CONVERGENCE RATE OF PARTICLE SYSTEM FOR SECOND-ORDER PDES ON WASSERSTEIN SPACE *

ERHAN BAYRAKTAR†, IBRAHIM EKREN‡, AND XIN ZHANG §

Abstract. In this paper, we provide a convergence rate for particle approximations of a class of second-order PDEs on Wasserstein space. We show that, up to some error term, the infinite-dimensional inf(sup)-convolution of the finite-dimensional value function yields a super (sub)-viscosity solution to the PDEs on Wasserstein space. Hence, we obtain a convergence rate using a comparison principle of such PDEs on Wasserstein space. Our argument is purely analytic and relies on the regularity of value functions established in [20].

Key words. Wasserstein space, second-order PDEs, viscosity solutions, comparison principle

MSC codes. 49L25, 60H30, 93E20

1. Introduction.

In this paper, we consider particle approximations of the mean field PDE on the d-dimensional Torus $\mathbb{T}^d$

$$\begin{cases} -\partial_t v(t,\mu) = \int_{\mathbb{T}^d} H(x, D_\mu v(t,\mu,x),\mu)\,\mu(dx) + \int_{\mathbb{T}^d} \operatorname{tr}(D^2_{x\mu} v)(t,\mu,x)\,\mu(dx) \\ \qquad\qquad\qquad + a\operatorname{tr}(\mathcal{H}v)(t,\mu), \\ \quad v(T,\mu) = G(\mu), \end{cases} \tag{1.1}$$

where a is a nonnegative constant, $\mathcal{H}v(t,\mu)$ is the partial Hessian defined as in [4, 20]

$$\mathcal{H}v(t,\mu) := \int_{\mathbb{T}^d} D^2_{x\mu} v(t,\mu,x)\,\mu(dx) + \int_{\mathbb{T}^d}\int_{\mathbb{T}^d} D^2_{\mu\mu} v(t,\mu,x,y)\,\mu(dx)\mu(dy),$$

and H, G are respectively the Hamiltonian and the terminal condition

$$H = H(x,p,\mu) : \mathbb{T}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{T}^d) \to \mathbb{R}, \quad G = G(\mu) : \mathcal{P}(\mathbb{T}^d) \to \mathbb{R}.$$

Denote $\boldsymbol{x} = (x^1, \dots, x^N) \in \mathbb{T}^{dN}$ and its empirical measure by $\mu^{\boldsymbol{x}} := \frac{1}{N}\sum_{i=1}^N \delta_{x^i}$. The particle approximation of (1.1) is provided by

$$\begin{cases} -\partial_t v^N(t,\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^N H(x^i, N D_{x^i} v^N(t,\boldsymbol{x}), \mu^{\boldsymbol{x}}) + \sum_{i=1}^N \Delta_{x^i} v^N(t,\boldsymbol{x}) \\ \qquad\qquad\qquad + a\sum_{i,j=1}^N \operatorname{tr}(D^2_{x^i x^j} v^N)(t,\boldsymbol{x}), \\ \quad v^N(T,\boldsymbol{x}) = G(\mu^{\boldsymbol{x}}) \end{cases} \tag{1.2}$$

see [20, 25] and the references therein. We provide a convergence rate of $v^N \to v$ based on the comparison principle and the regularity results obtained in [20].

It has been observed in [4, 20] that the partial Hessian term is equal to the second order derivative of v in the barycenter of measures. More precisely, taking

$$V(t,z,\mu) := v(t, (I_d + z)_\# \mu), \quad (t,z,\mu) \in [0,T] \times \mathbb{T}^d \times \mathcal{P}(\mathbb{T}^d),$$

*Submitted to the editors DATE.

Funding: E. Bayraktar is partially supported by the National Science Foundation and by the Susan M. Smith chair. I. Ekren is supported in part by NSF Grant DMS 2406240.

†Department of Mathematics, University of Michigan erhan@umich.edu

‡Department of Mathematics, University of Michigan iekren@umich.edu

§Department of Finance and Risk Engineering, New York University xz1662@nyu.edu

we have that $\nabla_z^2 V(t,z,\mu) = \mathcal{H}v(t,(I_d+z)_\#\mu)$. We say v is a viscosity solution to (1.1) if V is a viscosity solution to

$$\begin{cases} -\partial_t V(t,z,\mu) = \int_{\mathbb{T}^d} H^e(x,z,D_\mu V(t,z,\mu,x),\mu)\,\mu(dx) \\ \qquad\qquad + \int_{\mathbb{T}^d} \operatorname{tr}(D^2_{x\mu}V)(t,z,\mu,x)\,\mu(dx) + a\Delta_z V(t,z,\mu), \\ V(T,z,\mu) = G^e(z,\mu), \end{cases} \tag{1.3}$$

where

$$H^e(x,z,p,\mu) := H(x+z,p,(I_d+z)_\#\mu), \quad G^e(z,\mu) := G((I_d+z)_\#\mu).$$

Similarly, define $V^N(t,z,\boldsymbol{x}) := v^N(t,z+\boldsymbol{x})$ where $z+\boldsymbol{x} := (z+x^1,\dots,z+x^N)$. Denoting the Fourrier-Wasserstein distance in [4, 20, 32] by ρ_*, we show that the inf-convolution of V^N with respect to ρ_* given by

$$\overline{V}^{N,\epsilon}(t,z,\mu) := \inf_{(s,w,\boldsymbol{x})\in[0,T]\times\mathbb{T}^{d+dN}} \left(V^N(s,w,\boldsymbol{x}) + \frac{1}{2\epsilon}|t-s|^2 + \frac{1}{2\epsilon}|z-w|^2 + \frac{1}{2\epsilon}\rho_*^2(\mu^{\boldsymbol{x}},\mu) \right).$$

is a viscosity supersolution to (1.3) up to some error term $E(\epsilon,N)$, and therefore by comparison for all $(t,z,\boldsymbol{x}) \in [0,T]\times\mathbb{T}^{d+dN}$ we have

$$V^N(t,z,\boldsymbol{x}) + E(\epsilon,N) \geq \overline{V}^{N,\epsilon}(t,z,\mu^{\boldsymbol{x}}) + E(\epsilon,N) \geq V(t,z,\mu^{\boldsymbol{x}}).$$

By the same token, we obtain the other direction

$$V^N(t,z,\boldsymbol{x}) - E(\epsilon,N) \leq V(t,z,\mu^{\boldsymbol{x}}), \quad (t,z,\boldsymbol{x}) \in [0,T]\times\mathbb{T}^{d+dN},$$

and hence

$$|V^N(t,z,\boldsymbol{x}) - V(t,z,\mu^{\boldsymbol{x}})| \leq E(\epsilon,N).$$

Then choosing ϵ properly as a function of N, one obtains that $E(\epsilon,N) \approx C\alpha^{1/3}(N)$ where C is a positive constant independent of N and

$$\alpha(N) = \begin{cases} N^{-1/2}, & \text{if } d=1, \\ N^{-1/2}\log(N), & \text{if } d=2, \\ N^{-1/d}, & \text{if } d>2. \end{cases}$$

Let us denote by $B_R \subset \mathbb{R}^d$ the closed ball at the origin with radius R and $k_* = \lfloor d/2 \rfloor + 3$. Using the definitions of functional spaces in [20], we make the following assumptions on the coefficients H, G.

Assumption 1.1. *(i) G is k_*-times continuously differentiable and Lipschitz with respect to C^{-k_*};*

(ii) H is k_*-times continuously differentiable in all variables, and satisfies the regularity condition, with $C_H > 0$

$$|H(x,p,\mu) - H(x',p',\mu')| \leq C_H(1+|p|+|p'|)(|x-x'|+|p-p'|+W_1(\mu,\mu'))$$

for any $x,x' \in \mathbb{T}^d$, $p,p' \in \mathbb{R}^d$, $\mu,\mu' \in \mathcal{P}(\mathbb{T}^d)$. For each $R>0$, there is a constant C_R such that for each $(x,p) \in \mathbb{T}^d \times B_R$

$$|H(x,p,\mu) - H(x,p,\mu')| \leq C_R\|\mu-\mu'\|_{C^{-k_*}},$$

and for each $\mu \in \mathcal{P}(\mathbb{T}^d)$,

$$\sup_{\mu\in\mathcal{P}(\mathbb{T}^d)} \|H(\cdot,\cdot,\mu)\|_{C^{k_*}(\mathbb{T}^d\times B_R)} \leq C_R.$$

THEOREM 1.1. *Under Assumption 1.1, v^N converges to v with the rate $\alpha^{1/3}(N)$*

$$\sup_{(t,\boldsymbol{x})\in[0,T]\times\mathbb{T}^{dN}} \left|v^N(t,\boldsymbol{x}) - v(t,\mu^{\boldsymbol{x}})\right| \leq C\alpha^{1/3}(N),$$

where C is a constant independent of N.

