SIAM J. MATH. ANAL.
Vol. 53, No. 2, pp. 1320–1356

FINITE DIMENSIONAL APPROXIMATIONS OF HAMILTON–JACOBI–BELLMAN EQUATIONS IN SPACES OF PROBABILITY MEASURES*

WILFRID GANGBO†, SERGIO MAYORGA‡, AND ANDRZEJ ŚWIĘCH§

Abstract. We prove that viscosity solutions of Hamilton–Jacobi–Bellman (HJB) equations, corresponding either to deterministic optimal control problems for systems of n particles or to stochastic optimal control problems for systems of n particles with a common noise, converge locally uniformly to the viscosity solution of a limiting HJB equation in the space of probability measures. We prove uniform continuity estimates for viscosity solutions of the approximating problems which may be of independent interest. We pay special attention to the case when the Hamiltonian is convex in the gradient variable and equations are of first order and provide a representation formula for the solution of the limiting first order HJB equation. We also propose an intrinsic definition of viscosity solution on the Wasserstein space.

Key words. Hamilton–Jacobi–Bellman equations in infinite dimension, viscosity solutions, mean field control, master equation, Wasserstein space, Wasserstein gradient, Wasserstein Laplacian

AMS subject classifications. 35D40, 35F21, 35R15, 49L25, 26E15

DOI. 10.1137/20M1331135

1. Introduction.

We consider the problem of approximation of Hamilton–Jacobi–Bellman (HJB) equations in spaces of probability measures by equations in finite dimensional spaces. More precisely, we study if an appropriately interpreted viscosity solution of such HJB equations can be approximated locally uniformly by viscosity solutions of finite dimensional problems. This is related to the problem of whether value functions of variational or optimal control problems in spaces of probability measures can be approximated by value functions corresponding to problems for finite particle systems. Similar convergence problems have been recently studied in the context of mean field games [15, 16, 18, 20, 21, 28, 37, 41, 50, 53]. In particular it was proved in [18] that classical solutions of finite dimensional second order Nash systems converge, in a suitable sense, to classical solutions of the corresponding master equations. Also, convergence of functionals of empirical measures of the marginal laws of particle systems for McKean–Vlasov stochastic differential equations was studied recently in [20, 23] using calculus in the space of measures, stochastic analysis, and partial differential equations in the space of measures. Explicit convergence estimates were obtained in [20, 23]. The problems investigated there are different from the one here. They studied the case of independent noises and no controls so they dealt with partial differential equations which are linear, have slightly different form, and have smooth solutions. We refer to the references in [20, 23] for

*Received by the editors April 13, 2020; accepted for publication (in revised form) November 2, 2020; published electronically March 8, 2021.
https://doi.org/10.1137/20M1331135
Funding: The work of the first author was supported by National Science Foundation grant DMS-1700202 and AFOSR grant FA9550-18-1-0502.
†Department of Mathematics, University of California at Los Angeles, Los Angeles, CA 90095 USA (wgangbo@math.ucla.edu).
‡Department of Mathematics, Baylor University, Waco, TX 76798 USA (sergio_mayorgatatari@baylor.edu).
§School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332 USA (swiech@math.gatech.edu).

the discussion of other earlier results in this direction. Regularity and convergence problems for finite dimensional approximations of first order HJB equations in spaces of probability measures were studied in [37, 39, 41, 52], either when solutions were regular or when the Hamiltonian was quadratic. Some results are also mentioned in [15] without proofs, while some results may be considered to be part of the folklore of the theory. In this paper we want to investigate the problem rigorously from the point of view of viscosity solutions.

We will be concerned with first and second order degenerate HJB equations of the form

$$\begin{cases} \partial_t \mathcal{U} - \kappa \Delta_w \mathcal{U} + \mathcal{H}(\mu, \mu, \nabla_\mu \mathcal{U}) + \mathcal{F}(\mu) = 0 \quad \text{in } (0,T) \times \mathcal{P}_2(\mathbb{R}^d), \\ \mathcal{U}(0,\mu) = \mathcal{U}_0(\mu) \quad \text{on } \mathcal{P}_2(\mathbb{R}^d), \end{cases} \tag{1.1}$$

where $\Delta_w \mathcal{U}$ is the partial Laplacian of $\mathcal{U}$ (see [24]), $T > 0, \kappa \geq 0, \mathcal{P}_2(\mathbb{R}^d)$ is the Wasserstein space of probability measures on $\mathbb{R}^d$ with bounded second moments, $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, and, for $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$, $\xi \in L^2_{\mu_1}(\mathbb{R}^d; \mathbb{R}^d)$, the Hamiltonian $\mathcal{H}$ is defined by

$$\mathcal{H}(\mu_1, \mu_2, \xi) = \int_{\mathbb{R}^d} H(x, \mu_2, \xi(x)) \mu_1(dx)$$

for some function $H : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ which satisfies Hypothesis 2.1 below. We bring to the reader's attention the fact that the dependence of $\mathcal{H}$ in (1.1) on the first μ is linear, whereas the dependence on the second μ comes from the integrand function H. Moreover, the function $\mathcal{F}$ cannot be simply absorbed into $\mathcal{H}$: otherwise, the restriction of μ to averages of Dirac masses would not yield (1.2) below. Indeed, part of the goal of this paper is to justify that the approximating finite dimensional problems should have the form

$$\begin{cases} \partial_t u_n - \kappa \mathrm{Tr}(A_n D^2 u_n) + \frac{1}{n} \sum_{i=1}^n H(x_i, \frac{1}{n-1} \sum_{j \neq i}^n \delta_{x_j}, n D_{x_i} u_n) \\ \qquad\qquad + \mathcal{F}(\frac{1}{n} \sum_{i=1}^n \delta_{x_i}) = 0 \quad \text{in } (0,T) \times (\mathbb{R}^d)^n, \\ u_n(0, x_1, \ldots, x_n) = \mathcal{U}_0(\frac{1}{n} \sum_{i=1}^n \delta_{x_i}) \quad \text{on } (\mathbb{R}^d)^n, \end{cases} \tag{1.2}$$

where for $n \in \mathbb{N}$, A_n is the $nd \times nd$ matrix composed of n^2 block matrices I_d.

Equation (1.1) will be interpreted in the L^2 sense, that is, we will look at the "lifted" version of (1.1) in the space

$$E := L^2(\Omega; \mathbb{R}^d),$$

where Ω is an atomless probability space. This technique was introduced in [15, 50] and its detailed exposition and recent developments can be found in [20, 21, 43]. Without loss of generality we can assume that $\Omega = (0,1)$ with the standard Lebesgue measure $\mathcal{L}_1$. We denote by $\langle \cdot, \cdot \rangle_1$ the inner product in $L^2(\Omega; \mathbb{R})$ and, for $X, Y \in L^2(\Omega; \mathbb{R}^d)$, we set

$$\langle X, Y \rangle_d := \big(\langle X_1, Y_1 \rangle_1, \ldots, \langle X_d, Y_d \rangle_1 \big),$$

$$\langle X, Y \rangle := \langle X_1, Y_1 \rangle_1 + \cdots + \langle X_d, Y_d \rangle_1 \quad \text{and} \quad |X| = \sqrt{\langle X, X \rangle},$$

where $X_1, \ldots, X_d$, $Y_1, \ldots, Y_d$ are the components of X and Y, respectively. We denote the canonical basis in $\mathbb{R}^d$ by $\{e_1, \ldots, e_d\}$ and consider its elements as constant functions in E. We define the functions $U_0, F : E \to \mathbb{R}$ by

$$U_0(X) = \mathcal{U}_0(X_\sharp \mathcal{L}_1), \quad F(X) = \mathcal{F}(X_\sharp \mathcal{L}_1),$$

where $\sharp$ denotes pushforward. Thus $X_\sharp\mathcal{L}_1$ is the law of the random vector X and is an element of $\mathcal{P}_2(\mathbb{R}^d)$. If $U : E \to \mathbb{R}$ is twice differentiable and such that $\mathcal{U}(\mu) = U(X)$ if μ is the law of X, then we have the crucial formula

$$\Delta_w \mathcal{U}(\mu) = \sum_{k=1}^{d} \langle D^2 U(X) e_k, e_k \rangle;$$

see section 5.

For $X, P \in E$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, we define

$$\tilde{H}(X, \mu, P) := \int_\Omega H(X(\omega), \mu, P(\omega)) d\omega.$$

We rely on the theory of viscosity solutions in Hilbert spaces and consider functions $U : [0,T) \times E \to \mathbb{R}$ which are viscosity solutions of

$$\begin{cases} \partial_t U - \kappa \sum_{k=1}^{d} \langle D^2 U e_k, e_k \rangle + \tilde{H}(X, X_\sharp\mathcal{L}_1, DU) + F(X) = 0 \quad \text{in } (0,T) \times E, \\ U(0,X) = U_0(X) \quad \text{on } E. \end{cases} \tag{1.3}$$

Here, DU, D^2U stand for the Fréchet derivatives of U with respect to the X variable. We refer the readers to [29] for the theory of viscosity solutions in Hilbert spaces and extensive references.

We have the following definition.

Definition 1.1. *Let $\mathcal{U} : [0,T) \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and define $U : [0,T) \times E \to \mathbb{R}$ by $U(t,X) = \mathcal{U}(t, X_\sharp\mathcal{L}_1)$.*

(i) *We say that $\mathcal{U}$ is an L-viscosity subsolution of* (1.1) *on the Wasserstein space if U is a viscosity subsolution of* (1.3) *on $[0,T) \times E$.*

(ii) *We say that $\mathcal{U}$ is an L-viscosity supersolution of* (1.1) *on the Wasserstein space if U is a viscosity supersolution of* (1.3) *on $[0,T) \times E$.*

(ii) *When $\mathcal{U}$ is both an L-viscosity subsolution and an L-viscosity supersolution of* (1.1) *on the Wasserstein space, we say that it is an L-viscosity solution of* (1.1) *on the Wasserstein space.*

We remark that another definition of viscosity solution to HJB master equations in the Wasserstein space, also called an L-viscosity solution, was introduced in [59]. The definition in [59] uses the framework of path dependent PDE and is not related to Definition 1.1 here, even though both notions have the same name.

