x}}^N)](x^i) - D_{y_j} \left[D_{-s}\psi\left(m_{\mathbf{x}}^N + \frac{\lambda}{N} D_{y_j}(\delta_{x^i})\right) \right](x^i) \leq \frac{C\lambda}{N}. \quad (6.10)$$

Next, we define $\mathbf{x}^\lambda = (x^1, \dots, x^i - \lambda e^j, x^{i+1}, \dots, x^N)$, and we note that

$$\left\| m_{\mathbf{x}}^N + \frac{\lambda}{N} D_{y_j}(\delta_{x^i}) - m_{\mathbf{x}^\lambda}^N \right\|_{-s} = \frac{1}{N} \sup_{\|f\|_s=1} \left(f(x^i) - f(x^i - \lambda e^j) - \lambda D_{y_j} f(x^i) \right) \leq \frac{C\lambda^2}{N},$$

where we used again the Sobolev embedding of H^s into $\mathcal{C}^2(\mathbb{T}^d)$. Combining this with (6.10) and Lemma 2.11, we have

$$\begin{aligned} (D_m \psi)^j(m_{\mathbf{x}}^N, x^i) - (D_m \psi)^j(m_{\mathbf{x}^\lambda}^N, x^i) &= D_{y_j} \left[\frac{\delta \psi}{\delta m}(m_{\mathbf{x}}^N) \right](x^i) - D_{y_j} \left[\frac{\delta \psi}{\delta m}(m_{\mathbf{x}^\lambda}^N) \right](x^i) \\ &= D_{y_j} [D_{-s}\psi(m_{\mathbf{x}}^N)](x^i) - D_{y_j} [D_{-s}\psi(m_{\mathbf{x}^\lambda}^N)](x^i) \\ &= D_{y_j} [D_{-s}\psi(m_{\mathbf{x}}^N)](x^i) - D_{y_j} \left[D_{-s}\psi\left(m_{\mathbf{x}}^N + \frac{\lambda}{N} D_{y_j}(\delta_{x^i})\right) \right](x^i) \\ &\quad + D_{y_j} \left[D_{-s}\psi\left(m_{\mathbf{x}}^N + \frac{\lambda}{N} D_{y_j}(\delta_{x^i})\right) \right](x^i) - D_{y_j} [D_{-s}\psi(m_{\mathbf{x}^\lambda}^N)](x^i) \\ &\leq \frac{C\lambda}{N} + \frac{C[\psi]_{\mathcal{C}^{1,1}(H^{-s})} \lambda^2}{N}. \end{aligned}$$

Thus, if \mathbf{x} is such that the derivatives appearing in (6.8) exist, then we must have

$$\begin{aligned} D_{y_j} \left[y \mapsto (D_m \psi)^j \left(\frac{1}{N} \sum_{k \neq i} \delta_{x^k} + \frac{1}{N} \delta_y, x^i \right) \right](x^i) \\ = \lim_{\lambda \rightarrow 0^+} \frac{1}{\lambda} \left((D_m \psi)^j(m_{\mathbf{x}}^N, x^i) - (D_m \psi)^j(m_{\mathbf{x}^\lambda}^N, x^i) \right) \leq \frac{C}{N}, \end{aligned}$$

and summing over j gives the result. \square

Lemma 6.6. *If $\phi = \phi(q) : H^{-s} \rightarrow \mathbb{R}$ is $\mathcal{C}_{loc}^{1,1}(H^{-s})$ and C_S -semi-concave with respect to $\|\cdot\|_{-s}$, and Assumption 6.3 is in force, then Φ^∞ is a classical solution of (6.7). Moreover, Φ^∞ can be extended to a function*

$$\Phi^\infty = \Phi^\infty(t, q) : [0, T] \times H^{-s} \rightarrow \mathbb{R}.$$

such that Φ is CC_S -semi-concave, with C independent of ϕ .

We remark that in the statement of Lemma 6.6 we use the standard notation $\mathcal{C}_{loc}^{1,1}(H^{-s})$ for the set of \mathcal{C}^1 functions ϕ such that $\nabla_{-s}\phi$ is Lipschitz on each bounded subset of H^{-s} .

Proof. For any $q_0 \in \mathcal{C}^\infty(\mathbb{T}^d)$ we can find a unique classical solution to the equation

$$\partial_t q_t + \operatorname{div}_x(\alpha(t, \cdot) q_t) - \Delta_x q_t = 0 \text{ in } (t_0, T) \times \mathbb{T}^d, \quad q(t_0) = q_0,$$

and moreover, following the proof of Lemma 3.3 given in Appendix A.1, there is some $C := C(\sup_{t \in [0, T]} \|\alpha(t, \cdot)\|_{s-1}) > 0$ such that, for any two solutions $(q_t^{t_0, q_1})_{t \in [t_0, T]}$, $(q_t^{t_0, q_2})_{t \in [t_0, T]}$ starting respectively from q_1 and q_2 it holds

$$\sup_{t \in [t_0, T]} \|q_t^{t_0, q_1} - q_t^{t_0, q_2}\|_{-s} \leq C \|q_1 - q_2\|_{-s}. \quad (6.11)$$

As a consequence, we can extend by density the map $q_0 \mapsto (q_t^{t_0, q_0})_{t \in [t_0, T]} \in \mathcal{C}([t_0, T], H^{-s})$ to the whole of H^{-s} so that (6.11) is satisfied for any $q_0 \in H^{-s}$. We define Φ^∞ over the whole of $[0, T] \times H^{-s}(\mathbb{T}^d)$ by

$$\Phi^\infty(t_0, q_0) = \phi(q_T^{t_0, q_0}).$$

The map $\Phi^\infty(t_0, \cdot)$ is the composition map $\phi \in \mathcal{C}_{loc}^{1,1}(H^{-s})$ and the bounded linear map $H^{-s} \ni q_0 \mapsto q_T^{t_0, q_0} \in H^{-s}$. Therefore $\Phi^\infty(t_0, \cdot) \in \mathcal{C}_{loc}^{1,1}(H^{-s})$. In particular, the derivatives in the measure argument appearing in equation (6.7) exist. Moreover, if ϕ is C_S semi-concave then $\Phi^\infty(t_0, \cdot)$ is $C_S C$ semi-concave for some $C := C(\sup_{t \in [0, T]} \|\alpha(t, \cdot)\|_{s-1}) > 0$. It is then plain to check, by expanding Φ^∞ along solutions to the Fokker-Planck equation that $\Phi^\infty(\cdot, q_0)$ is differentiable whenever $q_0 \in \mathcal{P}(\mathbb{T}^d)$ and that Φ^∞ solves (6.7) over $[0, T] \times \mathcal{P}(\mathbb{T}^d)$. \square

Proof of Proposition 6.4. We start by defining, for each $\epsilon > 0$, the sup-convolution $\phi^\epsilon(q) : H^{-s} \rightarrow \mathbb{R}$ as in (4.1), i.e.

$$\phi^\epsilon(q) = \sup_{m \in \mathcal{P}(\mathbb{T}^d)} \left\{ \phi(q) - \frac{1}{2\epsilon} \|q - m\|_{-s}^2 \right\}.$$

By Proposition 4.3, we know that $\phi^\epsilon \in \mathcal{C}_{loc}^{1,1}(H^{-s})$. By the proof of the same proposition, it is clear that for all $\epsilon < \frac{1}{2C_S}$, ϕ^ϵ is $2C_S$ -semi-concave, where C_S denotes the semi-concavity constant of ϕ . Thus we can apply Lemma 6.6 to see that the map $\Phi^{\infty, \epsilon}$ given by (6.4)–(6.5) with ϕ^ϵ instead of ϕ is a classical solution of (6.7) (with ϕ^ϵ as boundary condition), and extends to a map $\Phi^{\infty, \epsilon} : [0, T] \times H^{-s} \rightarrow \mathbb{R}$ such that for each $t \in [0, T]$, $\Phi^{\infty, \epsilon}(t, \cdot)$ is semi-concave with constant C'_S depending only on ϕ only through C_S . Moreover, since $\Phi^{\infty, \epsilon} \in \mathcal{C}^1([0, T] \times \mathcal{P}(\mathbb{T}^d))$, the projections $\Phi^{\infty, \epsilon, N} : [0, T] \times (\mathbb{T}^d)^N \rightarrow \mathbb{R}$ given by

$$\Phi^{\infty, \epsilon, N}(t, \mathbf{x}) = \Phi^{\infty, \epsilon}(t, m_{\mathbf{x}}^N), \quad (t, \mathbf{x}) \in [0, T] \times \mathcal{P}(\mathbb{T}^d),$$

have derivatives

$$\partial_t \Phi^{\infty, \epsilon, N}(t, \mathbf{x}) = \partial_t \Phi^{\infty, \epsilon}(t, m_{\mathbf{x}}^N), \quad D_{x^i} \Phi^{\infty, \epsilon, N}(t, \mathbf{x}) = \frac{1}{N} D_m \Phi^{\infty, \epsilon}(t, m_{\mathbf{x}}^N, x^i),$$

which are continuous and bounded on $[0, T] \times (\mathbb{T}^d)^N$. In addition, we can apply (the reasoning of) Lemma 5.13 to conclude that $\Phi^{\infty, \epsilon, N}$ belongs to $W_{L^\infty}^{1,2}([0, T] \times \mathbb{T}^d)$, with derivatives $D_{x^i x^j} \Phi^{\infty, \epsilon, N}$ existing almost everywhere. Next, following the proof of Lemma 5.14, we note that for almost every (t, \mathbf{x}) , the derivatives

$$D_{y^j} \left[y \mapsto (D_m \Phi^{\infty, \epsilon})^j \left(\frac{1}{N} \sum_{k \neq i} \delta_{x^k} + \frac{1}{N} \delta_y, x^i \right) \right] (x^i) \quad (6.12)$$

exist for each $j = 1, \dots, d$, and that at any such point we have, for any $i = 1, \dots, N$,