Our result provides the first convergence rate for particle approximations of partially second-order PDEs on the Wasserstein space. Such equations arise in mean-field control problems when the control of common noise is state-independent, as well as in stochastic control with partial observation; see, e.g., [4]. In the former case, the particle approximation justifies the mean-field formulation of controlling a large population of agents; see, e.g., [12, 14] and references therein.

Without the common noise, i.e. setting $a = 0$ in (1.1), [19] obtains the convergence rate of $1/\sqrt{N}$ under some convexity assumptions. The argument relies on the fact that super-convolution of semi-concave functions provides desired regularity. While [19] shows better convergence rate, our argument is purely analytic and doesn't rely on the semi-concavity/convexity of H. We would like to mention that the argument of our result makes use of the regularity of v^N established in [20] where only the convergence

$$\sup_{(t,\boldsymbol{x})\in[0,T]\times\mathbb{T}^{dN}} \left|v^N(t,\boldsymbol{x}) - v(t,\mu^{\boldsymbol{x}})\right| \to 0$$

is provided but not its convergence rate. Moreover, using the same method of the infinite dimensional inf/sup-convolution and the comparison result from [4], one should also be able to show the convergence rate for the same type of equations on $\mathbb{R}^d$. But to illustrate the main idea of the argument, we choose to work on $\mathbb{T}^d$ to avoid the technical issue of non-compactness.

The remainder of the paper is organized as follows. We will discuss some related literature in Section 1.1, and introduce notations in Section 1.2. In Section 2, we will present the definition of viscosity of solution and some preliminary results. The main result will be proved in Section 3.

1.1. Related Literature. PDEs on Wasserstein space appear in mean field games and McKean-Vlasov control problems [10, 14, 13, 24, 3, 16], and also in filtering problems [28, 26, 1, 4, 5]. Various notions of differentiability for functions on Wasserstein space have been defined, and in this paper we adopt the one introduced by Lions in [10]. It is stronger than the geometric definition of differentiability, and allows a version of Itô's formula which is crucial for control problems.

The comparison principle of PDEs on Wasserstein space has attracted lots of attention. Viscosity solutions of first-order PDEs on the Wasserstein space have been studied in [35, 32, 31, 8, 2, 16, 6, 21]. It is worth noting that the Fourier-Wasserstein metric ρ_* was first used in the study of viscosity solution by [31]. The comparison principle of partially second-order equations, in which the second-order derivative in measure appears in the form of partial Hessian, have been studied in [4, 20, 23, 6]. Let us mention that [23, 6] adopted different notions of differentiability on Wasserstein space. Fully second-order PDEs on the Wasserstein space are related to measure-valued martingale optimization problems. [17] proves a uniqueness result for equations that are exact limit of finite dimensional approximations. PDEs on the Wasserstein space also appear in mean-field optimal stopping problems [35, 34, 29].

Convergence of particle system in mean field control problems were studied in [36, 33, 2, 15] based on viscosity theory, while [27] provided a probabilistic argument. The convergence rate for first-order PDEs on Wasserstein space was obtained in [11,

19, 9]. Assuming the existence of smooth solution to mean-field PDEs, [25] got the optimal convergence rate by a verification argument.

1.2. Notations. Define $\mathbb{T}^d = \mathbb{R}^d/(2\pi\mathbb{Z})^d$, and take Fourier basis

$$e_l(x) := (2\pi)^{-d/2} e^{il\cdot x}, \quad x \in \mathbb{T}^d,\, k \in \mathbb{Z}^d.$$

For any complex number $z \in \mathbb{C}$, we denote its complex conjugate by z^*. For any $f \in L^2(\mathbb{T}^d)$, we define $F_l(f) := \int_{\mathbb{T}^d} f(x) e_l^*(x)\, dx$. For any $k \in \mathbb{N}$, we define,

$$\|f\|_k^2 := \sum_{l\in\mathbb{Z}^d} (1+|l|^2)^k |F_l(f)|^2$$

and the Sobolev space

$$\mathbb{H}^k(\mathbb{T}^d) = \left\{ f \in L^2(\mathbb{T}^d) : \|f\|_k < \infty \right\}.$$

The space of k-th continuously differentiable function is denoted by $C^k(\mathbb{T}^d)$ with the norm defined as

$$\|f\|_{C^k} = \sum_{j\in\mathbb{N}^d : |j|\le k} \|D^j f\|_{L^\infty}.$$

For any signed Borel measure η on $\mathbb{T}^d$, we define

$$\|\eta\|_{-k} = \sup_{\|f\|_k \le 1} \int_{\mathbb{T}^d} f(x)\, \eta(dx),$$

$$F_l(\eta) = \int_{\mathbb{T}^d} e_l(x)\, d\eta, \quad l \in \mathbb{Z}^d,$$

and also

$$\|\eta\|_{C^{-k}} = \sup_{\|f\|_{C^k} \le 1} \int_{\mathbb{T}^d} f(x)\, \eta(dx).$$

Then for any $\mu, \nu \in \mathcal{P}(\mathbb{T}^d)$, we define $\rho_{-k}(\mu,\nu) = \|\mu - \nu\|_{-k}$. Throughout the paper, we take $k_* = \lfloor d/2 \rfloor + 3$ and denote $\rho_* = \rho_{-k_*}$.

Throughout the paper, we adopt Lions differentiability for functions defined on $\mathcal{P}(\mathbb{T}^d)$; see e.g. [14, Chapter 5].

2. Definition and preliminaries.

2.1. Definition of viscosity solution. First we introduce the notion of viscosity solution from [20]. Suppose $v : [0,T] \times \mathcal{P}(\mathbb{T}^d) \to \mathbb{R}$ is a smooth solution to (1.1). Define $V(t,z,\mu) : [0,T] \times \mathbb{T}^d \times \mathcal{P}(\mathbb{T}^d) \to \mathbb{R}$ via

$$V(t,z,\mu) := v(t, (I_d + z)_\# \mu). \tag{2.1}$$

It can be easily verified that

$$D_\mu V(t,z,\mu)(x) = D_\mu v(t, (I_d+z)_\#\mu)(x+z),$$
$$D^2_{x\mu} V(t,z,\mu)(x) = D^2_{x\mu} v(t, (I_d+z)_\#\mu)(x+z),$$
$$\Delta_z V(t,z,\mu) = \mathrm{tr}(\mathcal{H}v)(t, (I_d+z)_\#\mu),$$

and V satisfies (1.3).

We say v is a viscosity solution of (1.1) if V is a viscosity solution of (1.3). More precisely, we have the set of test functions.

DEFINITION 2.1. *Denote by* $C^{1,2,2}_{\mathfrak{p}}([0,T]\times\mathbb{T}^d\times\mathcal{P}(\mathbb{T}^d))$ *the set of continuous functions* $\Phi(t,z,\mu):[0,T]\times\mathbb{T}^d\times\mathcal{P}(\mathbb{T}^d)\to\mathbb{R}$ *such that the derivatives*

$$(\partial_t\Phi, D_z\Phi, D^2_{zz}\Phi)(t,z,\mu):[0,T]\times\mathbb{T}^d\times\mathcal{P}(\mathbb{T}^d)\to\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^{d\times d}$$

as well as

$$(D_\mu\Phi, D^2_{x\mu}\Phi)(t,z,\mu,x):[0,T]\times\mathbb{T}^d\times\mathcal{P}(\mathbb{T}^d)\times\mathbb{T}^d\to\mathbb{R}^d\times\mathbb{R}^{d\times d}$$

exist and are continuous.