We also propose an intrinsic definition of viscosity solution on the Wasserstein space and show in section 5 that the notion of L-viscosity solution provides a way to select particular intrinsic viscosity solutions. Only when the Hamiltonian is convex in the momentum variables and $\kappa = 0$ is it known that the notions of L-viscosity solution and intrinsic viscosity solution are equivalent [43].

The main result of the manuscript is the following convergence theorem.

Theorem 1.2. *Let Hypothesis* 2.1 *be satisfied and let $\kappa \geq 0$. Suppose that for $n \geq 1$ the functions $u_n : [0,T) \times (\mathbb{R}^d)^n \to \mathbb{R}$ are the viscosity solutions of* (1.2). *Then, for every bounded set B in $\mathcal{P}_2(\mathbb{R}^d)$,*

$$\lim_{n\to\infty} \sup \left\{ \left| u_n(t,x_1,\dots,x_n) - \mathcal{U}\left(t, \frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right)\right| : \right.$$
$$\left. (t,x_1,\dots,x_n) \in [0,T)\times(\mathbb{R}^d)^n, \frac{1}{n}\sum_{i=1}^n \delta_{x_i} \in B \right\} = 0,$$

where $\mathcal{U}$ is the unique L-viscosity solution of (1.1) *on the Wasserstein space.*

The assumptions of Hypothesis 2.1 will be introduced in section 2. To prove Theorem 1.2 we first obtain appropriate uniform continuity estimates for the solutions u_n of (1.2). This is done in Theorem 3.3 for a more general case when the second order coefficients A_n may depend on x. Theorem 3.3 is the main technical result of the paper and is of independent interest. We then convert the functions u_n into functions of empirical measures by defining new functions

$$\mathcal{V}_n(t,\mu_{\mathbf{x}}) := u_n(t,\mathbf{x}), \quad \text{where } \mu_{\mathbf{x}} := \frac{1}{n}\sum_{i=1}^n \delta_{x_i},\ \mathbf{x}=(x_1,\dots,x_n),$$

which are well defined since the functions u_n are invariant with respect to permutations of the variables of $\mathbf{x}$. The estimates of Theorem 3.3 guarantee that $\mathcal{V}_n$ are uniformly continuous in the topology of $[0,T]\times\mathcal{P}_r(\mathbb{R}^d)$, where $1<r<2$. We then extend $\mathcal{V}_n$ to $[0,T]\times\mathcal{P}_2(\mathbb{R}^d)$, preserving its modulus of continuity, and then use the Arzelà–Ascoli theorem to pass to the limit, along a subsequence, to a function $\mathcal{V}$ defined on $[0,T]\times\mathcal{P}_2(\mathbb{R}^d)$. We then prove directly that its "lifted" version $V:[0,T]\times E\to\mathbb{R}$ is a viscosity solution of (1.3). Uniqueness of viscosity solutions of (1.3) then guarantees that the whole sequence $\mathcal{V}_n$ converges to $\mathcal{V}$. Thus we completely avoid dealing with (1.1) in the space of probability measures which may not have a unique viscosity solution in the sense of [38] (see [43]). In section 6 we show that if $\kappa=0$ and H is convex in the gradient variable, then the functions u_n, which are value functions of optimal control problems for n-particle systems, converge to the value function of a variational problem in $\mathcal{P}_2(\mathbb{R}^d)$, thus giving a representation formula for the solution of (1.1). Finally we prove a few technical results in the appendix.

Equations (1.2) correspond either to deterministic optimal control problems for systems of n particles or to stochastic optimal control problems for systems of n particles with a common noise. Theorem 1.2 solves the problem of convergence for a large class of general first order HJB equations, even though the identification of the limit as a value function is only obtained for the convex case and $\kappa=0$. However, using the methods of this paper we were not able to obtain a result similar to Theorem 1.2 for other stochastic particle systems, for instance, systems of n particles with nonconstant diffusion coefficients, in which case the matrices A_n are functions like in Hypothesis 3.1, or for systems of n particles with independent noises, in which case $A_n = I_{nd}$. We also remark that some assumptions of Hypothesis 2.1 could be changed or relaxed while some may pose a bigger problem. This is worth investigating. The main challenge is in proving uniform continuity estimates of Theorem 3.3. We do not consider other cases here as Hypothesis 2.1 is sufficiently general and we do not want to overburden the presentation with too many technicalities. Our main goal is to convey the basic ideas. The readers can explore various generalizations.

HJB equations and master equations for mean field games or mean field control problems in spaces of probability measures have been studied a lot in recent years using various approaches. We refer the readers to [6, 7, 8, 9, 10, 11, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 28, 37, 38, 39, 41, 43, 45, 46, 52, 53, 54, 57, 58, 59]. Equations related

to differential games were studied in [25, 47], equations related to control problems with partial observation were studied in [4, 5], and HJB equations mostly related to large deviations and fluid dynamics problems were investigated with slightly different techniques in [30, 31, 32, 33, 34]. In particular an abstract method of relaxed limits for viscosity solutions was introduced in [31] for applications in large deviations and this technique was recently generalized in [48, 49]. HJB equations in metric spaces were studied by various techniques in [1, 12, 13, 39, 40, 44, 51, 55, 56]. Finally we refer the readers to [26, 35] for an introduction to the theory of viscosity solutions of partial differential equations in $\mathbb{R}^d$ and to [29] and the references there for the overview of the theory of viscosity solutions of second order HJB equations in Hilbert spaces.

2. Notation, assumptions, and definitions. We denote by $\mathcal{P}_r(\mathbb{R}^d), r \geq 1$, the space of Borel probability measures on $\mathbb{R}^d$ with finite rth moments, equipped with the Wasserstein r-metric

$$d_r(\mu,\nu) := \inf_{\gamma\in\Gamma(\mu,\nu)} \left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^r \gamma(dx,dy)\right)^{\frac{1}{r}},$$

where $\Gamma(\mu,\nu)$ is the set of all Borel probability measures γ on $\mathbb{R}^d\times\mathbb{R}^d$ with marginals μ,ν. The set of optimal measures in $\Gamma(\mu,\nu)$ will be denoted by $\Gamma_0(\mu,\nu)$. When $\mathbf{x}=(x_1,\dots x_n)\in(\mathbb{R}^d)^n$ we set

$$\mu_{\mathbf{x}} := \frac{1}{n}\sum_{i=1}^n \delta_{x_i} \quad \text{and} \quad |\mathbf{x}|_r = \frac{1}{n^{1/r}}\left(\sum_{i=1}^n |x_i|^r\right)^{1/r}.$$

We have

$$d_r(\mu_{\mathbf{x}},\mu_{\mathbf{y}}) = \inf_{\sigma} |\mathbf{x}-\mathbf{y}_\sigma|_r,$$

where the infimum is taken over all permutations σ of $\{1,\dots n\}$ and $\mathbf{y}_\sigma = (y_{\sigma(1)},\dots y_{\sigma(n)})$.

We use $\mathcal{L}_1$ to denote the Lebesgue measure on $\mathbb{R}$. If $X\in L^r(\Omega;\mathbb{R}^d)$, then its L^r norm will be denoted by $|X|_r$. If $r=2$ we will just write $|X|$ as in the introduction. For $X\in L^r(\Omega;\mathbb{R}^d)$, $\text{law}(X) := X_\sharp\mathcal{L}_1$ denotes the measure in $\mathcal{P}_r(\mathbb{R}^d)$ which is the pushforward of $\mathcal{L}_1$ by X.

When $\mu\in\mathcal{P}_2(\mathbb{R}^d)$, we denote as $L^2_\mu(\mathbb{R}^d;\mathbb{R}^d)$ the set of Borel vector fields $\xi:\mathbb{R}^d\to\mathbb{R}^d$ such that $\int_{\mathbb{R}^d}|\xi|^2\mu(dx)<\infty$. The tangent space at μ, denoted by $T_\mu\mathcal{P}_2(\mathbb{R}^d)$, is the closure of $\nabla C_c^\infty(\mathbb{R}^d)$ in $L^2_\mu(\mathbb{R}^d;\mathbb{R}^d)$.

For $x\in\mathbb{R}^m$ we will also be using the notation $|x|$ to denote the standard Euclidean norm in $\mathbb{R}^m$ and we will write $x\cdot y$ for $x,y\in\mathbb{R}^m$ to denote the dot product in $\mathbb{R}^m$.

If A is a matrix or a bounded operator in a Hilbert space, we will write $\|A\|$ to denote the operator norm of A. We denote by $S(m)$ the set of $m\times m$ symmetric matrices. If $A\in S(m)$, $\text{Tr}(A)$ means the trace of A.

For an open set $\mathscr{O}\subset\mathbb{R}^m$, we will write $C^1(\mathscr{O}), C^2(\mathscr{O})$ for the standard spaces of once and twice continuously differentiable functions on $\mathscr{O}$.

If W is a Hilbert space, we denote by $C^{1,2}((0,T)\times W)$ the space of functions $\varphi:(0,T)\times W\to\mathbb{R}$ such that $\partial_t\varphi, D\varphi, D^2\varphi$ are continuous on $(0,T)\times W$, where $D\varphi, D^2\varphi$ stand for the Fréchet derivatives of φ with respect to the Hilbert space variable.

Throughout the paper we will always identify a Hilbert space with its dual. Thus, with this identification, $D\varphi:(0,T)\times W\to W$ and $D^2\varphi:(0,T)\times W\to S(W)$, where $S(W)$ is the space of bounded self-adjoint operators in W.

We make the following assumptions about the Hamiltonian function H.

Hypothesis 2.1. Let $1 < r < 2$.

(i) The function $H : \mathbb{R}^d \times \mathcal{P}_r(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ is such that

(2.1)
$$|H(x,\nu,p) - H(x,\nu,q)| \le C(1+|p|+|q|)|p-q| \quad \forall\, p,q,x \in \mathbb{R}^d, \nu \in \mathcal{P}_r(\mathbb{R}^d),$$

(2.2)
$$|H(x,\mu,p) - H(y,\nu,p)| \le \sigma\big((|x-y| + d_r(\mu,\nu))(1+|p|)\big) \quad \forall\, p,x,y \in \mathbb{R}^d, \mu, \nu \in \mathcal{P}_r(\mathbb{R}^d)$$

for some concave modulus of continuity σ, and

(2.3)
$$|H(x,\mu,p)| \le C(1+|p|^2) \quad \forall\, p,x \in \mathbb{R}^d, \mu \in \mathcal{P}_r(\mathbb{R}^d).$$

(ii) The functions $\mathcal{U}_0, \mathcal{F} \in UC_b(\mathcal{P}_r(\mathbb{R}^d))$ (the space of bounded and uniformly continuous functions).