$$\begin{aligned} \Delta_{x^i} \Phi^{\infty, \epsilon, N}(t, \mathbf{x}) &= \frac{1}{N} \operatorname{tr} (D_y D_m \Phi^{\infty, \epsilon}(t, m_{\mathbf{x}}^N, x^i)) \\ &\quad + \frac{1}{N} \sum_{j=1}^d \left(D_{y_j} \left[y \mapsto (D_m \Phi^{\infty, \epsilon})^j \left(\frac{1}{N} \sum_{k \neq i} \delta_{x^k} + \frac{1}{N} \delta_y, x^i \right) \right] (x^i) \right), \end{aligned}$$

and so at such a point

$$\begin{aligned} & -\partial_t \Phi^{\infty, \epsilon, N}(t, \mathbf{x}) - \sum_{i=1}^N \Delta_{x^i} \Phi^{\infty, \epsilon, N}(t, \mathbf{x}) - \sum_{i=1}^N \alpha(t, x^i) \cdot D_{x^i} \Phi^{\infty, \epsilon, N}(t, \mathbf{x}) \\ &= -\partial_t \Phi^{\infty, \epsilon}(t, m_{\mathbf{x}}^N) - \frac{1}{N} \sum_{i=1}^N \operatorname{tr} (D_y D_m \Phi^{\infty, \epsilon}(t, m_{\mathbf{x}}^N, x^i)) - \frac{1}{N} \sum_{i=1}^N \alpha(t, x^i) \cdot D_m \Phi^{\infty, \epsilon}(t, m_{\mathbf{x}}^N) \\ &\quad - \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^d \left(D_{y_j} \left[y \mapsto (D_m \Phi^{\infty, \epsilon})^j \left(\frac{1}{N} \sum_{k \neq i} \delta_{x^k} + \frac{1}{N} \delta_y, x^i \right) \right] (x^i) \right) \\ &= -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^d \left(D_j \left[y \mapsto (D_m \Phi^{\infty, \epsilon})^j \left(\frac{1}{N} \sum_{k \neq i} \delta_{x^k} + \frac{1}{N} \delta_y, x^i \right) \right] (x^i) \right) \geq -\frac{C}{N}, \end{aligned}$$

where C depends on ϕ only through C_S and we have used both the PDE for $\Phi^{\infty, \epsilon}$ and Lemma 6.5. To summarize, at this stage we know that $\Phi^{\infty, \epsilon, N} \in W_{\infty}^{1,2}([0, T] \times \mathbb{T}^d)$ satisfies

$$-\partial_t \Phi^{\infty, \epsilon, N}(t, \mathbf{x}) - \sum_{i=1}^N \Delta_{x^i} \Phi^{\infty, \epsilon, N}(t, \mathbf{x}) - \sum_{i=1}^N \alpha(t, x^i) \cdot D_{x^i} \Phi^{\infty, \epsilon, N}(t, \mathbf{x}) \geq -\frac{C}{N},$$

for almost every $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$, while it is clear that the function $\Phi^{N, \epsilon}$ defined by (6.4)–(6.5) with ϕ^ϵ instead of ϕ satisfies the PDE (6.6) (with $\mathbf{x} \mapsto \phi^\epsilon(m_{\mathbf{x}}^N)$ as terminal condition) in a classical sense on $[0, T] \times (\mathbb{T}^d)^N$. This is enough to conclude by the comparison principle (which can be justified in this setting by the Itô-Krylov formula, as in the proof of Proposition 5.10) that

$$\Phi^{N, \epsilon}(t, \mathbf{x}) \leq \Phi^{\infty, \epsilon}(t, m_{\mathbf{x}}^N) + \frac{C}{N},$$

for each $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$ and a constant C depending on ϕ only through C_S (and in particular C is independent of ϵ). Finally, it is clear that $\Phi^{N, \epsilon} \rightarrow \Phi^N$, $\Phi^{\infty, \epsilon} \rightarrow \Phi^{\infty}$ uniformly as $\epsilon \downarrow 0$, and this completes the proof. \square

6.3. The “easy inequality” in the H^{-s} -regular case.

Proposition 6.7. *Let Assumption 2.4 hold. Then there is a constant C such that for each $N \in \mathbb{N}$,*

$$V^N(t, \mathbf{x}) \leq U(t, m_{\mathbf{x}}^N) + \frac{C}{N},$$

for all $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$.

Proof. Fix $(t_0, \mathbf{x}_0) \in [0, T] \times (\mathbb{T}^d)^N$, and let α be an optimizer for the mean field control problem starting from $(t_0, m_{\mathbf{x}_0}^N)$. We define $\mathbf{X} = (X^i)_{i=1, \dots, N}$ by

$$dX_t^i = \alpha(t, X_t^i) dt + \sqrt{2} dW_t^i, \quad t \in [t_0, T]; \quad X_{t_0}^i = x_{t_0}^i,$$

and define m by

$$\partial_t m_t = \Delta_x m_t - \operatorname{div}_x (m_t \alpha(t, \cdot)), \quad t \in [t_0, T]; \quad m_{t_0} = m_{\mathbf{x}_0}^N.$$

Notice that by linearity of the Fokker-Planck equation, $m_t = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(X_t^i)$, so that

$$\mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N L(X_t^i, \alpha(t, X_t^i)) \right] = \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{T}^d} L(x, \alpha(t, x)) \mathcal{L}(X_t^i)(dx) = \int_{\mathbb{T}^d} L(x, \alpha(t, x)) m_t(dx).$$

Using this, we can write

$$\begin{aligned} V^N(t_0, \mathbf{x}_0) &\leq \mathbb{E} \left[\int_{t_0}^T \left(\frac{1}{N} \sum_{i=1}^N L(X_t^i, \alpha(t, X_t^i)) + \mathcal{F}(m_{\mathbf{X}_t}^N) \right) dt + \mathcal{G}(m_{\mathbf{X}_T}^N) \right] \\ &= \int_{t_0}^T \int_{\mathbb{R}^d} L(x, \alpha(t, x)) m_t(dx) dt + \mathbb{E} \left[\int_{t_0}^T \mathcal{F}(m_{\mathbf{X}_t}^N) dt + \mathcal{G}(m_{\mathbf{X}_T}^N) \right] \\ &= U(t_0, m_{\mathbf{x}_0}) + \int_{t_0}^T E_t^1 dt + E^2, \end{aligned}$$

where

$$E_t^1 := \mathbb{E}[\mathcal{F}(m_{\mathbf{X}_t}^N)] - \mathcal{F}(m_t), \quad E^2 := \mathbb{E}[\mathcal{G}(m_{\mathbf{X}_T}^N)] - \mathcal{G}(m_T).$$

But using the Lipschitz and semi-concavity of \mathcal{F} and \mathcal{G} in H^{-s} and applying Proposition 6.4 (with $\phi = \mathcal{F}(t, \cdot)$ and $\phi = \mathcal{G}$), we see that

$$E_t^1 \leq \frac{C}{N}, \quad E^2 \leq \frac{C}{N},$$

and so

$$V^N(t_0, \mathbf{x}_0) \leq U(t_0, m_{\mathbf{x}_0}^N) + \frac{C}{N},$$

which completes the proof. □

Proof of Theorem 2.7. Combine Propositions 6.7 and 5.11. □

7. PROOFS FOR THE EXAMPLES AND THE CONVEX CASE

Lemma 7.1. *Suppose that \mathcal{F} and \mathcal{G} are convex and d_1 -Lipschitz, and H satisfies the conditions of Assumption 2.1. Then*

$$U(t, m_{\mathbf{x}}^N) \leq V^N(t, \mathbf{x}), \tag{7.1}$$

for all $N \in \mathbb{N}$ and $(t, \mathbf{x}) \in [0, T] \times (\mathbb{T}^d)^N$.

Proof. Let us first start with the observation that by Jensen's inequality, we have

$$\mathcal{F} \left(\frac{1}{N} \sum_{i=1}^N \mathcal{L}(X^i) \right) \leq \mathbb{E}[\mathcal{F}(m_{\mathbf{X}}^N)], \quad \mathcal{G} \left(\frac{1}{N} \sum_{i=1}^N \mathcal{L}(X^i) \right) \leq \mathbb{E}[\mathcal{G}(m_{\mathbf{X}}^N)], \tag{7.2}$$

for any $N \in \mathbb{N}$ and any $\mathbf{X} = (X^1, \dots, X^N)$ a random vector taking values in $(\mathbb{T}^d)^N$.

Now fix $(t_0, \mathbf{x}_0) \in [0, T] \times (\mathbb{T}^d)^N$, and let $\boldsymbol{\alpha} = (\alpha^1, \dots, \alpha^N)$ denote an admissible control for the N -particle control problem started from (t_0, \mathbf{x}_0) . We aim to build a control for the mean field control problem out of the control $\boldsymbol{\alpha}$. To do this, we start by using a mimicking result from [BS13] to conclude that we can find for each $i \in \{1, \dots, N\}$ a bounded and measurable function

$$\hat{\alpha}^i = \hat{\alpha}^i(t, x) : [t_0, T] \times \mathbb{T}^d \rightarrow \mathbb{R}^d$$

such that

- for almost every $t \in [0, T]$, $\hat{\alpha}^i(t, \cdot)$ is a version of $\mathbb{E}[\alpha_t^i | X_t^i = \cdot]$

- for each $t \in [0, T]$, we have

$$\mathcal{L}(X_t^i) = \mathcal{L}(\widehat{X}_t^i),$$

where X^i denotes the solution to the SDE

$$d\widehat{X}_t^i = \widehat{\alpha}^i(t, \widehat{X}_t^i)dt + dW_t^i, \quad t \in [t_0, T]; \quad \widehat{X}_{t_0}^i = x_0^i.$$