DEFINITION 2.2. *An upper semi-continuous function* $v:[0,T]\times\mathcal{P}(\mathbb{T}^d)\to\mathbb{R}$ *is called a viscosity subsolution to* (1.1) *if its extension* V *via* (2.1) *is a viscosity subsolution to* (1.3)*, i.e.,* $V(T,z,\mu)\geq G^e(z,\mu)$*, and for any* $\Phi\in C^{1,2,2}_{\mathfrak{p}}([0,T]\times\mathbb{T}^d\times\mathcal{P}(\mathbb{T}^d))$ *such that* $V-\Phi$ *obtains a local maximum at* $(t_0,z_0,\mu_0)\in[0,T)\times\mathbb{T}^d\times\mathcal{P}(\mathbb{T}^d)$*, we have*

$$\begin{aligned}-\partial_t\Phi(t_0,z_0,\mu_0)\leq&\int_{\mathbb{T}^d}H^e(x,z_0,D_\mu\Phi(t_0,z_0,\mu_0,x),\mu)\,\mu_0(dx)\\&+\int_{\mathbb{T}^d}\mathrm{tr}(D^2_{x\mu}\Phi)(t_0,z_0,\mu_0,x)\,\mu_0(dx)+a\Delta_z\Phi(t_0,z_0,\mu_0),\end{aligned}$$

Similarly, we define viscosity supersolution. A continuous function

$$v:[0,T]\times\mathcal{P}(\mathbb{T}^d)\to\mathbb{R}$$

is called a viscosity solution if its extension V *via* (2.1) *is a viscosity subsolution and supersolution to* (1.3) *at the same time.*

2.2. Preliminary results. As the HJB equation (1.2) of particle system is uniformly elliptic, one can show the regularity of solution. The following result is from [20, Lemma 3.1, Theorem 3.2].

LEMMA 2.3. *Under Assumption 1.1, there exists a unique classical solution* v^N *to* (1.2) *and positive constants* C *independent of* N *such that for any* $N\in\mathbb{N}$*,* $i\in\{1,2,\dots,N\}$*,* $k\leq k_*$*,*

$$\left|D^k_{x^i}v^N(t,\boldsymbol{x})\right|\leq\frac{C}{N},\quad\forall\,(t,\boldsymbol{x})\in[0,T]\times\mathbb{T}^{dN}.$$

In addition, for all $0\leq s<t\leq T$ *and* $\boldsymbol{x},\boldsymbol{y}\in\mathbb{T}^{dN}$*, we have*

$$\left|v^N(t,\boldsymbol{x})-v^N(s,\boldsymbol{y})\right|\leq C\left(\sqrt{t-s}+W_1(\mu^{\boldsymbol{x}},\mu^{\boldsymbol{y}})\right).$$

Let us provide an approximation of v^N that has been used in [20, 16]. Given any $\mu\in\mathcal{P}(\mathbb{T}^d)$, denote by $\mu^{\otimes N}$ the N fold product of probability measure μ. Let us define $\hat{v}^N:[0,T]\times\mathcal{P}(\mathbb{T}^d)$ via

$$\hat{v}^N(t,\mu)=\int_{\mathbb{T}^{dN}}v^N(t,\boldsymbol{y})\,\mu^{\otimes N}(d\boldsymbol{y}).$$

LEMMA 2.4. *Under Assumption 1.1, the following inequality holds for any* $\boldsymbol{x},\boldsymbol{y}\in\mathbb{T}^{dN}$ *with a positive constant* C *independent of* N*,*

$$\left|\hat{v}^N(t,\mu)-\hat{v}^N(t,\nu)\right|\leq C\rho_*(\mu,\nu).$$

Proof. It can be verified that $\hat{v}^N$ is linearly differentiable, and thanks to the symmetry its derivative is given by

$$\frac{\delta \hat{v}^N}{\delta \mu}(t,\mu,x) = \sum_{i=1}^{N} \int_{\mathbb{T}^{d(N-1)}} v^N(t, y^1, \dots, y^{i-1}, x, y^i, \dots, y^{N-1})\, \mu^{\otimes(N-1)}(dy^1, \dots, dy^{N-1})$$

In light of Lemma 2.3, we get an estimate of the $\mathbb{H}^{k_*}$ norm

$$\sup_{(t,\mu)\in[0,T]\times\mathbb{T}^d} \|\frac{\delta \hat{v}^N}{\delta \mu}(t,\mu,\cdot)\|_{k_*} \leq \sum_{k=1,\dots,k_*} \sum_{i=1}^{N} \|D^k_{x^i} v^N\|_\infty \leq C.$$

Therefore, by the definition of linear derivative, one immediately concludes that

$$\left|\hat{v}^N(t,\mu) - \hat{v}^N(t,\nu)\right| \leq C\|\mu-\nu\|_{-k_*} = C\rho_*(\mu,\nu). \qquad \square$$

The next proposition shows that $\hat{v}^N$ is indeed close to v^N and thus almost Lipschitz.

PROPOSITION 2.5. *For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{T}^{dN}$, the following estimates of the value function holds*

$$|v^N(t,\boldsymbol{x}) - \hat{v}^N(t,\mu^{\boldsymbol{x}})| \leq C\alpha(N)$$

where C is a positive constant independent of N. Together with Lemma 2.4, immediately we obtain that

$$|v^N(t,\boldsymbol{x}) - v^N(t,\boldsymbol{y})| \leq C\left(\rho_*(\mu^{\boldsymbol{x}},\mu^{\boldsymbol{y}}) + \alpha(N)\right).$$

Proof. Let us prove $\left|v^N(t,\boldsymbol{x}) - \hat{v}^N(t,\mu^{\boldsymbol{x}})\right| \leq C\alpha(N)$. Indeed, according to the definition of $\hat{v}^N$ and Lemma 2.3

$$\begin{aligned}\left|v^N(t,\boldsymbol{x}) - \hat{v}^N(t,\mu^{\boldsymbol{x}})\right| &\leq \int_{\mathbb{R}^{d\times N}} |v(t,\boldsymbol{x}) - v(t,\boldsymbol{y})|\, (\mu^{\boldsymbol{x}})^{\otimes N}(d\boldsymbol{y}) \\ &\leq C\int_{\mathbb{R}^{d\times N}} W_1(\mu^{\boldsymbol{x}},\mu^{\boldsymbol{y}})\, (\mu^{\boldsymbol{x}})^{\otimes N}(d\boldsymbol{y}) = C\mathbb{E}\left[W_1(\mu^{\boldsymbol{x}},\hat{\mu}^{\boldsymbol{x}})\right],\end{aligned}$$

where $\hat{\mu}^{\boldsymbol{x}}$ denotes the empirical measure of $\mu^{\boldsymbol{x}}$ with N samples. According to [22], it is bounded from above by $\alpha(N)$. $\square$

By Arzelà–Ascoli theorem, Lemma 2.4 implies there exists a subsequence of $\hat{v}^N$ uniformly converging to a function v, which is also ρ_*-Lipschitz. Moreover according to Proposition 2.5, v is also a limiting point of v^N, and hence is a viscosity solution to (1.1) by a standard argument. Finally, the uniqueness of viscosity solution has been proved in [20] adapting the techniques in [4].

PROPOSITION 2.6. *Under Assumption 1.1 for* (1.1)*, any upper semi-continuous subsolution to is smaller than lower semi-continuous supersolution. Then there is a unique viscosity solution v to* (1.1) *on $\mathcal{P}(\mathbb{T}^d)$.*

At the end of section, we provide a simple lemma quantifying the denseness of empirical measures in $\mathcal{P}(\mathbb{T}^d)$.