We notice that it easily follows from (2.1) and (2.3) that for all $X, P, Q \in E$,

$$|\tilde{H}(X,\mu,P) - \tilde{H}(X,\mu,Q)| \le C(1+|P|+|Q|)|P-Q| \tag{2.4}$$

and

$$|\tilde{H}(X,\mu,P)| \le C(1+|P|^2). \tag{2.5}$$

Moreover, by the concavity of σ and Jensen's inequality, we obtain that for all $X, Y, P \in E$, $\mu, \nu \in \mathcal{P}_r(\mathbb{R}^d)$,

$$\begin{aligned}
|\tilde{H}(X,\mu,P) - \tilde{H}(Y,\nu,P)| &\le \int_\Omega \sigma\big((|X(\omega)-Y(\omega)| + d_r(\mu,\nu))(1+|P(\omega)|)\big)d\omega \\
&\le \sigma\left(\int_\Omega (|X(\omega)-Y(\omega)| + d_r(\mu,\nu))(1+|P(\omega)|)d\omega\right) \\
&\le \sigma\left((|X-Y| + d_r(\mu,\nu))(1+|P|)\right). \qquad (2.6)
\end{aligned}$$

Let m_1 be a modulus of continuity for $\mathcal{U}_0$ and $\mathcal{F}$. Since

$$|U_0(X) - U_0(Y)| = |\mathcal{U}_0(X_\sharp \mathcal{L}_1) - \mathcal{U}_0(Y_\sharp \mathcal{L}_1)| \le m_1\left(d_r(X_\sharp \mathcal{L}_1, Y_\sharp \mathcal{L}_1)\right) \le m_1\left(|X-Y|_r\right)$$

we conclude

$$|U_0(X) - U_0(Y)| \le m_1\left(|X-Y|\right). \tag{2.7}$$

Similarly,

$$|F(X) - F(Y)| \le m_1\left(|X-Y|\right). \tag{2.8}$$

Moreover, for $\mathbf{x}, \mathbf{y} \in (\mathbb{R}^d)^n$,

$$|\mathcal{U}_0(\mu_{\mathbf{x}}) - \mathcal{U}_0(\mu_{\mathbf{y}})| \le m_1\left(|\mathbf{x}-\mathbf{y}|_r\right), \tag{2.9}$$

$$|\mathcal{F}(\mu_{\mathbf{x}}) - \mathcal{F}(\mu_{\mathbf{y}})| \le m_1\left(|\mathbf{x}-\mathbf{y}|_r\right). \tag{2.10}$$

Let W be a real Hilbert space with an inner product $\langle \cdot, \cdot \rangle$ and the norm $|\cdot|$. We recall the definitions of parabolic second order jets (see [26, section 8] and [27, section 3]).

Given $u : (0,T) \times W \to \mathbb{R}$ and $(\bar{t}, \bar{x}) \in (0,T) \times W$, the parabolic second order superjet of u at $(\bar{t}, \bar{x})$, $\mathcal{P}^{2,+}u(\bar{t}, \bar{x})$ is defined by

$$\mathcal{P}^{2,+}u(\bar{t}, \bar{x}) := \Big\{ (a, p, S) \in \mathbb{R} \times W \times S(W) :$$

$$\limsup_{(s,y) \to (\bar{t}, \bar{x})} \frac{u(s,y) - u(\bar{t}, \bar{x}) - a(s - \bar{t}) - \langle p, y - \bar{x} \rangle - \frac{1}{2} \langle S(y - \bar{x}), y - \bar{x} \rangle}{|s - \bar{t}| + |y - \bar{x}|^2} \leq 0 \Big\}. \tag{2.11}$$

The parabolic second order subjet of u at $(\bar{t}, \bar{x})$, $\mathcal{P}^{2,-}u(\bar{t}, \bar{x})$, is defined by reversing the inequality and replacing lim sup by lim inf in (2.11). The closure $\overline{\mathcal{P}}^{2,+}u(\bar{t}, \bar{x})$ of $\mathcal{P}^{2,+}u(\bar{t}, \bar{x})$, is defined as follows:

$$\overline{\mathcal{P}}^{2,+}u(\bar{t}, \bar{x}) := \Big\{ (a, p, S) \in \mathbb{R} \times W \times S(W) : \text{there exist } (t_n, x_n) \text{ and}$$

$$(a_n, p_n, S_n) \in \mathcal{P}^{2,+}u(t_n, x_n) \text{ s.t. } (t_n, x_n, u(t_n, x_n), p_n, S_n) \to (\bar{t}, \bar{x}, u(\bar{t}, \bar{x}), a, p, S) \Big\}.$$

The closure $\overline{\mathcal{P}}^{2,-}u(\bar{t}, \bar{x})$ of $\mathcal{P}^{2,-}u(\bar{t}, \bar{x})$ is defined similarly. We recall the definition of viscosity solution of an equation

$$\begin{cases} \partial_t u + G(t, x, u, Du, D^2 u) = 0 & \text{in } (0,T) \times W, \\ u(0,x) = g(x) & \text{on } W, \end{cases} \tag{2.12}$$

where $G : (0,T) \times W \times \mathbb{R} \times W \times S(W) \to \mathbb{R}$ is continuous.

Definition 2.2. *An upper semicontinuous function $u : [0,T) \times W \to \mathbb{R}$ is a viscosity subsolution of (2.12) if $u(0,x) \leq g(x)$ on W and whenever $\varphi \in C^{1,2}((0,T) \times W)$ and $u - \varphi$ has a local maximum at $(t,x) \in (0,T) \times W$, then*

$$\partial_t \varphi(t,x) + G(t, x, u(t,x), D\varphi(t,x), D^2\varphi(t,x)) \leq 0.$$

A lower semicontinuous function $u : [0,T) \times W \to \mathbb{R}$ is a viscosity supersolution of (2.12) if $u(0,x) \geq g(x)$ on W and whenever $\varphi \in C^{1,2}((0,T) \times W)$ and $u - \varphi$ has a local minimum at $(t,x) \in (0,T) \times W$, then

$$\partial_t \varphi(t,x) + G(t, x, u(t,x), D\varphi(t,x), D^2\varphi(t,x)) \geq 0.$$

A function u is a viscosity solution of (2.12) if it is a viscosity subsolution of (2.12) and a viscosity supersolution of (2.12).

It is easy to see that $(a, p, S) \in \mathcal{P}^{2,+}u(\bar{t}, \bar{x})$ (respectively, $(a, p, S) \in \mathcal{P}^{2,-}u(\bar{t}, \bar{x})$) if and only if there exists $\varphi \in C^{1,2}((0,T) \times W)$ such that $u - \varphi$ has a local maximum at $(\bar{t}, \bar{x})$ (respectively, $u - \varphi$ has a local minimum at $(\bar{t}, \bar{x})$) and

$$a = \partial_t \varphi(\bar{t}, \bar{x}), \quad p = D\varphi(\bar{t}, \bar{x}), \quad S = D^2\varphi(\bar{t}, \bar{x}).$$

The proof when $W = \mathbb{R}^m$ is in [35, Lemma 4.1], and it easily generalizes to the case of an infinite dimensional Hilbert space. Thus, since G is continuous, Definition 2.2 is equivalent to the definition using the closures of parabolic jets.

PROPOSITION 2.3. *An upper semicontinuous function* $u : [0,T) \times W \to \mathbb{R}$ *is a viscosity subsolution of* (2.12) *if* $u(0,x) \le g(x)$ *on* W *and*

$$a + G(t,x,u(t,x),p,S) \le 0 \quad \forall\ (t,x) \in (0,T) \times W \text{ and } (a,p,S) \in \overline{\mathcal{P}}^{2,+}u(t,x).$$

A lower semicontinuous function $u : [0,T) \times W \to \mathbb{R}$ *is a viscosity supersolution of* (2.12) *if* $u(0,x) \ge g(x)$ *on* W *and*

$$a + G(t,x,u(t,x),p,S) \ge 0 \quad \forall\ (t,x) \in (0,T) \times W \text{ and } (a,p,S) \in \overline{\mathcal{P}}^{2,-}u(t,x).$$

Remark 2.4. If (2.12) is of first order, that is, if $G : (0,T) \times W \times \mathbb{R} \times W \to \mathbb{R}$, then the test functions φ in Definition 2.2 are replaced by functions $C^1((0,T) \times W)$ and the parabolic second order superjet and subjets of u, $\mathcal{P}^{2,+}u$, and $\mathcal{P}^{2,-}u$ are replaced by the first order superdifferentials and subdifferentials D^+u and D^-u, respectively. Proposition 2.3 is then still true if $\overline{\mathcal{P}}^{2,+}u(t,x)$ and $\overline{\mathcal{P}}^{2,-}u(t,x)$ are replaced by the closures $\overline{D^+}u(t,x)$ and $\overline{D^-}u(t,x)$ which are defined similarly as the closures of the parabolic jets.

3. Estimates for finite dimensional equations. In this section we will consider a more general version of (1.2) with second order coefficients A_n depending on $\mathbf{x}$ or being more general constant matrices.

Hypothesis 3.1. For $\mathbf{x} = (x_1, \dots x_n)$, let $A_n(\mathbf{x}), n = 1, 2, \dots$, the $nd \times nd$ matrix composed of n^2 block matrices $a(x_i)a^*(x_j)$, $i, j = 1, 2, \dots n$, such that the function $a : \mathbb{R}^d \to S(d)$ is bounded and there exists $L \ge 0$ such that

$$\|a(x) - a(y)\| \le L|x - y| \quad \forall\ x, y \in \mathbb{R}^d. \tag{3.1}$$

The proof of the main theorem of this section uses the following simplified version of a well-known lemma (cf. Lemma 3.80 of [29]).