Next, we fix an \mathcal{F}_{t_0} -measurable random variable ξ with $\mathcal{L}(\xi) = m_{x_0}^N$. We choose ξ in such a way that we can find a partition of Ω by \mathcal{F}_{t_0} -measurable sets A^1, \dots, A^N such that $\mathbb{P}[A^i] = \frac{1}{N}$ and $\xi = x_0^i$ on A^i for each $i \in \{1, \dots, N\}$. (For instance, if we require ξ to take values in $\{x_0^1, \dots, x_0^N\}$, then A^i is just taken as the pre-image of x_0^i by ξ .) Now, we build a control for the mean field problem (started from $(t_0, m_{x_0}^N)$) as follows. First, we define a function

$$\widehat{\alpha} = \widehat{\alpha}(\omega, t, x) : \Omega \times [t_0, T] \times \mathbb{T}^d \rightarrow \mathbb{R}^d$$

by the formula

$$\widehat{\alpha}(\omega, t, x) = \sum_{i=1}^N 1_{A^i}(\omega) \widehat{\alpha}^i(t, x).$$

We denote by \widehat{X} the solution of the SDE

$$d\widehat{X}_t = \widehat{\alpha}(\omega, t, \widehat{X}_t)dt + dW_t, \quad t \in [t_0, T]; \quad \widehat{X}_{t_0} = \xi, \quad (7.3)$$

and we define $\widehat{\alpha}_t := \widehat{\alpha}(\omega, t, X_t)$. It is easy to argue that in fact (7.3) has a unique strong solution thanks to boundedness of the functions $\widehat{\alpha}^i$, and it can be built for instance by setting $\widehat{X}_t = \sum_{i=1}^N \widehat{X}_t^i 1_{A^i}$, where \widehat{X}_t^i solves

$$d\widehat{X}_t^i = \widehat{\alpha}^i(t, \widehat{X}_t^i)dt + dW_t^i, \quad t \in [t_0, T]; \quad \widehat{X}_{t_0}^i = x_0^i. \quad (7.4)$$

We note that we clearly have

$$\mathcal{L}(\widehat{X}_t) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(\widehat{X}_t^i) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(X_t^i).$$

Moreover, we notice that for almost every t ,

$$\begin{aligned} \mathbb{E}[L(\widehat{X}_t, \widehat{\alpha}_t)] &= \frac{1}{N} \sum_{i=1}^N \mathbb{E}[L(\widehat{X}_t^i, \widehat{\alpha}^i(t, \widehat{X}_t^i))] = \frac{1}{N} \sum_{i=1}^N \mathbb{E}[L(X_t^i, \widehat{\alpha}^i(t, X_t^i))] \\ &= \frac{1}{N} \sum_{i=1}^N \mathbb{E}[L(X_t^i, \mathbb{E}[\alpha_t^i | X_t^i])] \leq \frac{1}{N} \sum_{i=1}^N \mathbb{E}[L(X_t^i, \alpha_t^i)]. \end{aligned} \quad (7.5)$$

We now conclude by using (7.2) and (7.5) to estimate

$$\begin{aligned} U(t, m_{x_0}^N) &\leq \mathbb{E} \left[\int_{t_0}^T \left(L(\widehat{X}_t, \widehat{\alpha}_t) + \mathcal{F}(\mathcal{L}(\widehat{X}_t)) \right) dt + \mathcal{G}(\mathcal{L}(\widehat{X}_T)) \right] \\ &\leq \mathbb{E} \left[\int_{t_0}^T \left(\frac{1}{N} \sum_{i=1}^N L(X_t^i, \alpha_t^i) + \mathcal{F}(\mathcal{L}(\widehat{X}_t)) \right) dt + \mathcal{G}(\mathcal{L}(\widehat{X}_T)) \right] \\ &= \mathbb{E} \left[\int_{t_0}^T \left\{ \frac{1}{N} \sum_{i=1}^N L(X_t^i, \alpha_t^i) + \mathcal{F} \left(\frac{1}{N} \sum_{i=1}^N \mathcal{L}(X_t^i) \right) \right\} dt + \mathcal{G} \left(\frac{1}{N} \sum_{i=1}^N \mathcal{L}(X_T^i) \right) \right] \\ &\leq \mathbb{E} \left[\int_{t_0}^T \left(\frac{1}{N} \sum_{i=1}^N L(X_t^i, \alpha_t^i) + \mathcal{F}(m_{\mathbf{X}_t}^N) \right) dt + \mathcal{G}(m_{\mathbf{X}_T}^N) \right]. \end{aligned}$$

Taking the infimum over the admissible controls leads to (7.1). \square

Proof of Proposition 2.8. One inequality is provided by Lemma 7.1, and the other is given by Proposition 6.2. \square

Proof of Proposition 2.10. Since \mathcal{F} and \mathcal{G} depend only on \bar{m} and L is convex, we can conclude that the infimum in the definition of U can be restricted to deterministic controls, i.e.

$$U(t, m) = \inf_{\alpha = \alpha(s): [t, T] \rightarrow \mathbb{R}^d} \left\{ \int_t^T \left(L(\alpha(s)) + \mathcal{F}(\mathcal{L}(X_s)) \right) ds + \mathcal{G}(\mathcal{L}(X_T)) \right\},$$

subject to

$$dX_s = \alpha(s) ds + \sqrt{2} dW_s, \quad s \in [t, T]; \quad \mathcal{L}(X_t) = m_0,$$

with the infimum taken over square-integrable functions α . Indeed, for any admissible control α_t (in open-loop formulation) we can get a better reward by taking the deterministic control $t \mapsto \mathbb{E}[\alpha_t]$. But if we set $\bar{m}_s = \overline{\mathcal{L}(X_s)} = \mathbb{E}[X_s]$, then we have the ODE

$$\frac{d}{dt} \bar{m}_s = \alpha(s), \quad s \in [t, T]; \quad \bar{m}_t = \bar{m}, \quad (7.6)$$

and so we can further rewrite U as

$$U(t, m) = \inf_{\alpha = \alpha(s): [t, T] \rightarrow \mathbb{R}^d} \left\{ \int_t^T \left(L(\alpha(s)) + F(\bar{m}_s) \right) ds + \mathcal{G}(\bar{m}_T) \right\},$$

subject to (7.6), and it now follows from dynamic programming that $U(t, m) = u(t, \bar{m})$.

We next argue that $V^N(t, \mathbf{x}) = v^N(t, \frac{1}{N} \sum_{i=1}^N x^i)$. For this, we note the convexity of L together with the form of \mathcal{F} and \mathcal{G} easily implies that the infimum in the definition of V^N can be restricted to open-loop feedbacks such that $\alpha_s^i = \alpha_s$, for each i and some square-integrable process α . That is, we have

$$V^N(t, \mathbf{x}) = \inf_{\alpha \in \mathcal{A}} \mathbb{E} \left[\int_t^T \left(L(\alpha_s) + \mathcal{F}(m_{\mathbf{X}_s}^N) \right) ds + \mathcal{G}(m_{\mathbf{X}_T}^N) \right]$$

with

$$dX_s^i = \alpha_s ds + \sqrt{2} dW_s^i, \quad s \in [t, T]; \quad X_t^i = x^i.$$

Next, we notice that if we set $\bar{X}_s = \frac{1}{N} \sum_{i=1}^N X_s^i$ for $s \in [t, T]$, then we have the SDE

$$d\bar{X}_s = \alpha_s ds + \frac{1}{\sqrt{N}} d\bar{W}_s^N, \quad X_t = \bar{m}_{\mathbf{x}}^N := \frac{1}{N} \sum_{i=1}^N x_i, \quad (7.7)$$

where $(\bar{W}_s^N := \frac{1}{\sqrt{N}} \sum_{i=1}^N W_s^i)_{s \in [t, T]}$ is a Brownian motion. Thus we can write

$$V^N(t, \mathbf{x}) = \inf_{\alpha} \mathbb{E} \left[\int_t^T \left(L(\alpha_s) + F(\bar{X}_s) \right) ds + G(\bar{X}_T) \right],$$

subject to (7.7). It again follows from classical representation result that $V^N(t, \mathbf{x}) = v^N(t, \frac{1}{N} \sum_{i=1}^N x^i)$. \square

Proof of Proposition 2.9. On the one hand, it is obvious that

$$U(0, \delta_0) = 0,$$

since this is exactly the cost incurred by playing the control $\alpha = 0$. Assume now that T is large enough that the density of \mathcal{N}_T is bounded above by 1, i.e. $T \geq \frac{1}{2\pi}$. Using the Cole-Hopf transform

and Feynman-Kac formula to solve the PDE (HJB(N)) for V^N , we have a simple representation formula for $V^N(0, 0)$, which reads

$$V^N(0, 0) = -\frac{1}{N} \log \int_{(\mathbb{R}^d)^N} e^{-Nd_1(m_{\xi}^N, \mathcal{N}_T)} \mathcal{N}_T^{\otimes N}(d\mathbf{x}) = -\frac{1}{N} \log \mathbb{E} \left[e^{-Nd_1(m_{\xi}^N, \mathcal{N}_T)} \right],$$

with $(\xi^i)_{i \in \mathbb{N}}$ being i.d.d. with common law \mathcal{N}_T , and with $m_{\xi}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i}$. Our goal is to use this formula to prove that for infinitely many N , $V^N(0, 0) \geq cN^{-1/d}$, with $c > 0$ being independent of N . Clearly it suffices to show that there are constants C and c such that

$$\mathbb{E} \left[e^{-Nd_1(m_{\xi}^N, \mathcal{N}_T)} \right] \leq C e^{-cN^{1-1/d}}, \quad (7.8)$$

for infinitely many N . In order to establish (7.8), we fix $N \in \mathbb{N}$ such that $N = M^d$ for some $M \in \mathbb{N}$, and we consider a partition of the unit cube $\mathcal{C} = [0, 1]^d$ into N cubes $(\mathcal{C}^{N,i})_{i=1, \dots, N}$ of side length $N^{-1/d} = 1/M$. We next consider, for $p \in (0, 1)$, the event $A_{p,N}$ defined by

$$A_{p,N} := \left\{ \left| \{i : m_{\xi}^N(\mathcal{C}^{N,i}) > 0\} \right| > (1-p)N \right\} = \left\{ \left| \{i : \xi_j \in \mathcal{C}^{N,i} \text{ for some } 1 \leq j \leq N\} \right| > (1-p)N \right\}.$$

In other words, $A_{p,N}$ is the event that at least $(1-p)N$ of the N cubes $(\mathcal{C}^{N,i})_{i=1, \dots, N}$ are charged by the random measure m_{ξ}^N . We will now state two lemmas about the event $A_{p,N}$, both of which will be proved below.