LEMMA 2.7. *For any $\mu \in \mathcal{P}(\mathbb{T}^d)$, we have*

$$\inf_{\boldsymbol{x}\in\mathbb{T}^{dN}} \rho_*(\mu^{\boldsymbol{x}},\mu) \leq C\alpha(N).$$

Proof. Thanks to Sobolev embedding theorem [7, Corollary 9.13], as $k_* = \lfloor d/2 \rfloor + 2$ any $f \in \mathbb{H}^{k_*}(\mathbb{T}^d)$ is Lipschitz with coefficient proportional to $\|f\|_{k_*}$. Therefore, it can be easily seen that

$$\rho_*(\mu, \nu) \leq C W_1(\mu, \nu),$$

with a universal constant C. Let us consider an *i.i.d.* sequence X_n with distribution μ, and define $\hat{\mu}^N = \frac{1}{N}\delta_{X_n}$ to be the empirical measure of μ. With the sample complexity of W_1, see e.g. [22], we deduce that

$$\mathbb{E}[\rho_*(\hat{\mu}^N, \mu)] \leq C\alpha(N),$$

and hence

$$\inf_{\boldsymbol{x} \in \mathbb{T}^{dN}} \rho_*(\mu^{\boldsymbol{x}}, \mu) \leq C\alpha(N). \qquad \square$$

3. Proof of Theorem 1.1. We only prove the inequality

$$V^N(t, z, \boldsymbol{x}) \geq V(t, z, \mu^{\boldsymbol{x}}) - C\alpha^{1/3}(N), \tag{3.1}$$

and the proof for the other direction

$$V^N(t, z, \boldsymbol{x}) \leq V(t, z, \mu^{\boldsymbol{x}}) + C\alpha^{1/3}(N)$$

is similar. The argument is based on the fact that the inf convolution preserves the property of being a supersolution, and will be divided into several steps.

For any $z \in \mathbb{T}^d, \boldsymbol{x} \in \mathbb{T}^{dN}$, denote $z + \boldsymbol{x} := (z + x^1, \dots, z + x^N)$. Recall that v^N denotes the classical solution to (1.2). Introducing

$$V^N(t, z, \boldsymbol{x}) := v^N(t, z + \boldsymbol{x}),$$

let us consider the inf-convolution of V^N

$$\overline{V}^{N,\epsilon}(t, z, \mu) := \inf_{(s,w,\boldsymbol{x}) \in [0,T] \times \mathbb{T}^{d+dN}} \left(V^N(s, w, \boldsymbol{x}) + \frac{1}{2\epsilon}|t - s|^2 + \frac{1}{2\epsilon}|z - w|^2 + \frac{1}{2\epsilon}\rho_*^2(\mu^{\boldsymbol{x}}, \mu) \right).$$

It is clear that $V^N(t, z, \boldsymbol{x}) \geq \overline{V}^{N,\epsilon}(t, z, \mu^{\boldsymbol{x}})$ for every $x \in \mathbb{T}^{dN}$. We prove that $\overline{V}^{N,\epsilon}$ is a viscosity supersolution to (1.1) up to some error.

Suppose $\Phi : [0, T] \times \mathbb{T}^d \times \mathcal{P}(\mathbb{T}^d) \to \mathbb{R}$ is a regular test function in $C^{1,2,2}_{\mathfrak{p}}([0, T] \times \mathbb{T}^d \times \mathcal{P}(\mathbb{T}^d))$ and we have a local strict minimum (t_0, z_0, μ_0) at $\overline{V}^{N,\epsilon} - \Phi$ with $t_0 < T$, and we aim to show that

$$\begin{aligned} -\partial_t \Phi(t_0, z_0, \mu_0) + E(N, \epsilon) \geq & \int_{\mathbb{T}^d} H^e(x, z_0, D_\mu \Phi(t_0, z_0, \mu_0, x), \mu_0)\, \mu_0(dx) \\ & + \int_{\mathbb{T}^d} \mathrm{tr}(D^2_{x\mu}\Phi)(t_0, z_0, \mu_0, x)\, \mu_0(dx) + a\Delta_z \Phi(t_0, z_0, \mu_0), \end{aligned} \tag{3.2}$$

where $E(N, \epsilon)$ is an error term to be determined later.

By compactness, take $(s_0, w_0, \boldsymbol{x}_0)$ such that

$$\overline{V}^{N,\epsilon}(t_0, z_0, \mu_0) = V^N(s_0, w_0, \boldsymbol{x}_0) + \frac{1}{2\epsilon}|s_0 - t_0|^2 + \frac{1}{2\epsilon}|w_0 - z_0|^2 + \frac{1}{2\epsilon}\rho_*^2(\mu^{\boldsymbol{x}_0}, \mu_0).$$

Then we have the following crucial inequality, for every $(s, w, \boldsymbol{x})$ and (t, z, μ),

$$
\begin{aligned}
(3.3)\quad & V^N(s,w,\boldsymbol{x}) + \frac{1}{2\epsilon}\left(|s-t|^2 + |w-z|^2 + \rho_*^2(\mu^{\boldsymbol{x}},\mu)\right) - \Phi(t,z,\mu) \\
&\geq \overline{V}^{N,\epsilon}(t,z,\mu) - \Phi(t,z,\mu) \\
&\geq \overline{V}^{N,\epsilon}(t_0,z_0,\mu_0) - \Phi(t_0,z_0,\mu_0) \\
&= V^N(s_0,w_0,\boldsymbol{x}_0) + \frac{1}{2\epsilon}\left(|s_0-t_0|^2 + |w_0-z_0|^2 + \rho_*^2(\mu^{\boldsymbol{x}_0},\mu_0)\right) - \Phi(t_0,z_0,\mu_0).
\end{aligned}
$$

To make use of the viscosity property of V^N, let us define a finite dimensional test function Φ^ϵ by super-convolution,

$$
(3.4)\quad \Phi^\epsilon(s,w,\boldsymbol{x}) := \sup_{z\in\mathbb{T}^d}\left\{\Phi(t_0,z,\mu_0) - \frac{1}{2\epsilon}|w-z|^2\right\} - \frac{1}{2\epsilon}|s-t_0|^2 - \frac{1}{2\epsilon}\rho_*^2(\mu^{\boldsymbol{x}},\mu_0).
$$

Then inequality (3.3) implies that

$$
\begin{aligned}
V^N(s,w,\boldsymbol{x}) - \Phi^\epsilon(s,w,\boldsymbol{x}) &\geq \overline{V}^{N,\epsilon}(t_0,z_0,\mu_0) - \Phi(t_0,z_0,\mu_0) \\
(3.5)\qquad &\geq V^N(t_0,w_0,\boldsymbol{x}_0) - \Phi^\epsilon(s_0,w_0,\boldsymbol{x}_0).
\end{aligned}
$$

Therefore, $V^N - \Phi^\epsilon$ obtains a local minimum at $(t_0, w_0, \boldsymbol{x}_0)$, and we are going to invoke the viscosity property of V^N. The next lemma provides derivatives of Φ^ϵ.

LEMMA 3.1. *We have the following equalities, for every $w \in \mathbb{T}^d$ and every $\boldsymbol{x} \in \mathbb{T}^{dN}$*

$$
\begin{aligned}
\partial_s\Phi^\epsilon(s_0,w,\boldsymbol{x}) &= \frac{1}{\epsilon}(t_0-s_0), \\
ND_{x^i}\Phi^\epsilon(s_0,w,\boldsymbol{x}) &= \frac{1}{2\epsilon}D_\nu\rho_*^2(\mu^{\boldsymbol{x}},\mu_0)(x^i), && i=1,\dots,N, \\
ND^2_{x^i}\Phi^\epsilon(s_0,w,\boldsymbol{x}) &= \frac{1}{2\epsilon}D^2_{x\nu}\rho_*^2(\mu^{\boldsymbol{x}},\mu_0)(x^i) + \frac{1}{2\epsilon N}D^2_\nu\rho_*^2(\mu^{\boldsymbol{x}_0},\mu_0)(x^i,x^i), && i=1,\dots,N,
\end{aligned}
$$

and at (t_0, z_0, μ_0)

$$
\begin{aligned}
\partial_t\Phi(t_0,z_0,\mu_0) &= \frac{1}{\epsilon}(t_0-s_0), \\
D_\mu\Phi(t_0,z_0,\mu_0)(x^i) &= \frac{1}{2\epsilon}D_\nu\rho_*^2(\mu^{\boldsymbol{x}_0},\mu_0)(x^i), && i=1,\dots,N, \\
D^2_{x\mu}\Phi(t_0,z_0,\mu_0)(x^i) &= \frac{1}{2\epsilon}D^2_{x\nu}\rho_*^2(\mu^{\boldsymbol{x}_0},\mu_0)(x^i), && i=1,\dots,N.
\end{aligned}
$$

Moreover, $w \mapsto \Phi^\epsilon(s_0, w, \boldsymbol{x}_0)$ is second order differentiable a.e. *At every $w \in \mathbb{T}^d$ where it is second order differentiable, we have*

$$
\Delta_w\Phi^\epsilon(s_0,w,\boldsymbol{x}_0) \geq \Delta_z\Phi(t_0,z(w),\mu_0),
$$

where

$$
z(w) \in \operatorname*{arg\,max}_{z\in\mathbb{T}^d}\left\{\Phi(t_0,z,\mu_0) - \frac{1}{2\epsilon}|w-z|^2\right\}.
$$