LEMMA 3.2. *Let* $\delta > 0$*, and let* σ_1 *be a modulus of continuity. Then there exists a nondecreasing, concave,* C^2 *function* φ_δ *on* $[0, +\infty)$ *such that* $\varphi_\delta(0) < \delta$ *and*

$$\sigma_1(\varphi_\delta'(s)s + s) \le \varphi_\delta(s) \qquad \text{for } 0 \le s \le 2. \tag{3.2}$$

THEOREM 3.3. *Let Hypothesis* 2.1 *be satisfied. Let* $A_n(\mathbf{x})$ *satisfy Hypothesis* 3.1*, or let* $A_n(\mathbf{x}) = A_n$*, where* A_n *is any sequence of* $nd \times nd$ *symmetric matrices with constant coefficients such that* $A_n \ge 0$ *and* $\mathrm{Tr}(A_n) \le \tilde{L}n$ *for some* $\tilde{L} \ge 0$*. Suppose that for* $n \ge 1$ *the functions* $u_n : [0,T] \times (\mathbb{R}^d)^n \to \mathbb{R}$ *are the viscosity solutions of*

$$\begin{cases} \partial_t u_n - \kappa \mathrm{Tr}\big(A_n D^2 u_n\big) + \frac{1}{n}\sum_{i=1}^n H\big(x_i, \frac{1}{n-1}\sum_{j\ne i}^n \delta_{x_j}, nD_{x_i}u_n\big) \\ \qquad\qquad + \mathcal{F}\big(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\big) = 0 \quad \text{in} \quad (0,T) \times (\mathbb{R}^d)^n, \\ u_n(0, x_1, \dots, x_n) = \mathcal{U}_0\big(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\big) \quad \text{on } (\mathbb{R}^d)^n. \end{cases} \tag{3.3}$$

Then there exists a modulus of continuity ρ *such that for every* n*,*

$$|u_n(t,\mathbf{x}) - u_n(s,\mathbf{y})| \le \rho(|t-s| + |\mathbf{x}-\mathbf{y}|_r) \quad \forall \quad t, s \in [0,T], \mathbf{x}, \mathbf{y} \in (\mathbb{R}^d)^n. \tag{3.4}$$

Proof. We first note that if u is a bounded viscosity subsolution of (3.3) and v is a bounded viscosity supersolution of (3.3) and if we replace $u_n(t,\mathbf{x}) - u_n(t,\mathbf{y})$ by $u(t,\mathbf{x}) - v(t,\mathbf{y})$ in the proof of continuity estimates below, the same arguments work and we obtain that $u(t,\mathbf{x}) - v(t,\mathbf{x}) \le 0$ for all $(t,\mathbf{x})$. Thus the comparison theorem

holds for bounded viscosity subsolutions and bounded viscosity supersolutions of (3.3). Moreover, since the function

$$\mathbf{x} \longmapsto g(\mathbf{x}) = \mathcal{U}_0\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}\right)$$

is bounded and uniformly continuous on $(\mathbb{R}^d)^n$, if $(g_m)_{m=1}^{\infty} \subset C^2((\mathbb{R}^d)^n)$ have bounded first and second derivatives and $0 \le g_m - g \le a_m \to 0$ as $m \to \infty$, then for sufficiently large $C_m > 0$ the functions $g_m(\mathbf{x}) + C_m t$ and $g_m(\mathbf{x}) - a_m - C_m t$ are respectively viscosity supersolutions and viscosity subsolutions of (3.3) for all $m = 1, 2, \ldots$. Thus the functions

$$\overline{u}(t,\mathbf{x}) := \inf_{m\ge 1}(g_m(\mathbf{x}) + C_m t), \quad \underline{u}(t,\mathbf{x}) := \sup_{m\ge 1}(g_m(\mathbf{x}) - a_m - C_m t)$$

are respectively a viscosity supersolution and viscosity subsolution of (3.3) such that $\overline{u}(0,\mathbf{x}) = \underline{u}(0,\mathbf{x}) = g(\mathbf{x})$ on $(\mathbb{R}^d)^n$. Therefore the unique bounded continuous viscosity solution u_n of (3.3) can be constructed, for instance, by means of Perron's method (see [26]).

It is also easy to see that there exists $M > 0$ such that $\|u_n\|_\infty \le M, n = 1, 2, \ldots$. To show this, let $K > 0$ be such that $\|\mathcal{F}\|_\infty \le K, \|\mathcal{U}_0\|_\infty \le K$. Recall that we have $\|H(\cdot,\cdot,0)\|_\infty \le C$, where C is from (2.3). Then $w_1(t,\mathbf{x}) := -K - (C+K)t$ is a bounded viscosity subsolution of (3.3) and $w_2(t,\mathbf{x}) := K + (C+K)t$ is a bounded viscosity supersolution of (3.3). Therefore, by comparison we obtain that for every n

$$w_1 \le u_n \le w_2 \quad \text{on } [0,T]\times(\mathbb{R}^d)^n.$$

This gives the required bound with $M = K + (C+K)T$.

For $\delta > 0$, let φ_δ be the function from Lemma 3.2 applied to the modulus

$$\sigma_1(s) = (1+T)\sigma(3s) + m_1(s) + (2\kappa d L^2(1+T) + 2M + 1)s.$$

In particular we have

$$\varphi_\delta(1) \ge 2M+1, \quad \varphi_\delta(s) \ge m_1(s). \tag{3.5}$$

First, we are going to show that for every $\delta > 0$

$$u_n(t,\mathbf{x}) - u_n(t,\mathbf{y}) \le \varphi_\delta(|\mathbf{x}-\mathbf{y}|_r)(1+t) \tag{3.6}$$

if $t \in [0,T]$ and $(\mathbf{x},\mathbf{y}) \in (\mathbb{R}^d)^n$.

We define smooth approximations of $|\mathbf{z}|_r$. For $\gamma > 0$, let

$$\psi_\gamma(\mathbf{z}) = \frac{1}{n^{1/r}}\left(\sum_{i=1}^{n}(\gamma + |z_i|^2)^{\frac{r}{2}}\right)^{\frac{1}{r}}.$$

We now set

$$\varphi(t,\mathbf{x},\mathbf{y}) := \varphi_\delta\left(\psi_\gamma\left(\mathbf{x}-\mathbf{y}\right)\right)(1+t).$$

Suppose that there exist $\gamma, \mu > 0$ such that

$$\sup_{\mathbf{x},\mathbf{y}\in(\mathbb{R}^d)^n, t\in[0,T)}\left(u_n(t,\mathbf{x}) - u_n(t,\mathbf{y}) - \frac{\mu}{T-t} - \varphi(t,\mathbf{x},\mathbf{y})\right) > 0. \tag{3.7}$$

Then, for every $\alpha > 0$ small enough and $h(\mathbf{x}) := (1+|\mathbf{x}|^2)^{1/2}$,

(3.8)
$$\sup_{\mathbf{x},\mathbf{y}\in(\mathbb{R}^d)^n, t\in[0,T)} \left(u_n(t,\mathbf{x}) - u_n(t,\mathbf{y}) - \frac{\mu}{T-t} - \varphi(t,\mathbf{x},\mathbf{y}) - \alpha(h(\mathbf{x})+h(\mathbf{y})) \right) > 0.$$

Moreover, since u_n is bounded and $\alpha(h(\mathbf{x})+h(\mathbf{y})) \to +\infty$ as $x, y \to \infty$, the supremum of the above expression over $[0,T)\times(\mathbb{R}^d)^n\times(\mathbb{R}^d)^n$ is attained at some point $(\bar t, \bar{\mathbf{x}}, \bar{\mathbf{y}})$. Obviously $\bar t < T$ and it follows from the definition of the function $\varphi(t,\mathbf{x},\mathbf{y})$ and (3.5) that we must have $0 < \bar t$. It also follows from (3.5) that $\bar s = \psi_\gamma(\bar{\mathbf{x}} - \bar{\mathbf{y}}) < 1$. Also $\bar{\mathbf{x}} \neq \bar{\mathbf{y}}$ since if $\bar{\mathbf{x}} = \bar{\mathbf{y}}$, the expression in (3.8) is negative. We compute

$$D_{x_i}\varphi(\bar t, \bar{\mathbf{x}}, \bar{\mathbf{y}})) = \varphi'_\delta(\bar s)\frac{(\bar x_i - \bar y_i)(\gamma + |\bar x_i - \bar y_i|^2)^{\frac{r}{2}-1}}{n\bar s^{\frac{r}{r'}}}(1+\bar t),$$

where $1/r + 1/r' = 1$. Also, using the concavity of φ_δ, we have

$$D^2_{\mathbf{x}}\varphi(\bar t, \bar{\mathbf{x}}, \bar{\mathbf{y}})) = B_1 - B_2,$$

where B_1 is a diagonal matrix composed of n diagonal $d\times d$ blocks

$$B_{1i} = \varphi'_\delta(\bar s)\frac{(\gamma + |\bar x_i - \bar y_i|^2)^{\frac{r}{2}-1}}{n\bar s^{\frac{r}{r'}}}(1+\bar t)I_d,$$

and $B_2 \geq 0$ is a symmetric matrix. Therefore

$$D = D^2\varphi(\bar t, \bar{\mathbf{x}}, \bar{\mathbf{y}}) = \begin{pmatrix} B_1 & -B_1 \\ -B_1 & B_1 \end{pmatrix} - \begin{pmatrix} B_2 & -B_2 \\ -B_2 & B_2 \end{pmatrix} =: D_1 - D_2,$$

where $D^2\varphi$ above is the second derivative of φ with respect to the variables $(\mathbf{x},\mathbf{y})$. We now use Theorem 8.3 of [26] applied to the functions

$$u^1(t,x) := u_n(t,x) - \alpha h(x), \quad u^2(t,y) := -u_n(t,y) - \alpha h(y).$$