Lemma 7.2. *On the event $(A_{p,N})^c$, we have*

$$d_1(m_{\xi}^N, \mathcal{N}_T) \geq cpN^{-1/d},$$

for some $c > 0$ independent of p and N .

Lemma 7.3. *For $p > 0$ small enough that $p < \frac{(1-p)^2}{8}$ we have*

$$\mathbb{P}[A_{p,N}] \leq C e^{-c(p)N},$$

where C is some constant independent of N and $c(p) = \frac{(1-p)^2}{8} - p$.

With these lemmas in hand, we can easily conclude the proof, since for $p > 0$ small enough we have

$$\begin{aligned} \mathbb{E} \left[e^{-Nd_1(m_{\xi}^N, \mathcal{N}_T)} \right] &\leq \mathbb{P}[A_{p,N}] + (1 - \mathbb{P}[A_{p,N}]) e^{-cN^{1-1/d}} \\ &\leq C e^{-cN} + e^{-cN^{1-1/d}} \leq C e^{-cN^{1-1/d}}, \end{aligned}$$

for some c, C independent of N , which clearly implies (7.8). \square

We now prove the two lemmas used in the above example.

Proof of Lemma 7.2. Fix $\omega \in (A_{p,N})^c$. Define a set $I \subset \{1, \dots, N\}$ by

$$I := \left\{ i \in \{1, \dots, N\} : m_{\xi(\omega)}(\mathcal{C}^{N,i}) = 0 \right\},$$

and then set $\mathcal{C}^I = \cup_{i \in I} \mathcal{C}^{N,i}$. Then by the definition of $A_{p,N}$, we must have $|I| \geq pN$. Now let γ be any coupling between $m_{\xi(\omega)}^N$ and \mathcal{N}_T . That is, $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$, and we have

$$\gamma(A \times \mathbb{R}^d) = m_{\xi(\omega)}(A), \quad \gamma(\mathbb{R}^d \times A) = \mathcal{N}_T(A),$$

for any Borel set $A \subset \mathbb{R}^d$. Finally, for $i = 1, \dots, N$, let $\mathcal{C}_{1/2}^{N,i}$ denote the cube with the same center and half the side-length of $\mathcal{C}^{N,i}$. Then we have

$$\begin{aligned} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y| d\gamma(x, y) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y| \mathbf{1}_{x \in (\mathcal{C}^I)^c} d\gamma(x, y) \\ &\geq \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(|x - y| \mathbf{1}_{x \in (\mathcal{C}^I)^c} \sum_{j \in I} \mathbf{1}_{y \in \mathcal{C}_{1/2}^{N,j}} \right) d\gamma(x, y) \\ &\geq \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(cN^{-1/d} \mathbf{1}_{x \in (\mathcal{C}^I)^c} \sum_{j \in I} \mathbf{1}_{y \in \mathcal{C}_{1/2}^j} \right) d\gamma(x, y) \\ &= cN^{-1/d} \sum_{j \in I} \mathcal{N}_T(\mathcal{C}_{1/2}^j) \geq cpN^{-1/d}, \end{aligned}$$

where in the last step we used the fact that the density of \mathcal{N}_T is bounded from below on \mathcal{C} . Taking an infimum over the couplings γ shows that

$$d_1(m_\xi(\omega), \mathcal{N}_T) \geq cpN^{-1/d},$$

for each $\omega \in (A_{p,N})^c$, which completes the proof. \square

Proof of Lemma 7.3. First, we make a simplifying observation. Let $(U^i)_{i \in \mathbb{N}}$ denote a sequence of i.i.d. random variables each of which is uniformly distributed on the unit cube \mathcal{C} , i.e. $\mathcal{L}(U^i)$ is the Lebesgue measure restricted to \mathcal{C} . For $\mathbf{U} := (U^1, \dots, U^N)$, let $m_{\mathbf{U}}^N = \frac{1}{N} \sum_{i=1}^N \delta_{U^i}$, and define

$$B_{p,N} = \left\{ \left| \{i : m_{\mathbf{U}}^N(\mathcal{C}^{N,i}) > 0\} \right| > (1-p)N \right\} = \left\{ \left| \{i : U_j \in \mathcal{C}^{N,i} \text{ for some } 1 \leq j \leq N\} \right| > (1-p)N \right\}.$$

In other words, $B_{p,N}$ is defined exactly as $A_{p,N}$, but with the i.i.d. uniform random variables $(U^i)_{i=1, \dots, N}$ replacing the i.i.d. Gaussian random variables $(\xi^i)_{i=1, \dots, N}$. Thanks to the fact that we assumed T is large enough that the density of \mathcal{N}_T is bounded above by 1, it is clear that $\mathbb{P}[A_{p,N}] \leq \mathbb{P}[B_{p,N}]$, and so it suffices to prove the estimate

$$\mathbb{P}[B_{p,N}] \leq Ce^{-cN}, \quad (7.9)$$

for p small enough and C, c independent of N . We mention that $\mathbb{P}[B_{p,N}]$ is connected to the well-known ‘‘coupon collector problem’’ - in the language of this problem, $\mathbb{P}[B_{p,N}]$ is the probability that the coupon collector collects at least $(1-p)N$ distinct coupons in the first N draws from N distinct coupons.

To analyze the probability $\mathbb{P}[B_{p,N}]$, we first define an increasing sequence of random times $(T_m)_{m \in \mathbb{N}}$ by setting $T_0 := 0$ and

$$T_m := \inf \left\{ j \in \mathbb{N} : \left| \{i : m_{\mathbf{U}}^j(\mathcal{C}^{N,i}) > 0\} \right| = m \right\}, \quad m \in \mathbb{N},$$

and then we set $\tau_m := T_m - T_{m-1}$ for $m \in \mathbb{N} \setminus \{0\}$. So in the language of the coupon collector problem, T_m is the number of draws needed to find the m^{th} distinct coupon, and τ_m is the number of draws between the discovery of the m^{th} distinct coupon and that of the $(m-1)^{\text{th}}$ distinct coupon. It is easy to check that (using the notation $\lceil \cdot \rceil$ for the ceil part)

$$\mathbb{P}[B_{p,N}] \leq \mathbb{P}[\tau_1 + \dots + \tau_{\lceil (1-p)N \rceil} \leq N],$$

and that $(\tau_i)_{i \in \mathbb{N}}$ are independent, with $\tau_i \sim \text{Geom}(1 - \frac{i-1}{N})$ i.e. for $k \in \mathbb{N} \setminus \{0\}$,

$$\mathbb{P}[\tau^i = k] = \left(\frac{i-1}{N} \right)^{k-1} \left(1 - \frac{i-1}{N} \right).$$

Using the well-known formula for the moment generating function of a geometric random variable, we find that

$$\mathbb{E}[e^{-\lambda\tau_i}] = \frac{e^{-\lambda}\left(1 - \frac{i-1}{N}\right)}{1 - e^{-\lambda}\left(\frac{i-1}{N}\right)},$$

for each $\lambda > 0$. Thus, we see that for $\lambda > 0$, we have

$$\begin{aligned} \mathbb{P}[\tau_1 + \dots + \tau_{\lceil(1-p)N\rceil} \leq N] &= \mathbb{P}\left[e^{-\lambda(\tau_1 + \dots + \tau_{\lceil(1-p)N\rceil})} \geq e^{-\lambda N}\right] \\ &\leq e^{\lambda N} \mathbb{E}\left[e^{-\lambda(\tau_1 + \dots + \tau_{\lceil(1-p)N\rceil})}\right] \\ &= e^{\lambda N} \prod_{i=1}^{\lceil(1-p)N\rceil} \mathbb{E}\left[e^{-\lambda\tau_i}\right] \\ &= e^{\lambda N} \prod_{i=1}^{\lceil(1-p)N\rceil} \frac{e^{-\lambda}\left(1 - \frac{i-1}{N}\right)}{1 - e^{-\lambda}\left(\frac{i-1}{N}\right)} \\ &= \exp\left\{\lfloor pN \rfloor \lambda + \sum_{i=1}^{\lceil(1-p)N\rceil} \left(\log\left(1 - \frac{i-1}{N}\right) - \log\left(1 - e^{-\lambda}\left(\frac{i-1}{N}\right)\right)\right)\right\}. \end{aligned} \quad (7.10)$$

To go further, we note that

$$\begin{aligned} \log\left(1 - \frac{i-1}{N}\right) - \log\left(1 - e^{-\lambda}\frac{i-1}{N}\right) &= \sum_{n=1}^{\infty} n^{-1} e^{-\lambda n} \left(\frac{i-1}{N}\right)^n - \sum_{n=1}^{\infty} n^{-1} \left(\frac{i-1}{N}\right)^n \\ &\leq (e^{-\lambda} - 1) \left(\frac{i-1}{n}\right). \end{aligned} \quad (7.11)$$

Notice that for any $0 < \lambda \leq 1$, we have $\exp(-\lambda) - 1 \leq -\frac{\lambda}{4}$, and so combining (7.10) with (7.11) shows that for $0 < \lambda \leq 1$,

$$\begin{aligned} \mathbb{P}[\tau_1 + \dots + \tau_{\lceil(1-p)N\rceil} \leq N] &\leq \exp\left\{\lfloor pN \rfloor \lambda - (1 - e^{-\lambda}) \sum_{i=1}^{\lceil(1-p)N\rceil} \frac{i-1}{N}\right\} \\ &\leq \exp\left\{pN\lambda - (1 - e^{-\lambda})(1-p)^2 \frac{N}{2}\right\} \\ &\leq \exp\left\{pN\lambda - \frac{1}{8}\lambda N(1-p)^2\right\} = \exp\left\{\lambda N\left(p - \frac{(1-p)^2}{8}\right)\right\}. \end{aligned}$$

Fixing $\lambda = 1$, we see that if p small enough that $p < \frac{(1-p)^2}{8}$, then we have

$$\mathbb{P}[B_{p,N}] \leq \exp(-cN).$$

with $c = \left(p - \frac{(1-p)^2}{8}\right)$. This concludes the proof. \square

APPENDIX A. SOME AUXILIARY ESTIMATES FOR FINITE-DIMENSIONAL PDES

In this appendix, we first give some results about linear and semi-linear parabolic equations that were used in Section 3. We also give a sketch of proof of Lemma 4.1 in Subsection B.