Proof. Taking derivative with respect to s in (3.4),

$$
\partial_s\Phi^\epsilon(s_0,w,\boldsymbol{x}) = \frac{1}{\epsilon}(t_0-s_0).
$$

Inequality (3.3) implies that t_0 is a maximizer of

$$\mu \mapsto \Phi(t, z_0, \mu) - \frac{1}{2\epsilon}|s_0 - t|^2,$$

and thus

$$\partial_t \Phi(t_0, z_0, \mu_0) = \frac{1}{\epsilon}(t_0 - s_0) = \partial_s \Phi^\epsilon(s_0, w, \boldsymbol{x}_0).$$

Taking derivative with respect to $\boldsymbol{x}$ in (3.4),

$$\begin{aligned} D_{x^i}\Phi^\epsilon(s_0, w, \boldsymbol{x}) &= -\frac{1}{2\epsilon N} D_\mu \rho_*^2(\mu^{\boldsymbol{x}}, \mu_0)(x^i) \\ &= \frac{1}{2\epsilon N} D_\nu \rho_*^2(\mu^{\boldsymbol{x}}, \mu_0)(x^i), \end{aligned}$$

where $D_\nu \rho_*^2$ denotes Lions' derivative of $\eta \mapsto \rho_*^2(\mu, \eta)$. Taking derivative once more,

$$D^2_{x^i}\Phi^\epsilon(s_0, w, \boldsymbol{x}) = \frac{1}{2\epsilon N} D^2_{x\nu}\rho_*^2(\mu^{\boldsymbol{x}}, \mu_0)(x^i) + \frac{1}{2\epsilon N^2} D^2_\nu \rho_*^2(\mu^{\boldsymbol{x}_0}, \mu_0)(x^i, x^i).$$

Again due to (3.3), μ_0 is a maximizer of

$$\mu \mapsto \Phi(t_0, z_0, \mu) - \frac{1}{2\epsilon}\rho_*^2(\mu^{\boldsymbol{x}_0}, \mu),$$

and first order condition yields

$$D_\mu \Phi(t_0, z_0, \mu_0)(x^i) = \frac{1}{2\epsilon} D_\nu \rho_*^2(\mu^{\boldsymbol{x}_0}, \mu)(x^i), \quad x^i \in \mathbb{T}^d,$$

and hence

$$D^2_{x\mu}\Phi(t_0, z_0, \mu_0)(x^i) = \frac{1}{2\epsilon} D^2_{x\nu}\rho_*^2(\mu^{\boldsymbol{x}_0}, \mu)(x^i), \quad x^i \in \mathbb{T}^d.$$

The last claim is a property of super-convolution; see e.g. [18, Lemma A.5], and hence we finish proving the lemma. ∎

Without loss of generality, we assume that $V^N - \Phi^\epsilon$ obtains a strict local minimal at $(s_0, w_0, \boldsymbol{x}_0)$. Denoting $\theta_0 = (t_0, z_0, \mu_0)$, a direct application of Jensen's lemma ([18, Lemma A.3]) and Lemma 3.1 shows that

$$\left(\partial_t \Phi(\theta_0), D_\mu \Phi(\theta_0)(x_0^i), D^2_{x\mu}\Phi(\theta_0)(x_0^i) + \frac{1}{2\epsilon N} D^2_\nu \rho_*^2(\mu^{\boldsymbol{x}_0}, \mu_0)(x_0^i, x_0^i), \Delta_z \Phi(\theta_0)\right).$$

approximate derivatives of Φ^ϵ at $(s_0, w_0, \boldsymbol{x}_0)$. We summarize this key result in the following proposition.

PROPOSITION 3.2. *There exists a sequence of $\mathbb{T}^d \ni w_n \to w_0$ and $\mathbb{R}^d \ni p_n \to 0$ such that*

$$V^N(t_0, w, \boldsymbol{x}_0) - \Phi^\epsilon(t_0, w, \boldsymbol{x}_0) + \langle p_n, w\rangle \quad \textit{obtains a local minimum at } w_n,$$

and $w \mapsto V^N(t_0, w, \boldsymbol{x}_0) - \Phi^\epsilon(t_0, w, \boldsymbol{x}_0)$ is second-order differentiable at w_n. Moreover, for all $i = 1, \dots, N$, the limit point of

$$\left(\partial_s \Phi^\epsilon, N D_{x^i}\Phi^\epsilon, N D^2_{x^i}\Phi^\epsilon, \Delta_w \Phi^\epsilon\right)(s_0, w_n, \boldsymbol{x}_0),$$

as $n \to \infty$, is equal to

$$\left(\partial_t \Phi(\theta_0), D_\mu \Phi(\theta_0)(x_0^i), D^2_{x\mu}\Phi(\theta_0)(x_0^i) + \frac{1}{2\epsilon N} D^2_\nu \rho_*^2(\mu^{\boldsymbol{x}_0}, \mu_0)(x_0^i, x_0^i), \Delta_z \Phi(\theta_0)\right).$$

As v^N is a classical solution to (1.2), it can be easily seen that V^N is a classical solution to the following equation

$$\begin{cases} -\partial_t V^N(t,w,\boldsymbol{x}) = \sum_{i=1}^N \frac{H^e(x^i,w,ND_{x^i}V^N(t,w,\boldsymbol{x}),\mu^{\boldsymbol{x}})}{N} + \sum_{i=1}^N \Delta_{x^i} V^N(t,w,\boldsymbol{x}) \\ \qquad\qquad +a\Delta_w V^N(t,w,\boldsymbol{x}), \\ V^N(T,w,\boldsymbol{x}) = G^e(w,\mu^{\boldsymbol{x}}), \end{cases}$$

and thus also a viscosity solution. Remember that $\theta_0 = (t_0, z_0, \mu_0)$ and $V^N - \Phi^\epsilon$ obtains a local minimal at $(s_0, w_0, \boldsymbol{x}_0)$ in (3.5). Setting $\epsilon = \alpha(N)$ in Lemma 3.3 below, it can be seen that $s_0 < T$ for large enough N. Therefore due to Proposition 3.2 and the definition of viscosity solution

$$\begin{aligned} -\partial_t \Phi(\theta_0) \geq & \frac{1}{N}\sum_{i=1}^N H^e(x_0^i, w_0, D_\mu\Phi(\theta_0)(x_0^i), \mu^{\boldsymbol{x}_0}) \\ & + \sum_{i=1}^N \left(\frac{1}{N} \operatorname{tr} D^2_{x\mu}\Phi(\theta_0, x_0^i) + \frac{1}{2\epsilon N^2} D^2_\nu \rho^2_*(\mu^{\boldsymbol{x}_0}, \mu_0)(x_0^i, x_0^i) \right) + a\Delta_z \Phi(\theta_0). \end{aligned}$$

Comparing the inequality above with (3.2), we define the error term $E(N,\epsilon)$

$$\begin{aligned} E(N,\epsilon) :=& \frac{1}{N}\sum_{i=1}^N H^e(x_0^i, z_0, D_\mu\Phi(\theta_0)(x_0^i), \mu^{\boldsymbol{x}_0}) - H^e(x_0^i, w_0, D_\mu\Phi(\theta_0)(x_0^i), \mu^{\boldsymbol{x}_0}) \\ & + \int H^e(x, z_0, D_\mu\Phi(\theta_0)(x), \mu^{\boldsymbol{x}_0})\, \mu_0(dx) - \sum_{i=1}^N \frac{H^e(x_0^i, z_0, D_\mu\Phi(\theta_0)(x_0^i), \mu^{\boldsymbol{x}_0})}{N} \\ & + \int (H^e(x, z_0, D_\mu\Phi(\theta_0)(x), \mu_0) - H^e(x, z_0, D_\mu\Phi(\theta_0)(x), \mu^{\boldsymbol{x}_0}))\, \mu_0(dx) \\ & + \int \operatorname{tr}(D^2_{x\mu}\Phi)(\theta_0, x)\, (\mu_0 - \mu^{\boldsymbol{x}_0})(dx) - \sum_{i=1}^N \frac{1}{2\epsilon N^2} D^2_\nu \rho^2_*(\mu^{\boldsymbol{x}_0}, \mu_0)(x_0^i, x_0^i). \end{aligned} \tag{3.6}$$

It can be easily checked that (3.2) is satisfied with such $E(N,\epsilon)$. We will provide an upper bound of (3.6). To this end, let us first estimate $|s_0 - t_0|$, $|z_0 - w_0|$ and $\rho_*(\mu^{\boldsymbol{x}_0}, \mu_0)$.