We notice that, since u_n is a viscosity solution of (3.3), condition (8.5) of [26] is satisfied. Therefore, it follows from Theorem 8.3 of [26], applied with $\varepsilon = 1/(\|D_1\| + \|D_2\|)$, that there exist $b_1, b_2 \in \mathbb{R}$ and $S_1, S_2 \in S(nd)$ such that

$$\big(b_1, D_{\mathbf{x}}\varphi(\bar t, \bar{\mathbf{x}}, \bar{\mathbf{y}}) + \alpha Dh(\bar{\mathbf{x}}), S_1 + \alpha D^2 h(\bar{\mathbf{x}})\big) \in \overline{\mathcal{P}}^{2,+} u_n(\bar t, \bar{\mathbf{x}}),$$

$$\big(b_2, -D_{\mathbf{y}}\varphi(\bar t, \bar{\mathbf{x}}, \bar{\mathbf{y}}) - \alpha Dh(\bar{\mathbf{y}}), S_2 - \alpha D^2 h(\bar{\mathbf{y}})\big) \in \overline{\mathcal{P}}^{2,-} u_n(\bar t, \bar{\mathbf{y}}),$$

$$b_1 - b_2 = \varphi_\delta(\bar s) + \frac{\mu}{(T-\bar t)^2},$$

where

$$\begin{pmatrix} S_1 & 0 \\ 0 & -S_2 \end{pmatrix} \leq D + \frac{1}{\|D_1\| + \|D_2\|} D^2 \leq 2D_1, \tag{3.9}$$

where we used that

$$D^2 \leq (\|D_1\| + \|D_2\|)(D_1 + D_2).$$

Inequality (3.9) in particular implies that $S_1 \leq S_2$. Using the definition of viscosity solution and setting

$$Z(\bar{\mathbf{x}}, \bar{\mathbf{y}}) := \frac{1}{n}\sum_{i=1}^n H\left(\bar x_i,\ \frac{1}{n-1}\sum_{j\neq i}\delta_{\bar x_j},\ \varphi'_\delta(\bar s)\frac{(\bar x_i - \bar y_i)(\gamma + |\bar x_i - \bar y_i|^2)^{\frac{r}{2}-1}}{\bar s^{\frac{r}{r'}}}(1+\bar t) + n\alpha Dh(\bar{\mathbf{x}}) \right)$$

and

$$\tilde{Z}(\bar{\mathbf{x}},\bar{\mathbf{y}}) := \frac{1}{n}\sum_{i=1}^{n} H\left(\bar{y}_i,\ \frac{1}{n-1}\sum_{j\neq i}\delta_{\bar{y}_j},\ \varphi_\delta'(\bar{s})\frac{(\bar{x}_i-\bar{y}_i)(\gamma+|\bar{x}_i-\bar{y}_i|^2)^{\frac{r}{2}-1}}{\bar{s}^{\frac{r}{r'}}}(1+\bar{t}) - n\alpha Dh(\bar{\mathbf{y}})\right)$$

we now have

$$b_1 - \kappa \mathrm{Tr}(A_n(\bar{\mathbf{x}})(S_1+\alpha D^2h(\bar{\mathbf{x}}))) + Z(\bar{\mathbf{x}},\bar{\mathbf{y}}) + \mathcal{F}(\mu_{\bar{\mathbf{x}}}) \le 0 \tag{3.10}$$

and

$$b_2 - \kappa \mathrm{Tr}(A_n(\bar{\mathbf{y}})(S_2-\alpha D^2h(\bar{\mathbf{y}}))) + \tilde{Z}(\bar{\mathbf{x}},\bar{\mathbf{y}}) + \mathcal{F}(\mu_{\bar{\mathbf{y}}}) \ge 0. \tag{3.11}$$

We notice that (3.1) and (3.9) imply

$$\begin{aligned}
&\mathrm{Tr}(A_n(\bar{\mathbf{x}})S_1) - \mathrm{Tr}(A_n(\bar{\mathbf{y}})S_2)\\
&\quad\le 2\sum_{i=1}^{n}\mathrm{Tr}\left((a(x_i)-a(y_i))(a(x_i)-a(y_i))^*\varphi_\delta'(\bar{s})\frac{(\gamma+|\bar{x}_i-\bar{y}_i|^2)^{\frac{r}{2}-1}}{n\bar{s}^{\frac{r}{r'}}}(1+\bar{t})I_d\right)\\
&\quad\le \frac{1}{n}\sum_{i=1}^{n}\varphi_\delta'(\bar{s})\frac{2dL^2|\bar{x}_i-\bar{y}_i|^2(\gamma+|\bar{x}_i-\bar{y}_i|^2)^{\frac{r}{2}-1}}{\bar{s}^{\frac{r}{r'}}}(1+\bar{t}).
\end{aligned}$$

Note that if A_n is a constant matrix, then obviously $\mathrm{Tr}(A_nS_1)-\mathrm{Tr}(A_nS_2)\le 0$. Let us use the notation

$$\mu^i_{\mathbf{x}} := \frac{1}{n-1}\sum_{j\neq i}\delta_{x_j},\quad \mu^i_{\mathbf{y}} := \frac{1}{n-1}\sum_{j\neq i}\delta_{y_j}.$$

A simple calculation shows that $d_r(\mu^i_{\bar{\mathbf{x}}},\mu^i_{\bar{\mathbf{y}}})\le 2\bar{s}$. Subtracting (3.11) from (3.10) and using Hypothesis 2.1, the concavity of σ, and (2.10), we obtain

$$\begin{aligned}
\varphi_\delta(\bar{s})+\frac{\mu}{(T-\bar{t})^2} \le\ & \frac{\kappa}{n}\sum_{i=1}^{n}\varphi_\delta'(\bar{s})\frac{2dL^2|\bar{x}_i-\bar{y}_i|^2(\gamma+|\bar{x}_i-\bar{y}_i|^2)^{\frac{r}{2}-1}}{\bar{s}^{\frac{r}{r'}}}(1+\bar{t})\\
&+\frac{1}{n}\sum_{i=1}^{n}\sigma\left((|\bar{x}_i-\bar{y}_i|+d_r(\mu^i_{\bar{\mathbf{x}}},\mu^i_{\bar{\mathbf{y}}}))\Big(1+\varphi_\delta'(\bar{s})\frac{|\bar{x}_i-\bar{y}_i|(\gamma+|\bar{x}_i-\bar{y}_i|^2)^{\frac{r}{2}-1}}{\bar{s}^{\frac{r}{r'}}}(1+\bar{t})\Big)\right)\\
&+m_1(\bar{s})+\sigma_2(\alpha),
\end{aligned}$$

where $\lim_{\alpha\to 0}\sigma_2(\alpha)=0$. Thus,

$$\begin{aligned}
\varphi_\delta(\bar{s})+\frac{\mu}{(T-\bar{t})^2} \le\ & \frac{\kappa}{n}\sum_{i=1}^{n}\varphi_\delta'(\bar{s})\frac{2dL^2|\bar{x}_i-\bar{y}_i|^2(\gamma+|\bar{x}_i-\bar{y}_i|^2)^{\frac{r}{2}-1}}{\bar{s}^{\frac{r}{r'}}}(1+\bar{t})\\
&+\sigma\left(\frac{1}{n}\sum_{i=1}^{n}(|\bar{x}_i-\bar{y}_i|+2\bar{s})\left(1+\varphi_\delta'(\bar{s})\frac{|\bar{x}_i-\bar{y}_i|(\gamma+|\bar{x}_i-\bar{y}_i|^2)^{\frac{r}{2}-1}}{\bar{s}^{\frac{r}{r'}}}(1+\bar{t})\right)\right)\\
&+m_1(\bar{s})+\sigma_2(\alpha).
\end{aligned}$$

By Jensen's inequality

$$|\bar{\mathbf{x}}-\bar{\mathbf{y}}|_1 \le |\bar{\mathbf{x}}-\bar{\mathbf{y}}|_r \le \bar{s}$$

and also

$$\frac{1}{n}\sum_{i=1}^{n}\varphi_\delta'(\bar{s})\frac{|\bar{x}_i-\bar{y}_i|^2(\gamma+|\bar{x}_i-\bar{y}_i|^2)^{\frac{r}{2}-1}}{\bar{s}^{\frac{r}{r'}}} \le \varphi_\delta'(\bar{s})\bar{s}.$$

Using Schwarz's inequality, one verifies that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} 2\bar{s}\varphi_\delta'(\bar{s})\frac{|\bar{x}_i-\bar{y}_i|(\gamma+|\bar{x}_i-\bar{y}_i|^2)^{\frac{r}{2}-1}}{\bar{s}^{\frac{r}{r'}}} &\le \frac{1}{n}\sum_{i=1}^{n} 2\varphi_\delta'(\bar{s})\bar{s}^{1-r/r'}(\gamma+|\bar{x}_i-\bar{y}_i|^2)^{\frac{r-1}{2}}\\ &\le\ 2\varphi_\delta'(\bar{s})s^{1-\frac{r}{r'}}\left(\frac{1}{n}\sum_{i=1}^{n}(\gamma+|\bar{x}_i-\bar{y}_i|^2)^{\frac{r}{2}}\right)^{\frac{r-1}{r}}\\ &= 2\varphi_\delta'(\bar{s})s^{1-r/r'}\,\bar{s}^{r-1} = 2\varphi_\delta'(\bar{s})\bar{s}.\end{aligned}$$

Collecting these bounds into the inequality obtained for $\varphi_\delta(\bar{s})+\frac{\mu}{(T-t)^2}$ above, and using the subadditivity of σ and the definition of σ_1, we get

$$\begin{aligned}\varphi_\delta(\bar{s})+\frac{\mu}{T^2} &\le 2\kappa d L^2(1+T)\varphi_\delta'(\bar{s})\bar{s}+(1+T)\sigma(3\varphi_\delta'(\bar{s})\bar{s}+3\bar{s})+m_1(\bar{s})+\sigma_2(\alpha)\\ &\le \sigma_1(\varphi_\delta'(\bar{s})\bar{s}+\bar{s})+\sigma_2(\alpha).\end{aligned}$$

This gives a contradiction when we let $\alpha\to 0$, due to (3.2). Consequently, for all positive γ and μ, (3.7) is false. Letting $\gamma,\mu\to 0$, we obtain (3.6). Thus we have proved that for all $t\in[0,T]$, $(\mathbf{x},\mathbf{y})\in(\mathbb{R}^d)^n$,

$$|u_n(t,\mathbf{x})-u_n(t,\mathbf{y})|\le \inf_{\delta>0}\varphi_\delta(|\mathbf{x}-\mathbf{y}|_r)(1+T). \tag{3.12}$$