The following form of Grönwall's Lemma will be useful (see [YGD07]):

Lemma A.1. *Assume that $l \in L^\infty([0, T])$ satisfies, for some non-decreasing measurable $C_1 : [0, T] \rightarrow \mathbb{R}^+$, some $C_2 > 0$ and all $t \in [0, T]$,*

$$l(t) \leq C_1(t) + C_2 \int_t^T \frac{l(s)}{\sqrt{s-t}} ds,$$

then, there exists a constant $C(C_2, T)$ (depending on C_2 and T) such that, for all $t \in [0, T]$,

$$l(t) \leq C_1(t)C(C_2, T).$$

A.1. Linear PDEs.

Lemma A.2. *Suppose that $\alpha : [t_0, T] \times \mathbb{T}^d \rightarrow \mathbb{R}^d$ is Lipschitz continuous in x uniformly in t , take $g \in C^1(\mathbb{T}^d)$, $t_1 \in (t_0, T]$ and let v be the solution to*

$$-\partial_t v(t, x) - \Delta_x v(t, x) - \alpha(t, x) \cdot D_x v(t, x) = 0 \text{ in } (t_0, t_1) \times \mathbb{T}^d, \quad v(t_1, x) = g(x) \text{ in } \mathbb{T}^d. \quad (\text{A.1})$$

Then,

- If α satisfies

$$\sup_{t_0 \leq t \leq T} \|\alpha(t, \cdot)\|_{L^\infty} + \sup_{t_0 \leq t \leq T} \sqrt{T-t} \|D_x \alpha(t, \cdot)\|_{L^\infty} \leq C_1, \quad (\text{A.2})$$

for some $C_1 > 0$. Then, there is C'_1 depending only on C_1 such that

$$\sup_{t \in [t_0, t_1]} \|D_x v(t, \cdot)\|_{L^\infty} + \sup_{t \in [t_0, t_1]} \sqrt{t_1 - t} \|D_{xx}^2 v(t, \cdot)\|_{L^\infty} \leq C'_1 \|D_x g\|_{L^\infty}. \quad (\text{A.3})$$

- If we further assume that $g \in H^s$ and $\sup_{t_0 \leq t \leq T} \|\alpha(t, \cdot)\|_{s-1} \leq C_2$ for some $C_2 > 0$, then

$$\sup_{t \in [t_0, t_1]} \|v(t, \cdot)\|_{H^s} \leq C'_2 \|g\|_{H^s}, \quad (\text{A.4})$$

for some C'_2 depending only on C_2 .

Proof. We start with the proof of (A.3) Let us write P_t for the heat semi-group on \mathbb{T}^d , i.e. given $\phi : \mathbb{T}^d \rightarrow \mathbb{R}$,

$$(P_t \phi)(x) = \psi(t, x), \quad \text{where } \psi \text{ solves } \partial_t \psi = \Delta_x \psi, \quad \psi(0, x) = \phi(x).$$

We note for later use that we have the classical estimates for P :

$$\|P_t f\|_{L^\infty} \leq \|f\|_{L^\infty}, \quad \|D_x P_t f\|_{L^\infty} \leq \frac{C}{\sqrt{t}} \|f\|_{L^\infty}, \quad (\text{A.5})$$

$$\|P_t f\|_s \leq \|f\|_s, \quad \|P_t f\|_s \leq \frac{C}{\sqrt{t}} \|f\|_{s-1}. \quad (\text{A.6})$$

Notice that the solution v to (A.1) must satisfy for $(t, x) \in [t_0, t_1] \times \mathbb{T}^d$

$$v(t, x) = P_{t_1-t} g(x) + \int_t^{t_1} P_{s-t} [\alpha(s, \cdot) \cdot D_x v(s, \cdot)](x) ds. \quad (\text{A.7})$$

And therefore, differentiating (A.7) in x , it holds that

$$D_x v(t, x) = P_{t_1-t} D_x g(x) + \int_t^{t_1} D_x P_{s-t} [\alpha(s, \cdot) \cdot D_x v(s, \cdot)](x) ds,$$

which, in turn, using the smoothing properties (A.5), leads to

$$\|D_x v(t, \cdot)\|_{L^\infty} \leq \|D_x g\|_{L^\infty} + \sup_{t \in [t_0, t_1]} \|\alpha(t, \cdot)\|_{L^\infty} \int_t^{t_1} \frac{\|D_x v(s, \cdot)\|_{L^\infty}}{\sqrt{s-t}} ds,$$

and we conclude with Lemma A.1 that inequality

$$\sup_{t \in [t_0, t_1]} \|D_x v(t, \cdot)\|_{L^\infty} \leq C \|D_x g\|_{L^\infty} \quad (\text{A.8})$$

holds for some $C = C(C_1, T)$. For the estimate on $\|D_{xx}^2 v(t, \cdot)\|_\infty$, we differentiate (A.7) as follows

$$D_{xx}^2 v(t, x) = D_x P_{t_1-t} D_x g(x) + \int_t^{t_1} D_x P_{s-t} [D_x \alpha(s, \cdot) D_x v(s, \cdot) + D_{xx}^2 v(s, \cdot) \alpha(s, \cdot)](x) ds.$$

And therefore, using the smoothing properties (A.5) and the regularity condition (A.2) on α we have

$$\begin{aligned} \|D_{xx}^2 v(t, \cdot)\|_{L^\infty} &\leq \frac{\|D_x g\|_{L^\infty}}{\sqrt{t_1-t}} + \sup_{t \in [t_0, t_1]} \|D_x v(t, \cdot)\|_{L^\infty} \int_t^{t_1} \frac{\|D_x \alpha(s, \cdot)\|_{L^\infty}}{\sqrt{s-t}} ds \\ &\quad + \sup_{t \in [t_0, t_1]} \|\alpha(t, \cdot)\|_{L^\infty} \int_t^{t_1} \frac{\|D_{xx}^2 v(s, \cdot)\|_{L^\infty}}{\sqrt{t-s}} ds \\ &\leq \frac{\|D_x g\|_{L^\infty}}{\sqrt{t_1-t}} + C_1 \sup_{t \in [t_0, t_1]} \|D_x v(t, \cdot)\|_{L^\infty} \int_t^{t_1} \frac{1}{\sqrt{T-s}\sqrt{s-t}} ds \\ &\quad + C_1 \int_t^{t_1} \frac{\|D_{xx}^2 v(s, \cdot)\|_{L^\infty}}{\sqrt{t-s}} ds. \end{aligned}$$

Noticing that

$$\int_t^{t_1} \frac{1}{\sqrt{T-s}\sqrt{s-t}} ds \leq \int_t^T \frac{1}{\sqrt{T-s}\sqrt{s-t}} ds = \pi,$$

we get, applying Gronwall's Lemma A.1,

$$\|D_{xx}^2 v(t, \cdot)\|_{L^\infty} \leq C \left(\sup_{t \in [t_0, t_1]} \|D_x v(t, \cdot)\|_{L^\infty} + \frac{\|D_x g\|_{L^\infty}}{\sqrt{t_1-t}} \right),$$

for some $C = C(C_1, T)$. We conclude using the bound (A.8) on $D_x v$.

We continue with the proof of (A.4) under the corresponding assumption. Being $s-1 > d/2$, there is some C_s depending only on s such that $\|fg\|_{H^{s-1}} \leq C_s \|f\|_{H^{s-1}} \|g\|_{H^{s-1}}$ whenever $f, g \in H^{s-1}$. Therefore, using the smoothness assumed on α , we have for each fixed $r \in [t_0, t_1]$

$$\|\alpha(r, \cdot) \cdot D_x v(r, \cdot)\|_{H^{s-1}} \leq C \|D_x v(r, \cdot)\|_{H^{s-1}} \leq C \|v\|_{H^s}.$$

Combining this with the smoothing property (A.6), we have

$$\|P_{r-t}(\alpha(r, \cdot) \cdot D_x v(r, \cdot))\|_{H^s} \leq \frac{C}{\sqrt{r-t}} \|v(r, \cdot)\|_{H^s}. \quad (\text{A.9})$$

But now using (A.7), we have, for all $t \in [t_0, t_1]$,

$$\begin{aligned} \|v(t, \cdot)\|_{H^s} &\leq \|g\|_{H^s} + \int_t^{t_1} \|P_{r-t}(b(r, \cdot) \cdot D_x v(r, \cdot))\|_{H^s} dr \\ &\leq \|g\|_{H^s} + C \int_t^{t_1} \frac{\|v(r, \cdot)\|_{H^s}}{\sqrt{r-t}} dr. \end{aligned}$$

Using Grönwall Lemma A.1, we get (A.4). □

Now we can prove the stability Lemmas 3.3 and 3.5.