Lemma 3.3. *We have the following estimates,*

$$|t_0 - s_0| \leq C\epsilon^{2/3}, \quad |z_0 - w_0| \leq C\epsilon, \quad \rho_*(\mu_0, \mu^{\boldsymbol{x}_0}) \leq C\left(\epsilon + \alpha(N) + \sqrt{\epsilon\alpha(N)}\right),$$

where C is a constant independent of ϵ, N.

Proof. According to Lemma 2.3, we have

$$\left|V^N(t,z,\boldsymbol{x}) - V^N(s,w,\boldsymbol{x})\right| \leq C(\sqrt{|t-s|} + |w-z|).$$

Recall that

$$\begin{aligned} \overline{V}^{N,\epsilon}(t_0, z_0, \mu_0) =& V^N(s_0, w_0, \boldsymbol{x}_0) + \frac{1}{2\epsilon}|s_0 - t_0|^2 + \frac{1}{2\epsilon}|w_0 - z_0|^2 + \frac{1}{2\epsilon}\rho^2_*(\mu^{\boldsymbol{x}_0}, \mu_0) \\ \leq& V^N(t_0, w_0, \boldsymbol{x}_0) + \frac{1}{2\epsilon}|w_0 - z_0|^2 + \frac{1}{2\epsilon}\rho^2_*(\mu^{\boldsymbol{x}_0}, \mu_0). \end{aligned}$$

Therefore we get inequalities

$$\frac{1}{2\epsilon}|s_0-t_0|^2 \le V^N(t_0,w_0,\boldsymbol{x}_0)-V^N(s_0,w_0,\boldsymbol{x}_0) \le C\sqrt{|s_0-t_0|},$$

and hence $|s_0-t_0| \le C\epsilon^{2/3}$. By a similar argument, we obtain that $|w_0-z_0| \le C\epsilon$.

Thanks to Lemma 2.7, take $\boldsymbol{x}' \in \mathbb{T}^{dN}$ such that $\rho_*(\mu^{\boldsymbol{x}'},\mu_0) \le C\alpha(N)$. By the definition of $\overline{V}^{N,\epsilon}$,

$$\begin{aligned} &V^N(s_0,w_0,\boldsymbol{x}_0)+\frac{1}{2\epsilon}|s_0-t_0|^2+\frac{1}{2\epsilon}|w_0-z_0|^2+\frac{1}{2\epsilon}\rho_*^2(\mu^{\boldsymbol{x}_0},\mu_0)\\ &\le V^N(s_0,w_0,\boldsymbol{x}')+\frac{1}{2\epsilon}|s_0-t_0|^2+\frac{1}{2\epsilon}|w_0-z_0|^2+\frac{1}{2\epsilon}\rho_*^2(\mu^{\boldsymbol{x}'},\mu_0). \end{aligned}$$

and therefore in conjunction with Proposition 2.5 we have

$$\begin{aligned} &\frac{1}{2\epsilon}\rho_*^2(\mu^{\boldsymbol{x}_0},\mu_0)-\frac{1}{2\epsilon}\rho_*^2(\mu^{\boldsymbol{x}'},\mu_0) \le V^N(s_0,w_0,\boldsymbol{x}')-V^N(s_0,w_0,\boldsymbol{x}_0)\\ &\le C(\rho_*(\mu^{\boldsymbol{x}_0},\mu^{\boldsymbol{x}'})+\alpha(N)) \le C(\rho_*(\mu^{\boldsymbol{x}_0},\mu_0)+2\alpha(N)). \end{aligned}$$

Hence we obtain the inequality

$$\frac{1}{2\epsilon}\rho_*^2(\mu^{\boldsymbol{x}_0},\mu_0)-C\rho_*(\mu^{\boldsymbol{x}_0},\mu_0) \le \frac{1}{2\epsilon}C^2\alpha(N)^2+2C\alpha(N),$$

and thus

$$\rho_*(\mu^{\boldsymbol{x}_0},\mu_0) \le C\left(\epsilon+\alpha(N)+\sqrt{\epsilon\alpha(N)}\right), \qquad \square$$

where C is a constant independent of ϵ, N.

According to [31, Lemma 5.1,5.4], we have

$$\begin{aligned} \rho_*^2(\mu,\nu) &= \sum_{l\in\mathbb{Z}^d}(1+|l|^2)^{-k_*}|F_l(\mu-\nu)|^2,\\ D_\nu\rho_*^2(\mu,\nu)(x) &= -2i\sum_{l\in\mathbb{Z}^d} l(1+|l|^2)^{-k_*}F_l(\nu-\mu)e_l^*(x), \qquad (3.7)\\ D_\nu^2\rho_*^2(\mu,\nu)(x,y) &= -2\sum_{l\in\mathbb{Z}^d} l^2(1+|l|^2)^{-k_*}F_l(\nu-\mu)e_l^*(x)e_l(y). \end{aligned}$$

Therefore by Cauchy Schwarz inequality,

$$\begin{aligned} \frac{1}{2}|D_\nu\rho_*^2(\mu,\nu)(x)| &\le \rho_*(\mu,\nu)\sqrt{\sum_{l\in\mathbb{Z}^d}|l|^2(1+|l|^2)^{-k_*}},\\ \frac{1}{2}|D_\nu^2\rho_*^2(\mu,\nu)(x,y)| &\le \rho_*(\mu,\nu)\sqrt{\sum_{l\in\mathbb{Z}^d}|l|^4(1+|l|^2)^{-k_*}}. \qquad (3.8) \end{aligned}$$

where the constants $\sqrt{\sum_{l\in\mathbb{Z}^d}|l|^2(1+|l|^2)^{-k_*}}$ and $\sqrt{\sum_{l\in\mathbb{Z}^d}|l|^4(1+|l|^2)^{-k_*}}$ are finite due to our choice of k_*. Then according to Lemma 3.1 and Lemma 3.3,

$$|D_\mu\Phi(t_0,z_0,\mu_0)(x)| \le \frac{\sqrt{\sum_{l\in\mathbb{Z}^d}|l|^2(1+|l|^2)^{-k_*}}\left(\epsilon+\alpha(N)+\sqrt{\epsilon\alpha(N)}\right)}{\epsilon} =: R(\epsilon,N).$$

Here is an upper bound of $E(\epsilon,N)$ depending on $R(\epsilon,N)$.

LEMMA 3.4. *We have that*

$$E(\epsilon, N) \leq C_{R(\epsilon,N)}\epsilon + C_{R(\epsilon,N)}(1 + 1/(\epsilon N))\rho_*(\mu^{\boldsymbol{x}_0}, \mu_0) + \frac{1}{2\epsilon}\rho_*^2(\mu^{\boldsymbol{x}_0}, \mu_0)$$
$$+ \frac{\epsilon C_{R(\epsilon,N)}^2}{4}\left(\frac{\rho_*(\mu^{\boldsymbol{x}_0}, \mu_0)}{\epsilon} + \frac{\rho_*^{k_*-1}(\mu^{\boldsymbol{x}_0}, \mu_0)}{\epsilon^{k_*-1}}\right)^2,$$

where $C_{R(\epsilon,N)}$ is a constant that only depends on $R(\epsilon, N)$.

Proof. Let us denote the terms on the right-hand side of (3.6) by $(I), (II), (III)$, and (IV) from the top to the bottom, and we have the following estimate.

Estimate of (I): Recall that $H^e(x, z, p, \mu) := H(x + z, p, (I_d + z)_\# \mu)$, and hence

$$|H^e(x, z, p, \mu) - H^e(x, w, p, \mu)| \leq C_R(|w - z| + W_1((I_d + z)_\#\mu, (I_d + w)_\#\mu))$$
$$\leq 2C_{R(\epsilon,N)}|w - z|.$$

Together with Lemma 3.3, we conclude that $(I) \leq C_{R(\epsilon,N)}\epsilon$.