We will now obtain the continuity estimate with respect to t. We know by (2.9) that

$$|u_n(0,\mathbf{x})-u_n(0,\mathbf{y})|\le m_1\left(|\mathbf{x}-\mathbf{y}|_2\right).$$

Setting $v_n(\mathbf{x})=u_n(0,\sqrt{n}\mathbf{x})$ we thus have

$$|v_n(\mathbf{x})-v_n(\mathbf{y})|\le m_1\left(|\mathbf{x}-\mathbf{y}|\right).$$

Approximating the functions $v_n(\mathbf{x})$ by supinf-convolutions and then mollifying them, there exist constants $L_m, m=1,2,\cdots$ (independent of n), and C^2 functions $\tilde{\varphi}_m^n$ such that $0\le\tilde{\varphi}_m^n(\mathbf{x})-v_n(\mathbf{x})\le\frac{1}{m}$ on $(\mathbb{R}^d)^n$ and

$$|D\tilde{\varphi}_m^n|\le L_m,\quad \|D^2\tilde{\varphi}_m^n\|\le L_m.$$

Then if $\varphi_m^n(\mathbf{x})=\tilde{\varphi}_m^n(\frac{1}{\sqrt{n}}\mathbf{x})$, we have

$$0\le\varphi_m^n(\mathbf{x})-u_n(0,\mathbf{x})\le\frac{1}{m}\quad\text{on }(\mathbb{R}^d)^n \tag{3.13}$$

and

$$|D\varphi_m^n|\le\frac{L_m}{\sqrt{n}},\quad \|D^2\varphi_m^n\|\le\frac{L_m}{n}.$$

Recall that K is such that $\|\mathcal{F}\|_\infty\le K$. We set $C_m=\kappa\tilde{L}L_m+C+CL_m^2+K$, where $\tilde{L}$ is such that $\mathrm{Tr}(A_n(\mathbf{x}))\le\tilde{L}n$ and C is from Hypothesis 2.1. We define the functions

$$\psi_m^n(t,\mathbf{x}):=\varphi_m^n(\mathbf{x})+C_m t.$$

Then

$$\begin{aligned}
&\partial_t \psi_m^n(t,\mathbf{x}) - \kappa \mathrm{Tr}(A_n(\mathbf{x}) D^2 \psi_m^n(t,\mathbf{x})) + \frac{1}{n}\sum_{i=1}^n H(x_i, \mu_{\mathbf{x}}^i, nD_{x_i}\psi_m^n(t,\mathbf{x})) + \mathcal{F}(\mu_{\mathbf{x}}) \\
&\geq C_m - \frac{\kappa}{n}\mathrm{Tr}\left(A_n(\mathbf{x}) D^2 \tilde{\varphi}_m^n\left(\frac{1}{\sqrt{n}}\mathbf{x}\right)\right) + \frac{1}{n}\sum_{i=1}^n H\left(x_i, \mu_{\mathbf{x}}^i, \sqrt{n} D_{x_i}\tilde{\varphi}_m^n\left(\frac{1}{\sqrt{n}}\mathbf{x}\right)\right) - K \\
&\geq C_m - \frac{\kappa}{n}\mathrm{Tr}(A_n(\mathbf{x}))\left\| D^2 \tilde{\varphi}_m^n\left(\frac{1}{\sqrt{n}}\mathbf{x}\right)\right\| - \frac{1}{n}\sum_{i=1}^n C\left(1 + n\left|D_{x_i}\tilde{\varphi}_m^n\left(\frac{1}{\sqrt{n}}\mathbf{x}\right)\right|^2\right) - K \\
&\geq C_m - \kappa \tilde{L} L_m - C - C\sum_{i=1}^n \left|D_{x_i}\tilde{\varphi}_m^n\left(\frac{1}{\sqrt{n}}\mathbf{x}\right)\right|^2 - K \\
&\geq C_m - \kappa \tilde{L} L_m - C - C L_m^2 - K = 0.
\end{aligned}$$

Therefore the functions ψ_m^n are viscosity supersolutions of (3.3). Similarly, the functions

$$\tilde{\psi}_m^n(t,\mathbf{x}) := \varphi_m^n(\mathbf{x}) - \frac{1}{m} - C_m t$$

are viscosity subsolutions of (3.3). Therefore, by comparison, we have for every $n, m \geq 1$,

$$\varphi_m^n(\mathbf{x}) - \frac{1}{m} - C_m t \leq u_n(t,\mathbf{x}) \leq \varphi_m^n(\mathbf{x}) + C_m t.$$

Using (3.13), this implies that

$$\begin{aligned}
-\frac{1}{m} - C_m t \leq \varphi_m^n(\mathbf{x}) - \frac{1}{m} - C_m t - u_n(0,\mathbf{x}) &\leq u_n(t,\mathbf{x}) - u_n(0,\mathbf{x}) \\
&\leq \varphi_m^n(\mathbf{x}) + C_m t - u_n(0,\mathbf{x}) \\
&\leq C_m t + \frac{1}{m}.
\end{aligned}$$

Therefore we obtain

$$|u_n(t,\mathbf{x}) - u_n(0,\mathbf{x})| \leq \tilde{\rho}(t) := \inf\left\{\frac{1}{m} + C_m t : m = 1, 2, \cdots\right\}$$

for all $(t,\mathbf{x}) \in [0,T] \times (\mathbb{R}^d)^n, n = 1, 2, \cdots$. The function $\tilde{\rho}$ is independent of n. We then define for every $h \in (0,T)$ the functions

$$v_n^h(t,\mathbf{x}) = u_n(t+h,\mathbf{x}), \quad (t,\mathbf{x}) \in [0, T-h] \times (\mathbb{R}^d)^n.$$

The functions v_n^h are viscosity solutions of (3.3) on $(0, T-h) \times (\mathbb{R}^d)^n$ and

$$|v_n^h(0,\mathbf{x}) - u_n(0,\mathbf{x})| \leq \tilde{\rho}(h).$$

By comparison we thus obtain

$$|u_n(t+h,\mathbf{x}) - u_n(t,\mathbf{x})| = |v_n^h(t,\mathbf{x}) - u_n(t,\mathbf{x})| \leq \tilde{\rho}(h) \tag{3.14}$$

for $(t,\mathbf{x}) \in (0, T-h) \times (\mathbb{R}^d)^n$. We now let

$$\rho(s) = \tilde{\rho}(s) + \inf_{\delta > 0} \varphi_\delta(s)(1+T).$$

Combining (3.14) with (3.12), we obtain (3.4) for this modulus ρ, uniformly with respect to $n \in \mathbb{N}$. ∎

4. Proof of Theorem 1.2.

Proof of Theorem 1.2. *Step* 1. We define the functions

$$\mathcal{V}_n(t,\mu_{\mathbf{x}}) := u_n(t,\mathbf{x}), \quad \text{where} \quad \mu_{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}.$$

This function is well defined since the functions u_n are invariant with respect to permutations of the variables in $\mathbf{x}$. This follows from uniqueness of viscosity solutions of (1.2) as these equations are invariant with respect to permutations of the variables in $\mathbf{x}$. The function $\mathcal{V}_n(t,\cdot)$ is now defined on the subset of $\mathcal{P}_2(\mathbb{R}^d)$ that consists of averages of n Dirac point masses. It follows from (3.4) that

$$|\mathcal{V}_n(t,\mu_{\mathbf{x}}) - \mathcal{V}_n(s,\mu_{\mathbf{y}})| \le \rho(|t-s| + d_r(\mu_{\mathbf{x}},\mu_{\mathbf{y}}))\ \forall\ t,s\in[0,T], \mathbf{x},\mathbf{y}\in(\mathbb{R}^d)^n.$$

For each fixed n, we can extend the function $\mathcal{V}_n$ to a function on $[0,T]\times\mathcal{P}_r(\mathbb{R}^d)$, still denoted by $\mathcal{V}_n$, satisfying

$$|\mathcal{V}_n(t,\mu) - \mathcal{V}_n(s,\nu)| \le \rho(|t-s| + d_r(\mu,\nu))\ \forall\ t,s\in[0,T], \mu,\nu\in\mathcal{P}_r(\mathbb{R}^d). \tag{4.1}$$

Since sets

$$M_R^2 = \left\{\mu\in\mathcal{P}_r(\mathbb{R}^d) : \int_{\mathbb{R}^d}|x|^2\mu(dx)\le R\right\}$$

are relatively compact in $\mathcal{P}_r(\mathbb{R}^d)$, up to a subsequence (still denoted by $\mathcal{V}_n$), $\mathcal{V}_n$ converges uniformly on every set $[0,T]\times M_R^2$ to a function $\mathcal{V}:[0,T]\times\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}$ which satisfies the same estimate (4.1). Define now

$$\begin{aligned} V:[0,T]\times E &\longrightarrow \mathbb{R},\\ (t,X) &\longmapsto V(t,X) := \mathcal{V}(t,\mathrm{law}(X)).\end{aligned}$$

We will show that V is a viscosity solution of (1.3). Since (1.3) has a unique bounded viscosity solution U, we can then conclude that $V=U$. This will prove the theorem since the argument can be done for any subsequence of $\mathcal{V}_n$. The proof that (1.3) has a unique bounded viscosity solution is included in the appendix, Theorem 7.4.

Step 2. Let then $\varphi\in C^{1,2}((0,T)\times E)$ and suppose that $V-\varphi$ has a local maximum at $(t_0,X_0)\in(0,T)\times E$. By considering $\varphi(t,X)+(t-t_0)^2+|X-X_0|^2$ and modifying it outside of a neighborhood of (t_0,X_0) we can assume with no loss of generality that the maximum at (t_0,X_0) is strict and global. Being a strict maximum means that if $V(t_i,X_i)-\varphi(t_i,X_i)\to V(t_0,X_0)-\varphi(t_0,X_0)$, then $(t_i,X_i)\to(t_0,X_0)$. Denote $P_0=D\varphi(t_0,X_0)$. For each $\epsilon>0$ let $X_\epsilon,P_\epsilon\in E$ be such that X_ϵ,P_ϵ are continuous on $[0,1]$ and $|X_0-X_\epsilon|+|P_0-P_\epsilon|<\epsilon$.