Proof of Lemma 3.3. We argue by duality using Lemma A.2 above. Indeed, for any m^1, m^2 satisfying

$$\partial_t m_t^i = \Delta m_t^i - \operatorname{div}(m_t^i \alpha(t, \cdot)),$$

for $(t, x) \in [t_0, T] \times \mathbb{T}^d$, any $g \in H^s$ and any $t_1 \in [t_0, T]$, we have

$$\int_{\mathbb{T}^d} g(x)(m_{t_1}^1 - m_{t_1}^2)(dx) = \int_{\mathbb{T}^d} v(t_0, x)(m_{t_0}^1 - m_{t_0}^2)(dx),$$

where v solves (A.1). As a consequence,

$$\int_{\mathbb{T}^d} g(x)(m_{t_1}^1 - m_{t_1}^2)(dx) \leq \|v(t_0, \cdot)\|_{H^s} \|m_{t_0}^1 - m_{t_0}^2\|_{H^{-s}} \leq C \|g\|_{H^s} \|m_{t_0}^1 - m_{t_0}^2\|_{H^{-s}},$$

and so taking a supremum over g with $\|g\|_{H^s} \leq 1$ gives the result. \square

Proof of Lemma 3.5. We argue once again by duality to deduce that, for all $t_1 \in [t_0, T]$ and all 1-Lipschitz function g ,

$$\begin{aligned} \int_{\mathbb{T}^d} g(x)(m_{t_1}^1 - m_{t_1}^2)(dx) &= \int_{\mathbb{T}^d} v(t_0, x)(m_{t_0}^1 - m_{t_0}^2)(dx) \\ &\leq \|D_x v(t_0, \cdot)\|_{L^\infty} d_1(m_{t_0}^1, m_{t_0}^2) \leq C'_1 d_1(m_{t_0}^1, m_{t_0}^2), \end{aligned}$$

where v is solution to (A.1) with terminal data g and C'_1 is the constant appearing in (A.3). Taking the supremum over 1-Lipschitz functions g gives the first statement in Lemma 3.5. For the second statement we proceed similarly. Once again, let $t_1 \in (t_0, T]$, let g be a 1-Lipschitz function and v the solution to (A.1) with terminal data g . Using (A.3) we infer that

$$\begin{aligned} \int_{\mathbb{T}^d} g(x)(m_{t_1}^1 - m_{t_1}^2)(dx) &= \int_{\mathbb{T}^d} v(t_0, x)(m_{t_0}^1 - m_{t_0}^2)(dx) \\ &\leq \left(\|D_x v(t_0, \cdot)\|_{L^\infty} + \|D_{xx}^2 v(t_0, \cdot)\|_{L^\infty} \right) \|m_{t_0}^1 - m_{t_0}^2\|_{-2, \infty} \\ &\leq \frac{C'_1}{\sqrt{t_1 - t_0}} \|m_{t_0}^1 - m_{t_0}^2\|_{-2, \infty}. \end{aligned}$$

Taking the supremum over 1-Lipschitz functions g gives the result. \square

A.2. HJB equation.

Lemma A.3. *Assume that H satisfies (1) and (2) in Assumption (2.1). Take $f \in L^\infty([t_0, T], \mathcal{C}^2(\mathbb{T}^d))$ and $g \in \mathcal{C}^2(\mathbb{T}^d)$. Let $u \in C^{1,2}([t_0, T] \times \mathbb{T}^d)$ be the solution to*

$$-\partial_t u(t, x) + H(x, D_x u(t, x)) - \Delta_x u(t, x) = f(t, x) \text{ in } (t_0, T) \times \mathbb{T}^d, \quad u(T, x) = g(x) \text{ in } \mathbb{T}^d. \quad (\text{A.10})$$

Then $\|D_x u\|_{L^\infty} + \sup_{t \in [t_0, T]} \sqrt{T-t} \|D_{xx}^2 u(t, \cdot)\|_{L^\infty} \leq C$ for some $C = C(\|D_x f\|_{L^\infty}, \|D_x g\|_{L^\infty}) > 0$. If we further assume that $H \in \mathcal{C}^s(\mathbb{T}^d \times \mathbb{R}^d)$, $f \in L^\infty([0, T], \mathcal{C}^s(\mathbb{T}^d))$ and $g \in \mathcal{C}^s(\mathbb{T}^d)$ then

$$\sup_{t \in [t_0, T]} \|u(t, \cdot)\|_{\mathcal{C}^s} + \sup_{t \in [t_0, T]} \sqrt{T-t} \|u(t, \cdot)\|_{\mathcal{C}^{s+1}} \leq C'$$

for some $C' = C'(\sup_{t \in [t_0, T]} \|f(t, \cdot)\|_{\mathcal{C}^s}, \|g\|_{\mathcal{C}^s})$.

Proof. The Lipschitz estimate is standard and follows from the classical Bernstein method and the condition on the growth of $D_x H$ in Assumption (2.1). Therefore, there is $C = C(\|D_x f\|_{L^\infty}, \|D_x g\|_{L^\infty})$ such that

$$\|D_x u\|_{L^\infty} \leq C. \quad (\text{A.11})$$

Being a classical solution, u must satisfy, for all $(t, x) \in [t_0, T] \times \mathbb{T}^d$,

$$u(t, x) = P_{T-t} g(x) + \int_t^T P_{s-t} f(s, \cdot)(x) ds - \int_t^T P_{s-t} [H(\cdot, D_x u(s, \cdot))](x) ds.$$

Differentiating twice and integrating by parts leads, for all $(t, x) \in [t_0, T] \times \mathbb{T}^d$ to

$$D_{xx}^2 u(t, x) = D_x P_{T-t} D_x g(x) + \int_t^T D_x P_{s-t} D_x f(s, \cdot)(x) ds - \int_t^T D_x P_{s-t} D_x [H(\cdot, D_x u(s, \cdot))](x) ds. \quad (\text{A.12})$$

Using the smoothing properties (A.5), we deduce that inequality

$$\|D_{xx}^2 u(t, \cdot)\|_{L^\infty} \leq \frac{\|D_x g\|_{L^\infty}}{\sqrt{T-t}} + \int_t^T \frac{\|D_x f\|_{L^\infty}}{\sqrt{s-t}} ds + \int_t^T \frac{C(1 + \|D_{xx}^2 u(s, \cdot)\|_{L^\infty})}{\sqrt{s-t}} ds$$

holds for some $C = C(\|D_x u\|_{L^\infty}) > 0$ and for all $t \in [t_0, T]$. As a consequence, taking (A.11) into account we can find $C = C(\|D_x f\|_{L^\infty}, \|D_x g\|_{L^\infty}) > 0$ such that inequality

$$\|D_{xx}^2 u(t, \cdot)\|_{L^\infty} \leq C \left(1 + \frac{1}{\sqrt{T-t}} + \int_t^T \frac{\|D_{xx}^2 u(s, \cdot)\|_{L^\infty}}{\sqrt{s-t}} ds \right)$$

holds for all $t \in [t_0, T]$ and we deduce from Grönwall's Lemma (A.1) that

$$\|D_{xx}^2 u(t, \cdot)\|_{L^\infty} \leq C \left(1 + \frac{1}{\sqrt{T-t}} \right),$$

for some $C = C(\|D_x f\|_{L^\infty}, \|D_x g\|_{L^\infty}) > 0$.

For the second part of the lemma, we easily prove by induction that

$$\sup_{t \in [t_0, T]} \|u(t, \cdot)\|_{C^s} \leq C,$$

for some $C = C(\sup_{t \in [t_0, T]} \|f(t, \cdot)\|_{C^s}, \|g\|_{C^s}) > 0$ and we deduce the bound on $\|u(t, \cdot)\|_{C^{s+1}}$ as we got the bound on $\|D_{xx}^2 u(t, \cdot)\|_{L^\infty}$. □

APPENDIX B. A MOLLIFICATION ARGUMENT

We provide the proof of Lemma 4.1.

Proof. Step 1. We introduce the following notation. For any $n \geq 1$, we denote by $\mathcal{P}_n(\mathbb{T}^d)$ the collection of probability measures whose Fourier coefficients of order greater than n are null, i.e.

$$\mathcal{P}_n(\mathbb{T}^d) := \{m \in \mathcal{P}(\mathbb{T}^d) : \widehat{m}^{\mathbf{k}} = 0, \quad |\mathbf{k}|_\infty \geq n\},$$

where \mathbf{k} in the above notation is implicitly understood as a multi-index $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d$ of sup norm $|\mathbf{k}|_\infty := \max(|k_1|, \dots, |k_d|)$. Accordingly, the set $\mathcal{P}_n(\mathbb{T}^d)$ can be identified with a finite-dimensional set \mathcal{O}_n describing the Fourier coefficients of order less than n . The set \mathcal{O}_n can be described in an exhaustive manner by means of Bochner's theorem. We just refer to [CD22] for the details as the precise formulation of \mathcal{O}_n does not really matter for our needs. The only two things that we need are

- (1) The set \mathcal{O}_n can be regarded as a subset of $\mathbb{R}^{|D(n)|} \times \mathbb{R}^{|D(n)|}$ and it contains an open ball $\mathcal{B}_n \times \mathcal{B}_n$ centered at the origin of $\mathbb{R}^{|D(n)|} \times \mathbb{R}^{|D(n)|}$, where $D(n)$ is a certain subset of $\{\mathbf{k} \in \mathbb{Z}^d \setminus \{0\} : |\mathbf{k}|_\infty < n\}$ and $|D(n)|$ denotes its cardinal.
- (2) Denoting by

$$\begin{aligned} \mathcal{I}_n &: \left((a_{\mathbf{k}})_{\mathbf{k} \in D(n)}, (b_{\mathbf{k}})_{\mathbf{k} \in D(n)} \right) \in \mathbb{R}^{|D(n)|} \times \mathbb{R}^{|D(n)|} \\ &\mapsto \left(x \in \mathbb{T}^d \mapsto 1 + 2 \sum_{\mathbf{k} \in D(n)} (a_{\mathbf{k}} \cos(2\pi \mathbf{k} \cdot x) + b_{\mathbf{k}} \sin(2\pi \mathbf{k} \cdot x)) \right), \end{aligned}$$

the mapping \mathcal{I}_n is one-to-one from \mathcal{O}_n onto $\mathcal{P}_n(\mathbb{T}^d)$ (here, the additional 1 is regarded as the constant function on the torus, equal to 1). For $m \in \mathcal{P}_n(\mathbb{T}^d)$, its pre-image by \mathcal{I}_n writes

$$\mathcal{I}_n^{-1}(m) = \left(\mathcal{R}[\widehat{m}^{\mathbf{k}}], \mathfrak{I}[\widehat{m}^{\mathbf{k}}] \right)_{\mathbf{k} \in D(n)},$$

where $\mathcal{R}(z)$ and $\mathfrak{I}(z)$ denote the real and imaginary parts of a complex number z .