Estimate of (II): Let us denote $f(x) := H^e(x, z_0, \partial_\mu\Phi(\theta_0)(x), \mu^{\boldsymbol{x}_0})$. Then it is straightforward that $(II) = \int f(x)\,(\mu_0 - \mu^{\boldsymbol{x}_0})(dx)$. Due to Lemma 3.1,

$$D_\mu\Phi(\theta_0)(\cdot) = \frac{1}{2\epsilon}D_\nu\rho_*^2(\mu^{\boldsymbol{x}_0}, \mu_0)(\cdot),$$

According to [31, Lemma 5.4] the term on the right is in the Sobolev space $\mathbb{H}^{k_*-1}$. Indeed, according to (3.7)

$$\|D_\nu\rho_*^2(\mu^{\boldsymbol{x}_0}, \mu_0)(\cdot)\|_{k_*-1}^2 = 4\sum_{l\in\mathbb{Z}^d}(1 + |l|^2)^{k_*-1}|l|^2(1 + |l|^2)^{-2k_*}|F_l(\nu - \mu)|^2$$
$$\leq 4\rho_*^2(\mu^{\boldsymbol{x}_0}, \mu_0),$$

and hence

$$\|D_\mu\Phi(\theta_0)\|_{k_*-1} \leq \frac{1}{\epsilon}\rho_*(\mu^{\boldsymbol{x}_0}, \mu_0).$$

As $\|H^e(\cdot, z_0, \cdot, \mu^{\boldsymbol{x}_0})\|_{C^{k_*}(\mathbb{T}^d\times B_{R(\epsilon,N)})} \leq C_{R(\epsilon,N)}$, f is actually a composition $g \circ (I_d, h)$ of two functions g and (I_d, h) with $g = H^e(\cdot, z_0, \cdot, \mu^{\boldsymbol{x}_0}) \in C^{k_*}(\mathbb{T}^d \times B_{R(\epsilon,N)})$ and $h = D_\mu\Phi(\theta_0)(\cdot) \in \mathbb{H}^{k_*-1}(\mathbb{T}^d)$. Therefore according to the chain rule, [30, Remark 2, Section 5.2],

$$\|f\|_{k_*-1} = \|g \circ (I_d, h)\|_{k_*-1} \leq C(R_{\epsilon,N})\left(\|D_\mu\Phi(\theta_0)\|_{k_*-1} + \|D_\mu\Phi(\theta_0)\|_{k_*-1}^{k_*-1}\right)$$
$$\leq C(R_{\epsilon,N})\left(\frac{\rho_*(\mu^{\boldsymbol{x}_0}, \mu_0)}{\epsilon} + \frac{\rho_*^{k_*-1}(\mu^{\boldsymbol{x}_0}, \mu_0)}{\epsilon^{k_*-1}}\right)$$

where $C(R_{\epsilon,N})$ is a constant that only depends on $\|H^e(\cdot, z_0, \cdot, \mu^{\boldsymbol{x}_0})\|_{C^{k_*}(\mathbb{T}^d\times B_{R(\epsilon,N)})} \leq C_{R(\epsilon,N)}$. Therefore we obtain the estimate

$$(II) \leq C_{R(\epsilon,N)}\rho_{1-k_*}(\mu^{\boldsymbol{x}_0}, \mu_0)\left(\frac{\rho_*(\mu^{\boldsymbol{x}_0}, \mu_0)}{\epsilon} + \frac{\rho_*^{k_*-1}(\mu^{\boldsymbol{x}_0}, \mu_0)}{\epsilon^{k_*-1}}\right)$$

$$\leq \frac{1}{\epsilon}\rho_{1-k_*}^2(\mu^{\boldsymbol{x}_0},\mu_0) + \frac{\epsilon C_{R(\epsilon,N)}^2}{4}\left(\frac{\rho_*(\mu^{\boldsymbol{x}_0},\mu_0)}{\epsilon} + \frac{\rho_*^{k_*-1}(\mu^{\boldsymbol{x}_0},\mu_0)}{\epsilon^{k_*-1}}\right)^2.$$

Estimate of (III)*:* It is straightforward from Assumption 1.1 that (III) is bounded from above by $C_{R(\epsilon,N)}\rho_*(\mu^{\boldsymbol{x}_0},\mu_0)$.

Estimate of (IV)*:* According to Lemma 3.1, we compute

$$\begin{aligned}
(IV) =& -\frac{1}{\epsilon}\sum_{l\in\mathbb{Z}^d}\frac{|l|^2}{(1+|l|^2)^{k_*}}F_l(\mu_0-\mu^{\boldsymbol{x}_0})\int e_l^*(x)\,(\mu_0-\mu^{\boldsymbol{x}_0})(dx)\\
&-\sum_{i=1}^N\frac{1}{2\epsilon N^2}D_\nu^2\rho_*^2(\mu^{\boldsymbol{x}_0},\mu_0)(x_0^i,x_0^i)\\
=& -\frac{1}{\epsilon}\sum_{l\in\mathbb{Z}^d}\frac{|l|^2}{(1+|l|^2)^{k_*}}|F_l(\mu_0-\mu^{\boldsymbol{x}_0})|^2 - \sum_{i=1}^N\frac{1}{2\epsilon N^2}D_\nu^2\rho_*^2(\mu^{\boldsymbol{x}_0},\mu_0)(x_0^i,x_0^i)\\
=& -\frac{1}{\epsilon}\rho_{1-k_*}^2(\mu^{\boldsymbol{x}_0},\mu_0) + \frac{1}{\epsilon}\rho_*^2(\mu^{\boldsymbol{x}_0},\mu_0) - \sum_{i=1}^N\frac{1}{2\epsilon N^2}D_\nu^2\rho_*^2(\mu^{\boldsymbol{x}_0},\mu_0)(x_0^i,x_0^i)\\
\leq& -\frac{1}{\epsilon}\rho_{1-k_*}^2(\mu^{\boldsymbol{x}_0},\mu_0) + \frac{1}{\epsilon}\rho_*^2(\mu^{\boldsymbol{x}_0},\mu_0) + \frac{C}{\epsilon N}\rho_*(\mu^{\boldsymbol{x}_0},\mu_0),
\end{aligned}$$

where we apply the estimate (3.8) to get the last inequality. Summing up the estimates for $(I),(II),(III),(IV)$, we conclude the result. ∎

To finish the proof, we show that $\overline{V}^{\epsilon,N}(T,z,\mu) \geq G^e(z,\mu)$ up to some error. Suppose that

$$\overline{V}^{N,\epsilon}(T,z,\mu) = V^N(s_T,w_T,\boldsymbol{x}_T) + \frac{1}{2\epsilon}|T-s_T|^2 + \frac{1}{2\epsilon}|z-w_T|^2 + \frac{1}{2\epsilon}\rho_*^2(\mu^{\boldsymbol{x}_T},\mu).$$

By the same argument as in Lemma 3.3, it can be shown that

$$|T-s_T| \leq C\epsilon^{2/3}, \quad |z-w_T| \leq C\epsilon, \quad \rho_*(\mu_0,\mu^{\boldsymbol{x}_T}) \leq C\left(\epsilon+\alpha(N)+\sqrt{\epsilon\alpha(N)}\right),$$

where C is a constant independent of ϵ, N. Then thanks to the regularity of V^N from Lemma 2.3 and Proposition 2.5, we get that

$$\begin{aligned}
\overline{V}^{N,\epsilon}(T,z,\mu) &\geq V^N(T,z,\mu) - C(\epsilon^{1/3}+\epsilon) - C\alpha(N) - C\left(\epsilon+\alpha(N)+\sqrt{\epsilon\alpha(N)}\right)\\
&\geq G^e(z,\mu) - C(\epsilon^{1/3}+\epsilon+\alpha(N)+\sqrt{\epsilon\alpha(N)}).
\end{aligned}$$

Together with (3.2), $\overline{V}^{N,\epsilon}(t,z,\mu)+(T-t)E(\epsilon,N)+C(\epsilon^{1/3}+\epsilon+\alpha(N)+\sqrt{\epsilon\alpha(N)})$ is a viscosity supersolution to (1.1). Then due to the comparison principle in Proposition 2.6, we have that

$$\overline{V}^{N,\epsilon}(t,z,\mu)+(T-t)E(\epsilon,N)+C(\epsilon^{1/3}+\epsilon+\alpha(N)+\sqrt{\epsilon\alpha(N)}) \geq V(t,z,\mu).$$

Recalling that by the definition of inf-convolution, $V^N(t,z,\boldsymbol{x}) \geq \overline{V}^{N,\epsilon}(t,z,\mu^{\boldsymbol{x}})$, and hence

$$V^N(t,z,\boldsymbol{x}) \geq V(t,z,\mu^{\boldsymbol{x}}) - TE(\epsilon,N) - C(\epsilon^{1/3}+\epsilon+\alpha(N)+\sqrt{\epsilon\alpha(N)}).$$

Choosing $\epsilon = \alpha(N)$, it can be easily checked that $R(\epsilon, N)$ is bounded independent of N, and therefore

$$TE(\epsilon, N) + C(\epsilon^{1/3} + \epsilon + \alpha(N) + \sqrt{\epsilon\alpha(N)}) \leq C\alpha^{1/3}(N).$$

So that we finish the proof of (3.1).