For every n we denote $A_i^n=(\frac{i-1}{n},\frac{i}{n}), i=1,\dots,n$. We then consider the function $\varphi_n:(0,T)\times(\mathbb{R}^d)^n\to\mathbb{R}$ defined by

$$\varphi_n(t,\mathbf{x}) := \varphi\left(t,\sum_{i=1}^{n}x_i\mathbf{1}_{A_i^n}\right),$$

where $\mathbf{1}_{A_i^n}$ is the characteristic function of the set A_i^n.

Since the original maximum at (t_0,X_0) was strict it is easy to see that the functions $u_n-\varphi_n$ must have local maxima at points $(t_n,\mathbf{x}^n)=(t_n,x_1^n,\dots,x_n^n)$ such that

$$t_n\to t_0 \quad\text{and}\quad X^n=\sum_{i=1}^{n}x_i^n\mathbf{1}_{A_i^n}\to X_0.$$

In particular for sufficiently big n,

$$|t_n - t_0| + |X_0 - X^n| \leq \epsilon.$$

Now, by the chain rule,

$$D_{x_i}\varphi_n(t_n, \mathbf{x}^n) = \left\langle D\varphi\left(t_n, \sum_{i=1}^n x_i^n \mathbf{1}_{A_i^n}\right), \mathbf{1}_{A_i^n} e\right\rangle_d = \int_{A_i^n} D\varphi\left(t_n, \sum_{i=1}^n x_i^n \mathbf{1}_{A_i^n}\right),$$

where $e = (1, \dots, 1) \in \mathbb{R}^d$. If $\mathbf{x} = (x_1, \dots, x_n)$, we will denote $x_i = (x_{i1}, \dots, x_{id})$. Then

$$\operatorname{Tr}(A_n D^2\varphi_n) = \sum_{i,j=1}^n \operatorname{Tr}(D^2_{x_i x_j}\varphi_n) = \sum_{i,j=1}^n \sum_{k=1}^d \frac{\partial^2 \varphi_n}{\partial x_{ik}\partial x_{jk}}.$$

Now

$$\frac{\partial^2 \varphi_n}{\partial x_{ik}\partial x_{jk}}(t_n, \mathbf{x}^n) = \int_{A_i^n} \left(D^2\varphi\left(t_n, \sum_{i=1}^n x_i^n \mathbf{1}_{A_i^n}\right)\mathbf{1}_{A_j^n} e_k\right) \cdot e_k$$

so

$$\begin{aligned}\operatorname{Tr}(A_n D^2\varphi_n) &= \sum_{i,j=1}^n \sum_{k=1}^d \int_{A_i^n} \left(D^2\varphi\left(t_n, \sum_{i=1}^n x_i^n \mathbf{1}_{A_i^n}\right)\mathbf{1}_{A_j^n} e_k\right) \cdot e_k \\ &= \sum_{k=1}^d \left\langle D^2\varphi\left(t_n, \sum_{i=1}^n x_i^n \mathbf{1}_{A_i^n}\right) e_k, e_k\right\rangle.\end{aligned}$$

We note that if

$$\mathbf{x}, \mathbf{y} \in (\mathbb{R}^d)^n, \quad X = \sum_{i=1}^n x_i \mathbf{1}_{A_i^n} \quad \text{and} \quad Y = \sum_{i=1}^n y_i \mathbf{1}_{A_i^n},$$

then

$$|X - Y|^2 = n^{-1}\sum_{i=1}^n |x_i - y_i|^2, \qquad |X - Y|_r^r = n^{-1}\sum_{i=1}^n |x_i - y_i|^r.$$

Furthermore, choosing $\xi, \eta \in E$, we have

$$\begin{aligned}\left|\sum_{i=1}^n H\left(x_i, \mu_{\mathbf{x}}^i, n\int_{A_i^n}\xi\right) - \sum_{i=1}^n H\left(y_i, \mu_{\mathbf{y}}^i, n\int_{A_i^n}\eta\right)\right| &\leq \sum_{i=1}^n \left|H\left(x_i, \mu_{\mathbf{x}}^i, n\int_{A_i^n}\xi\right) - H\left(y_i, \mu_{\mathbf{y}}^i, n\int_{A_i^n}\eta\right)\right| \\ &\leq C\sum_{i=1}^n \left(1 + \left|n\int_{A_i^n}\xi\right| + \left|n\int_{A_i^n}\eta\right|\right)\int_{A_i^n} n|\xi - \eta| \\ &\quad + \sum_{i=1}^n \sigma\left((|x_i - y_i| + d_r(\mu_{\mathbf{x}}^i, \mu_{\mathbf{y}}^i))\left(1 + \left|n\int_{A_i^n}\xi\right|\right)\right).\end{aligned}$$

Thus,

$$\begin{aligned}&n^{-1}\left|\sum_{i=1}^n H\left(x_i, \mu_{\mathbf{x}}^i, n\int_{A_i^n}\xi\right) - \sum_{i=1}^n H\left(y_i, \mu_{\mathbf{y}}^i, n\int_{A_i^n}\eta\right)\right| \\ &\quad \leq C\int_\Omega |\xi - \eta| + C\left(\left(\sum_{i=1}^n n\left|\int_{A_i^n}\xi\right|^2\right)^{\frac{1}{2}} + \left(\sum_{i=1}^n n\left|\int_{A_i^n}\eta\right|^2\right)^{\frac{1}{2}}\right)\left(\sum_{i=1}^n n\left|\int_{A_i^n}|\xi - \eta|\right|^2\right)^{\frac{1}{2}} \\ &\quad + \sigma\left(\frac{1}{n}\sum_{i=1}^n \left(|x_i - y_i| + \left(\frac{n}{n-1}\right)^{\frac{1}{r}}|X - Y|_r\right)\left(1 + \left|n\int_{A_i^n}\xi\right|\right)\right).\end{aligned}$$

We conclude

$$n^{-1}\left|\sum_{i=1}^{n} H\Big(x_i,\mu^i_{\mathbf{x}},n\int_{A^n_i}\xi\Big)-\sum_{i=1}^{n} H\Big(y_i,\mu^i_{\mathbf{y}},n\int_{A^n_i}\eta\Big)\right|$$
$$\le C\left(\int_\Omega|\xi-\eta|^2\right)^{\frac12}+C\left(\left(\sum_{i=1}^n\int_{A^n_i}|\xi|^2\right)^{\frac12}+\left(\sum_{i=1}^n\int_{A^n_i}|\eta|^2\right)^{\frac12}\right)\left(\sum_{i=1}^n\int_{A^n_i}|\xi-\eta|^2\right)^{\frac12}$$
$$+\sigma\left(\left(\left(\frac1n\sum_{i=1}^n|x_i-y_i|^2\right)^{\frac12}+2|X-Y|_r\right)\left(1+\left(\sum_{i=1}^n n\left|\int_{A^n_i}\eta\right|^2\right)^{\frac12}\right)\right).$$

Finally, we have

(4.2)
$$n^{-1}\left|\sum_{i=1}^{n} H\Big(x_i,\mu^i_{\mathbf{x}},n\int_{A^n_i}\xi\Big)\sum_{i=1}^{n} H\Big(y_i,\mu^i_{\mathbf{y}},n\int_{A^n_i}\eta\Big)\right|\le C\Big(1+|\xi|+|\eta|\Big)|\xi-\eta|+\sigma\Big(3|X-Y|\,(1+|\xi|)\Big).$$

Set

$$X^n_\epsilon=\sum_{i=1}^n X_\epsilon\left(\frac in\right)\mathbf{1}_{A^n_i},\quad \mathbf{x}^n_\epsilon=(X_\epsilon(1/n),X_\epsilon(2/n),\dots,X_\epsilon(1)).$$

Using the definition of viscosity subsolution and (4.2) we now have, for every $\epsilon>0$,

$$0\ge\partial_t\varphi(t_n,X^n)-\kappa\sum_{k=1}^d\left\langle D^2\varphi\left(t_n,\sum_{i=1}^n x^n_i\mathbf{1}_{A^n_i}\right)e_k,e_k\right\rangle+\frac1n\sum_{i=1}^n H\left(x^n_i,\mu^i_{\mathbf{x}^n},n\int_{A^n_i}D\varphi(t_n,X^n)\right)+\mathcal{F}(\mu_{\mathbf{x}^n}).$$

Hence, for large n,

$$\begin{aligned}0&\ge\partial_t\varphi(t_0,X_0)-\kappa\sum_{k=1}^d\langle D^2\varphi(t_0,X_0)e_k,e_k\rangle+\frac1n\sum_{i=1}^n H\left(X_\epsilon\left(\frac in\right),\mu^i_{\mathbf{x}^n_\epsilon},n\int_{A^n_i}P_\epsilon\right)+F(X^n)\\&\quad-\rho_0(n)-C(1+|D\varphi(t_n,X^n)|+|P_\epsilon|)|D\varphi(t_n,X^n)-P_\epsilon|-\sigma\left(3|X^n_\epsilon-X^n|\,(1+|P_\epsilon|)\right)\\&\ge\partial_t\varphi(t_0,X_0)-\kappa\sum_{k=1}^d\langle D^2\varphi(t_0,X_0)e_k,e_k\rangle+\frac1n\sum_{i=1}^n H\left(X_\epsilon\left(\frac in\right),\mu^i_{\mathbf{x}^n_\epsilon},n\int_{A^n_i}P_\epsilon\right)+F(X_0)\\&\quad-m_1(\epsilon)-\rho_0(n)-C(1+2|P_0|+2\epsilon)2\epsilon-\sigma\left((3|X^n_\epsilon-X_\epsilon|+3\epsilon)\,|\,(1+|P_0|+\epsilon)\right),\end{aligned}$$

where $\lim_{n\to\infty}\rho_0(n)=0$. Since X_ϵ,P_ϵ are continuous on Ω, it follows that

$$\lim_{n\to\infty}\frac1n\sum_{i=1}^n H\left(X_\epsilon\left(\frac in\right),\mu^i_{\mathbf{x}^n_\epsilon},n\int_{A^n_i}P_\epsilon\right)=\tilde H(X_\epsilon,\mathrm{law}(X_\epsilon),P_\epsilon).$$