The next ingredient that is needed in the proof is the Féjer kernel of rank n , which we denote by $\varphi_n : \mathbb{T}^d \rightarrow \mathbb{R}$. We recall that φ_n is a (symmetric) probability density whose Fourier coefficients are given by

$$\widehat{\varphi}^{\mathbf{k}} = \prod_{j=1}^d \left(1 - \frac{k_j}{n}\right), \quad \text{if } |\mathbf{k}|_\infty < n,$$

$$\widehat{\varphi}^{\mathbf{k}} = 0, \quad \text{if } |\mathbf{k}|_\infty \geq n.$$

for any multi-index $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d$. In particular, for any $m \in \mathcal{P}(\mathbb{T}^d)$, the convolution $m * \varphi_n$ is an element of $\mathcal{P}_n(\mathbb{T}^d)$. Also, for $\eta \in [0, 1]$ and for $(\mathbf{a}, \mathbf{b}) := ((a_{\mathbf{k}})_{\mathbf{k} \in D(n)}, (b_{\mathbf{k}})_{\mathbf{k} \in D(n)}) \in \mathcal{B}_n$, the combination

$$(1 - \eta)m * \varphi_n + \eta \mathcal{I}_n(\mathbf{a}, \mathbf{b})$$

is an element of $\mathcal{P}_n(\mathbb{T}^d)$. Its pre-image by \mathcal{I}_n is

$$\begin{aligned} & \mathcal{I}_n^{-1} \left((1 - \eta)m * \varphi_n + \eta \mathcal{I}_n(\mathbf{a}, \mathbf{b}) \right) \\ &= \left(\left[(1 - \eta)\mathcal{R}[\widehat{m}^{\mathbf{k}}] \widehat{\varphi}_n^{\mathbf{k}} + \eta a_{\mathbf{k}} \right], \left[(1 - \eta)\mathfrak{I}[\widehat{m}^{\mathbf{k}}] \widehat{\varphi}_n^{\mathbf{k}} + \eta b_{\mathbf{k}} \right] \right)_{\mathbf{k} \in D(n)}. \end{aligned}$$

Step 2. For $\eta \in (0, 1)$, we now let:

$$\Phi^{\eta, n}(m) = \int_{\mathbb{R}^{|D(n)|} \times \mathbb{R}^{|D(n)|}} \Phi \left((1 - \eta)m * \varphi_n + \eta \mathcal{I}_n(\mathbf{a}, \mathbf{b}) \right) \rho_n(\mathbf{a}) \rho_n(\mathbf{b}) d\mathbf{a} d\mathbf{b},$$

for ρ_n a smooth kernel whose support is included in \mathcal{B}_n . By an obvious change of variable, we have

$$\begin{aligned} \Phi^{\eta, n}(m) &= \int_{\mathbb{R}^{|D(n)|} \times \mathbb{R}^{|D(n)|}} \left[\Phi(\mathcal{I}_n(\mathbf{a}, \mathbf{b})) \right. \\ &\quad \left. \times \rho_n \left(\left(\frac{a_{\mathbf{k}} - (1 - \eta)\mathcal{R}[\widehat{m}^{\mathbf{k}}] \widehat{\varphi}_n^{\mathbf{k}}}{\eta} \right)_{\mathbf{k} \in D(n)} \right) \rho_n \left(\left(\frac{b_{\mathbf{k}} - (1 - \eta)\mathfrak{I}[\widehat{m}^{\mathbf{k}}] \widehat{\varphi}_n^{\mathbf{k}}}{\epsilon} \right)_{\mathbf{k} \in D(n)} \right) \right] d\mathbf{a} d\mathbf{b}, \end{aligned}$$

from which we deduce that $\Phi^{\eta, n}$ is infinitely differentiable with respect to the real and imaginary parts of the Fourier coefficients of the inputs. In particular, the latter says that $\Phi^{\eta, n}$ can be regarded as a smooth function on \mathcal{O}_n . The analysis achieved in [CD22] shows that $\Phi^{\eta, n}$ is continuously differentiable in m . The derivative $[\delta\Phi^{\eta, n}/\delta m](m, \cdot)$ has, at any $m \in \mathcal{P}(\mathbb{T}^d)$, a finite number of non-zero Fourier modes and is thus smooth. This shows (3) in the statement.

Step 3. We now prove (2) in the statement. Lipschitz property (w.r.t. d_1) is just a consequence of the following two facts: Φ is c_1 -Lipschitz continuous w.r.t. d_1 and

$$d_1(m * \varphi_n, m' * \varphi_n) \leq d_1(m, m'), \quad m, m' \in \mathcal{P}(\mathbb{T}^d),$$

i.e. convolution by Féjer kernel is a contraction under d_1 . The proof of the above inequality is quite obvious. For two $m, m' \in \mathcal{P}(\mathbb{T}^d)$, we call π an optimal coupling between m, m' for the d_1 -distance. Then, the probability measure defined by

$$\pi'(A \times B) = \int_{\mathbb{T}^d \times \mathbb{T}^d} \pi((A + z) \times (B + z)) d\varphi_n(z),$$

for any two Borel subsets A and B of \mathbb{T}^d , is a coupling between $m * \varphi_n$ and $m' * \varphi_n$. Obviously,

$$d_1(m * \varphi_n, m' * \varphi_n) \leq \int_{\mathbb{T}^d \times \mathbb{T}^d} |x - y| d\pi'(x, y) = \int_{\mathbb{T}^d \times \mathbb{T}^d} |x - y| d\pi(x, y) = d_1(m, m').$$

As for the proof of the semi-concavity property, we already know that, for any $\lambda \in [0, 1]$,

$$\Phi(\lambda m' + (1 - \lambda)m) \geq \lambda \Phi(m') + (1 - \lambda)\Phi(m) - \frac{c_2}{2} \lambda(1 - \lambda) d_1^2(m, m'), \quad m, m' \in \mathcal{P}(\mathbb{T}^d).$$

Inserting this property in the definition of $\Phi^{\eta, n}$ and using again the contractive property of the convolution under d_1 , we deduce that $\Phi^{\eta, n}$ is also c_2 -semi-concave w.r.t. d_1 .

Step 4. We now prove (1). To do so, it suffices to observe that

$$\begin{aligned} d_1((1 - \eta)m * \varphi_n + \eta \mathcal{S}(\mathbf{a}, \mathbf{b}), m) &\leq (1 - \eta) d_1(m * \varphi_n, m) + \eta d_1(\mathcal{S}(\mathbf{a}, \mathbf{b}) * \varphi_n, m) \\ &\leq (1 - \eta) d_1(m * \varphi_n, m) + c_d \eta, \end{aligned}$$

for a constant c_d only depending on the dimension (this following from the fact that, on the torus, d_1 is bounded by the total variation distance up to a multiplicative constant).

It then remains to see that

$$d_1(m * \varphi_n, m) \leq \sup_{f_{1\text{-Lip}}} \int_{\mathbb{T}^d} f(x) d(m * \varphi_n - m)(x) \leq \sup_{f_{1\text{-Lip}}} \|f - f * \varphi_n\|_\infty.$$

The latter right-hand side tends to 0 as n tends to ∞ , see again [CD22]. \square

REFERENCES

- [AKT84] M. Ajtai, J. Komlós, and G. Tusnády. On optimal matchings. *Combinatorica*, 4(4):259–264, 1984.
- [Bar94] Guy Barles. *Solutions de viscosité des équations de Hamilton-Jacobi*. Springer Berlin, Heidelberg, 1994.
- [BC18] Ariela Briani and Pierre Cardaliaguet. Stable solutions in potential mean field game systems. *NoDEA Nonlinear Differential Equations Appl.*, 25(1):Paper No. 1, 26, 2018.
- [BCD97] Martino Bardi and Italo Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA, 1997. With appendices by Maurizio Falcone and Pierpaolo Soravia.
- [Ber21] Charles Bertucci. Monotone solutions for mean field games master equations : continuous state space and common noise. *arXiv*, <https://arxiv.org/abs/2107.09531>, 2021.
- [BFY15] Alain Bensoussan, Jens Frehse, and Sheung Chi Phillip Yam. The master equation in mean field theory. *J. Math. Pures Appl.*, 103(6):1441–1474, 2015.
- [BIRS20] Matteo Burzoni, Vincenzo Ignazio, A. Max Reppen, and H. M. Soner. Viscosity solutions for controlled McKean-Vlasov jump-diffusions. *SIAM J. Control Optim.*, 58(3):1676–1699, 2020.
- [BLPR17] Rainer Buckdahn, Juan Li, Shige Peng, and Catherine Rainer. Mean-field stochastic differential equations and associated PDEs. *The Annals of Probability*, 45(2):824 – 878, 2017.
- [BS13] Gerard Brunick and Steven Shreve. Mimicking an Itô process by a solution of a stochastic differential equation. *Annals of Applied Probability*, 23(4):1584–1628, 2013.
- [Cal18] J. Calder. Lecture notes on viscosity solutions. University of Minnesota., 2018.
- [CCDar] Jean-François Chassagneux, Dan Crisan, and François Delarue. A probabilistic approach to classical solutions of the master equation for large population equilibria. *Memoirs of the AMS*, To appear.
- [CD18a] René Carmona and François Delarue. *Probabilistic theory of mean field games with applications. II*, volume 84 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2018. Mean field games with common noise and master equations.
- [CD18b] René Carmona and François Delarue. *Probabilistic Theory of Mean Field Games with Applications I : Mean Field FBSDEs, Control, and Games*. Springer, 2018.
- [CD22] Alekos Cecchin and François Delarue. Weak solutions to the master equation of potential mean field games. *arXiv*, 2204.04315, 2022.
- [CDJS23] Pierre Cardaliaguet, Samuel Daudin, Joe Jackson, and Panagiotis Souganidis. An algebraic convergence rate for the optimal control of mckean-vlasov dynamics. *arXiv*, 2203.14554, 2023.
- [CDLL19] Pierre Cardaliaguet, François Delarue, Jean-Michel Lasry, and Pierre-Louis Lions. *The master equation and the convergence problem in mean field games*, volume 201 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2019.