Acknowledgments. We would like to thank Joe Jackson for his comments on the first version of the paper.

REFERENCES

[1] E. Bandini, A. Cosso, M. Fuhrman, and H. Pham, *Randomized filtering and bellman equation in wasserstein space for partial observation control problem*, Stochastic Processes and their Applications, 129 (2019), pp. 674–711.

[2] E. Bayraktar, A. Cecchin, and P. Chakraborty, *Mean field control and finite agent approximation for regime-switching jump diffusions*, Appl. Math. Optim., 88 (2023), pp. Paper No. 36, 35.

[3] E. Bayraktar, A. Cosso, and H. Pham, *Randomized dynamic programming principle and Feynman-Kac representation for optimal control of McKean-Vlasov dynamics*, Trans. Amer. Math. Soc., 370 (2018), pp. 2115–2160.

[4] E. Bayraktar, I. Ekren, and X. Zhang, *Comparison of viscosity solutions for a class of second order PDEs on the Wasserstein space*, To appear in Communications in Partial Differential Equations.

[5] E. Bayraktar, I. Ekren, and X. Zhang, *A PDE approach for regret bounds under partial monitoring*, Journal of Machine Learning Research, 24 (2023), pp. 1–24.

[6] C. Bertucci, *Stochastic optimal transport and Hamilton-Jacobi-Bellman equations on the set of probability measures*, To appear in AIHPC.

[7] H. Brezis, *Functional analysis, Sobolev spaces and partial differential equations*, Universitext, Springer, New York, 2011.

[8] M. Burzoni, V. Ignazio, A. M. Reppen, and H. M. Soner, *Viscosity solutions for controlled mckean–vlasov jump-diffusions*, SIAM Journal on Control and Optimization, 58 (2020), pp. 1676–1699.

[9] P. Cardaliaguet, S. Daudin, J. Jackson, and P. E. Souganidis, *An algebraic convergence rate for the optimal control of mckean–vlasov dynamics*, SIAM Journal on Control and Optimization, 61 (2023), pp. 3341–3369.

[10] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions, *The Master Equation and the Convergence Problem in Mean Field Games:(AMS-201)*, vol. 201, Princeton University Press, 2019.

[11] P. Cardaliaguet, J. Jackson, N. Mimikos-Stamatopoulos, and P. E. Souganidis, *Sharp convergence rates for mean field control in the region of strong regularity*, arXiv e-prints, (2023), p. arXiv:2312.11373.

[12] P. Cardaliaguet and C.-A. Lehalle, *Mean field game of controls and an application to trade crowding*, Math. Financ. Econ., 12 (2018), pp. 335–363.

[13] R. Carmona and F. Delarue, *Probabilistic theory of mean field games with applications. II*, vol. 84 of Probability Theory and Stochastic Modelling, Springer, Cham, 2018. Mean field games with common noise and master equations.

[14] R. Carmona and F. c. Delarue, *Probabilistic theory of mean field games with applications. I*, vol. 83 of Probability Theory and Stochastic Modelling, Springer, Cham, 2018. Mean field FBSDEs, control, and games.

[15] A. Cecchin, *Finite state N-agent and mean field control problems*, ESAIM Control Optim. Calc. Var., 27 (2021), pp. Paper No. 31, 33.

[16] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham, and M. Rosestolato, *Master Bellman equation in the Wasserstein space: uniqueness of viscosity solutions*, Trans. Amer. Math. Soc., 377 (2024), pp. 31–83.

[17] A. M. G. Cox, S. Källblad, M. Larsson, and S. Svaluto-Ferro, *Controlled measure-valued martingales: A viscosity solution approach*, The Annals of Applied Probability, 34 (2024), pp. 1987 – 2035.

[18] M. G. Crandall, H. Ishii, and P.-L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. (N.S.), 27 (1992), pp. 1–67.

[19] S. Daudin, F. Delarue, and J. Jackson, *On the optimal rate for the convergence problem in mean field control*, Journal of Functional Analysis, 287 (2024), p. 110660.

[20] S. Daudin, J. Jackson, and B. Seeger, *Well-posedness of Hamilton-Jacobi equations in the Wasserstein space: non-convex Hamiltonians and common noise*, To appear in Communications in Partial Differential Equations.

[21] S. Daudin and B. Seeger, *A comparison principle for semilinear Hamilton–Jacobi–Bellman equations in the Wasserstein space*, Calculus of Variations and Partial Differential Equations, 63 (2024), p. 106.

[22] N. Fournier and A. Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probab. Theory Related Fields, 162 (2015), pp. 707–738.

[23] W. Gangbo, S. Mayorga, and A. Swiech, *Finite dimensional approximations of hamilton–jacobi–bellman equations in spaces of probability measures*, SIAM Journal on Mathematical Analysis, 53 (2021), pp. 1320–1356.

[24] W. Gangbo, A. R. Mészáros, C. Mou, and J. Zhang, *Mean field games master equations with nonseparable Hamiltonians and displacement monotonicity*, Ann. Probab., 50 (2022), pp. 2178–2217.

[25] M. Germain, H. Pham, and X. Warin, *Rate of convergence for particle approximation of PDEs in Wasserstein space*, J. Appl. Probab., 59 (2022), pp. 992–1008.

[26] F. Gozzi and A. Świech, *Hamilton–Jacobi–Bellman equations for the optimal control of the Duncan–Mortensen–Zakai equation*, Journal of Functional Analysis, 172 (2000), pp. 466–510.

[27] D. Lacker, *Limit theory for controlled mckean–vlasov dynamics*, SIAM Journal on Control and Optimization, 55 (2017), pp. 1641–1672.

[28] M. Martini, *Kolmogorov equations on spaces of measures associated to nonlinear filtering processes*, Stochastic Processes and their Applications, 161 (2023), pp. 385–423.

[29] D. Possamaï and M. Talbi, *Mean-field games of optimal stopping: master equation and weak equilibria*, arXiv e-prints, (2023), p. arXiv:2307.09278.

[30] T. Runst and W. Sickel, *Sobolev spaces of fractional order, Nemytskij operators, and nonlinear partial differential equations*, vol. 3 of De Gruyter Series in Nonlinear Analysis and Applications, Walter de Gruyter & Co., Berlin, 1996.

[31] H. M. Soner and Q. Yan, *Viscosity solutions for McKean-Vlasov control on a torus*, SIAM J. Control Optim., 62 (2024), pp. 903–923.

[32] ———, *Viscosity solutions of the eikonal equation on the wasserstein space*, Applied Mathematics & Optimization, 90 (2024), p. 1.

[33] M. Talbi, *A finite-dimensional approximation for partial differential equations on wasserstein space*, Stochastic Processes and their Applications, 177 (2024), p. 104445.

[34] M. Talbi, N. Touzi, and J. Zhang, *Dynamic programming equation for the mean field optimal stopping problem*, SIAM J. Control Optim., 61 (2023), pp. 2140–2164.

[35] ———, *Viscosity solutions for obstacle problems on Wasserstein space*, SIAM J. Control Optim., 61 (2023), pp. 1712–1736.

[36] M. Talbi, N. Touzi, and J. Zhang, *From finite population optimal stopping to mean field optimal stopping*, The Annals of Applied Probability, 34 (2024), pp. 4237 – 4267.