- [Cec21] Alekos Cecchin. Finite state N -agent and mean field control problems. *ESAIM Control Optim. Calc. Var.*, 27:Paper No. 31, 33, 2021.
- [CF21] Paul-Eric Chaudru de Raynal and Noufel Frikha. From the backward kolmogorov pde on the wasserstein space to propagation of chaos for mckean-vlasov sdes. *Journal de Mathématiques Pures et Appliquées*, 156:1–124, 2021.
- [CF22] Paul-Eric Chaudru de Raynal and Noufel Frikha. Well-posedness for some non-linear sdes and related pde on the wasserstein space. *Journal de Mathématiques Pures et Appliquées*, 159:1–167, 2022.
- [CGK⁺21] Andrea Cosso, Fausto Gozzi, Idris Kharroubi, Huyên Pham, and Mauro Rosestolato. Master bellman equation in the wasserstein space: Uniqueness of viscosity solutions. *arXiv*, <https://arxiv.org/abs/2107.10535>, 2021.
- [CKT21] Giovanni Conforti, Richard Kraaij, and Daniela Tonon. Hamilton–jacobi equations for controlled gradient flows: the comparison principle. *arXiv*, <https://arxiv.org/abs/2111.13258>, 2021.
- [CL84] M. G. Crandall and P.-L. Lions. Two approximations of solutions of Hamilton-Jacobi equations. *Math. Comp.*, 43(167):1–19, 1984.
- [CLOS22] Giulia Cavagnari, Stefano Lisini, Carlo Orrieri, and Giuseppe Savaré. Lagrangian, Eulerian and Kantorovich formulations of multi-agent optimal control problems: equivalence and gamma-convergence. *J. Differential Equations*, 322:268–364, 2022.
- [CP21] Pierre Cardaliaguet and Alessio Porretta. An introduction to mean field game theory. In *Mean Field Games, chapter 1, Cetraro, Italy 2019, Cardaliaguet, Pierre, Porretta, Alessio (Eds.)*, LNM 2281, pages 203–248. Springer, 2021.
- [CS22] Pierre Cardaliaguet and Panagiotis Souganidis. Regularity of the value function and quantitative propagation of chaos for mean field control problems. *arXiv*, 2204.01314, 2022.
- [CST22] Jean-François Chassagneux, Lukasz Szpruch, and Alvin Tse. Weak quantitative propagation of chaos via differential calculus on the space of measures. *The Annals of Applied Probability*, 32(3):1929 – 1969, 2022.
- [Dje22] Mao Fabrice Djete. Extended mean field control problem: a propagation of chaos result, 2022.
- [DPT22] Mao Fabrice Djete, Dylan Possamai, and Xiaolu Tan. Mckean-Vlasov optimal control: Limit theory and equivalence between different formulations. *Mathematics of Operations Research*, 2022.
- [DSS13] Steffen Dereich, Michael Scheutzow, and Reik Schottstedt. Constructive quantization: approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(4):1183–1203, 2013.
- [DT21] François Delarue and Alvin Tse. Uniform in time weak propagation of chaos on the torus. *arXiv*, <https://arxiv.org/abs/2104.14973>, 2021.
- [Eva10] Lawrence C. Evans. Adjoint and compensated compactness methods for Hamilton-Jacobi PDE. *Arch. Ration. Mech. Anal.*, 197(3):1053–1088, 2010.
- [FG15] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields*, 162(3-4):707–738, 2015.
- [Fis17] Markus Fischer. On the connection between symmetric N -player games and mean field games. *Ann. Appl. Probab.*, 127(2):757–810, 2017.
- [Fle71] Wendell H. Fleming. Stochastic control for small noise intensities. *SIAM J. Control*, 9:473–517, 1971.
- [FLOS19] M. Fornasier, S. Lisini, C. Orrieri, and G. Savaré. Mean-field optimal control as gamma-limit of finite agent controls. *European J. Appl. Math.*, 30(6):1153–1186, 2019.
- [FM97] Begoña Fernandez and Sylvie Méléard. A Hilbertian approach for fluctuations on the McKean-Vlasov model. *Stochastic Process. Appl.*, 71(1):33–53, 1997.
- [FS86] W. H. Fleming and P. E. Souganidis. Asymptotic series and the method of vanishing viscosity. *Indiana Univ. Math. J.*, 35(2):425–447, 1986.
- [GMS21] Wilfrid Gangbo, Sergio Mayorga, and Andrzej Świech. Finite dimensional approximations of Hamilton-Jacobi-Bellman equations in spaces of probability measures. *SIAM J. Math. Anal.*, 53(2):1320–1356, 2021.
- [GPW22] Maximilien Germain, Huyên Pham, and Xavier Warin. Rate of convergence for particle approximation of pdes in wasserstein space. *Journal of Applied Probability*, 59(4):992–1008, 2022.
- [GS14] Diogo A. Gomes and João Saúde. Mean field games models—a brief survey. *Dyn. Games Appl.*, 4(2):110–154, 2014.
- [GS15] Wilfrid Gangbo and Andrzej Świech. Existence of a solution to an equation arising from the theory of mean field games. *J. Differential Equations*, 259(11):6573–6643, 2015.
- [HMC06] Minyi Huang, Roland P. Malhamé, and Peter E. Caines. Large population stochastic dynamic games: Closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Commun. Inf. Syst.*, 6(3):221–251, 2006.
- [HPM03] Minyi Huang, Caines PE, and R.P. Malhame. Individual and mass behaviour in large population stochastic wireless power control problems: centralized and nash equilibrium solutions, 2003.
- [JM98] B. Jourdain and S. Méléard. Propagation of chaos and fluctuations for a moderate model with smooth initial data. *Annales de l’Institut Henri Poincaré (B) Probability and Statistics*, 34:727–766, 12 1998.

- [Kol12] V. Kolokoltsov. Nonlinear markov games on a finite state space (mean-field and binary interactions). *Int. J. Stat. Prob.*, 1:77–91, 2012.
- [Kry80] Nicolai V. Krylov. *Controlled Diffusion Processes*. Springer-Verlag Berlin Heidelberg, 1980.
- [Lac17] Daniel Lacker. Limit theory for controlled mckean-vlasov dynamics. *SIAM J. Control Optim.*, 55:1641–1672, 2017.
- [Lac20] Daniel Lacker. On the convergence of closed-loop Nash equilibria to the mean field game limit. *Ann. Appl. Probab.*, 30(4):1693–1761, 2020.
- [Lioa] Pierre-Louis Lions. Cours au collège de france, equations aux dérivées partielles et applications. <https://www.college-de-france.fr/site/pierre-louis-lions/course-2010-2011.htm>, 2010-11.
- [Liob] Pierre-Louis Lions. Estimées nouvelles pour les équations quasilinéaires. Seminar in Applied Mathematics at the Collège de France. <http://www.college-de-france.fr/site/pierre-louis-lions/seminar-2014-11-14-11h15.htm>, 2014.
- [Lio82] Pierre-Louis Lions. *Generalized solutions of Hamilton-Jacobi equations*, volume 69 of *Research Notes in Mathematics*. Pitman (Advanced Publishing Program), Boston, Mass.-London, 1982.
- [LL86] Jean-Michel Lasry and Pierre-Louis Lions. A remark on regularization in Hilbert spaces. *Israel Journal of Mathematics*, 55(3):257–266, 1986.
- [LL06a] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. I. Le cas stationnaire. *C. R. Math. Acad. Sci. Paris*, 343(9):619–625, 2006.
- [LL06b] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. II. Horizon fini et contrôle optimal. *C. R. Math. Acad. Sci. Paris*, 343(10):679–684, 2006.
- [LL07] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Jpn. J. Math.*, 2(1):229–260, 2007.
- [LSZ20] Daniel Lacker, Mykhaylo Shkolnikov, and Jiacheng Zhang. Superposition and mimicking theorems for conditional McKean-Vlasov equations. *arXiv*, <https://arxiv.org/abs/2004.00099>, 2020.
- [Mél96] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, volume 1627 of *Lecture Notes in Math.*, pages 42–95. Springer, 1996.
- [SY22] H. Mete Soner and Qinxin Yan. Viscosity solutions for mckean-vlasov control on a torus. *arXiv*, 2212.11053, 2022.
- [Szn85] A.S. Sznitman. A fluctuation result for nonlinear diffusions. *Infinite Dimensional Analysis and Stochastic Processes*, pages 145–160, 1985.
- [TH81] H. Tanaka and M. Hitsuda. Central limit theorem for a simple diffusion model of interacting particles. *Hiroshima Mathematical Journal*, 11(2):415–423, 1981.
- [Tse21] Alvin Tse. Higher order regularity of nonlinear Fokker-Planck PDEs with respect to the measure component. *Journal de Mathématiques Pures et Appliquées*, 150:134–180, 2021.
- [WZ20] Cong Wu and Jianfeng Zhang. Viscosity solutions to parabolic master equations and McKean-Vlasov SDEs with closed-loop controls. *Ann. Appl. Probab.*, 30(2):936–986, 2020.
- [YGD07] Haiping Ye, Jianming Gao, and Yongsheng Ding. A generalized Gronwall inequality and its application to a fractional differential equation. *J. Math. Anal. Appl.*, 328(2):1075–1081, 2007.

(S. DAUDIN) UNIVERSITÉ CÔTE D’AZUR, CNRS, LABORATOIRE J.A. DIEUDONNÉ, 06108 NICE, FRANCE

Email address: SAMUEL.DAUDIN@UNIV-COTEDAZUR.FR

(F. DELARUE) UNIVERSITÉ CÔTE D’AZUR, CNRS, LABORATOIRE J.A. DIEUDONNÉ, 06108 NICE, FRANCE

Email address: FRANCOIS.DELARUE@UNIV-COTEDAZUR.FR

(J. JACKSON) DEPARTMENT OF MATHEMATICS, THE UNIVERSITY OF TEXAS AT AUSTIN, UNITED STATES

Email address: JJACKS01@UTEXAS.EDU