Thus, letting $n\to\infty$ above we obtain

$$\begin{aligned}&\partial_t\varphi(t_0,X_0)-\kappa\sum_{k=1}^d\langle D^2\varphi(t_0,X_0)e_k,e_k\rangle+\tilde H(X_\epsilon,\mathrm{law}(X_\epsilon),P_\epsilon)\\&\le m_1(\epsilon)+C(1+2|P_0|+2\epsilon)2\epsilon+2\sigma\left(3\epsilon\,(1+|P_0|+\epsilon)\right).\end{aligned}$$

Finally, letting $\epsilon\to0$ we conclude that

$$\partial_t\varphi(t_0,X_0)-\kappa\sum_{k=1}^d\langle D^2\varphi(t_0,X_0)e_k,e_k\rangle+\tilde H(X_0,\mathrm{law}(X_0),D\varphi(t_0,X_0))+F(X_0)\le0.$$

Thus V is a viscosity subsolution of (1.3). Reasoning in the same way, we can prove that V is a supersolution of (1.3). ∎

EXAMPLE 4.1. *The following is an example of a typical particle system that leads to simple equations of type* (1.2). *Let* $G : \mathbb{R}^d \to \mathbb{R}^d$ *be a bounded even function such that*

$$|G(x) - G(y)| \leq L|x - y| \quad \forall x, y \in \mathbb{R}^d$$

and let W *be a standard Wiener process in* $\mathbb{R}^d$. *Let* $T > 0$. *We consider a system of* n *particles whose states are given by the SDE with common noise*

$$\begin{cases} dX_i(s) = \frac{1}{n-1} \sum_{j \neq i} G(X_i(s) - X_j(s))ds + \sqrt{2\kappa} dW(s), \quad t \leq s \leq T, \\ X_i(t) = x_i. \end{cases}$$

If we define

$$u_n(t, \mathbf{x}) = \mathbb{E}\left[- \int_t^T \mathcal{F}\left(\frac{1}{n} \sum_{i=1}^n \delta_{X_i(s)} \right) ds + \mathcal{U}_0\left(\frac{1}{n} \sum_{i=1}^n \delta_{X_i(T)} \right) \right]$$

(where $\mathbb{E}$ *above is the expectation with respect to a probability measure on some reference probability space), then the function* u_n *is the viscosity solution of the terminal value problem*

$$\begin{cases} -\partial_t u_n - \kappa \mathrm{Tr}(A_n D^2 u_n) - \frac{1}{n-1} \sum_{i=1}^n \sum_{j \neq i} G(x_i - x_j) \cdot D_{x_i} u_n \\ \qquad + \mathcal{F}(\frac{1}{n} \sum_{i=1}^n \delta_{x_i}) = 0 \qquad \text{in } (0,T) \times (\mathbb{R}^d)^n, \\ u_n(T, x_1, \ldots, x_n) = \mathcal{U}_0(\frac{1}{n} \sum_{i=1}^n \delta_{x_i}) \qquad \text{on } (\mathbb{R}^d)^n, \end{cases}$$

where A_n *is as in* (1.2). *In this example the Hamiltonian* H *is defined by*

$$H(x, \nu, p) = -p \cdot \int_{\mathbb{R}^d} G(x - y)\nu(dy).$$

It is obvious that H *satisfies* (2.1) *and* (2.3). *We point out that the boundedness of* G *is needed here to guarantee* (2.3). *Regarding* (2.2), *let* $x, y, p \in \mathbb{R}^d, \mu, \nu \in \mathcal{P}_r(\mathbb{R}^d)$, *and let* γ *be a Borel probability measure on* $\mathbb{R}^d \times \mathbb{R}^d$ *with marginals* μ, ν. *Then*

$$\begin{aligned} |H(x, \mu, p) - H(y, \nu, p)| &\leq L|x - y||p| + \left| \int_{\mathbb{R}^d} G(x - z)\mu(dy) \cdot p - \int_{\mathbb{R}^d} G(x - w)\nu(dw) \cdot p \right| \\ &\leq L|x - y||p| + \left| \int_{\mathbb{R}^d} (G(x - z) - G(x - w))\gamma(dz, dw) \right| |p| \\ &\leq L|x - y||p| + L|p| \int_{\mathbb{R}^d} |z - w|\gamma(dz, dw) \\ &\leq L|p| \left(|x - y| + \left(\int_{\mathbb{R}^d} |z - w|^r \gamma(dz, dw) \right)^{\frac{1}{r}} \right). \end{aligned}$$

Since this holds for every γ *we thus obtain*

$$|H(x, \mu, p) - H(y, \nu, p)| \leq L|p| \left(|x - y| + d_r(\mu, \nu) \right).$$

5. *L*-viscosity solutions versus viscosity solutions on the Wasserstein space. In this section, we consider either

$$\mathcal{U} : [0, T) \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \quad \text{or} \quad \mathcal{U} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$$

and

$$U : [0, T) \times E \to \mathbb{R} \quad \text{or} \quad U : E \to \mathbb{R}$$

such that $U(t,X) = \mathcal{U}(t, X_\sharp \mathcal{L}_1)$ or $U(X) = \mathcal{U}(X_\sharp \mathcal{L}_1)$. Recall

$$E := L^2((0,1), \mathbb{R}^d).$$

When $\mathcal{U}$ is differentiable at $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $\nabla_\mu \mathcal{U}(\mu)$, the Wasserstein gradient of $\mathcal{U}$ at μ satisfies

$$\nabla_\mu \mathcal{U}(\mu) \in T_\mu \mathcal{P}_2(\mathbb{R}^d). \tag{5.1}$$

Assume next that $\nabla_\mu \mathcal{U}$ is differentiable at (q,μ) in the sense of [37] and $\nabla_{\mu\mu}\mathcal{U}(\mu)(q,\cdot)$, the Wasserstein gradient of $\nabla_\mu \mathcal{U}(\mu)$ at μ belongs to $L^\infty_{\mu\otimes\mu}(\mathbb{R}^{2d}, \mathbb{R}^{d\times d})$. We have

$$\big(\nabla_{\mu\mu}\mathcal{U}(\mu)\big)^T(q,x) = \big(\nabla_{\mu\mu}\mathcal{U}(\mu)\big)(x,q) \qquad \forall q, x \in \mathbb{R}^d. \tag{5.2}$$

If we assume that $\mathcal{U}$ is twice differentiable in the sense of [37], then the map $(q,\mu) \mapsto \nabla_\mu \mathcal{U}(\mu)(q)$ has a first order Taylor expansion on an appropriate set [24]. Furthermore, $\nabla_\mu \mathcal{U}(\mu)$ is Lipschitz on the support of μ and there exists a symmetric matrix $A_1(\mu) \in L^\infty_\mu(\mathbb{R}^d, \mathbb{R}^{d\times d})$ which coincides almost everywhere with $\nabla_q\big(\nabla_\mu \mathcal{U}(\mu)\big)$. In this manuscript, we adopt the notation and terminology of [37] by defining $\nabla^2_w \mathcal{U}(\mu)$, the Wasserstein second differential of $\mathcal{U}$ at μ, as

$$\nabla^2_w \mathcal{U}(\mu)(\xi,\xi_*) = \int_{\mathbb{R}^d} A_1(\mu)(q)\xi(q)\cdot\xi_*(q)\mu(dq) + \int_{\mathbb{R}^{2d}} \nabla_{\mu\mu}\mathcal{U}(\mu)(q_*,q)\xi(q)\cdot\xi_*(q_*)\mu(dq)\mu(dq_*)$$

if $\xi, \xi_* \in L^2_\mu(\mathbb{R}^d, \mathbb{R}^d)$. Note that the ordering (q_*, q) in the last integral is not a typo. By abuse of notation, we identify the bilinear forms $\nabla^2_w \mathcal{U}(\mu)$ with the operators

$$\xi \longmapsto A_1(\mu)\xi + \int_{\mathbb{R}^d} \nabla_{\mu\mu}\mathcal{U}(\mu)(\cdot,q)\xi(q)\mu(dq)$$

which, by (5.2) and the fact that A_1 is symmetric μ-almost everywhere, turns out to be self-adjoint on $L^2_\mu(\mathbb{R}^d, \mathbb{R}^d)$.

The relation $U(X) = \mathcal{U}(X_\sharp \mathcal{L}_1)$ expresses the fact that U is invariant under the set of maps which preserve Lebesgue measure. This is what imposes a special structure on the second differential of U at X when it exists. When U is twice differentiable at X then for any $\zeta \in E$, $D^2U(X)(\zeta)(\cdot)$ belongs to E and

$$D^2U(X)(\zeta)(\cdot) = A_1(X_\sharp \mathcal{L}_1)(X)\ \zeta + \int_{(0,1)} \nabla_{\mu\mu}\mathcal{U}(X_\sharp \mathcal{L}_1)\big(X, X(\omega)\big)\zeta(\omega)d\omega.$$

Given an arbitrary orthonormal basis $\{e_1, \dots, e_d\}$ of $\mathbb{R}^d$, we identify each e_k with the constant function which assumes the value e_k everywhere. Abusing notation we write $e_k : \mathbb{R}^d \to \mathbb{R}^d$. Note that if $X \in E$, then $e_k \circ X \equiv e_k$, and so we may also consider e_k to be the constant function $e_k : (0,1) \to \mathbb{R}^d$. If E_0 is the finite dimensional space spanned by $\{e_1, \dots, e_d\}$, we have the orthogonal decomposition

$$E = E_0 \oplus E_0^\perp.$$

The partial trace of the operator $\zeta \to D^2U(X)(\zeta)$ on E_0 is

$$\triangle_{\mathbb{R}^d} U(X) = \sum_{k=1}^d \langle D^2U(X)e_k, e_k\rangle.$$

We have the relation

$$\triangle_{\mathbb{R}^d} U(X) = \Delta_w \mathcal{U}(\mu),$$

where Δ_w is the partial Wasserstein Laplacian [24]. This relation will allow us to compare viscosity solutions on the Wasserstein space and the Hilbert space.

5.1. Domain of definition of Wasserstein Hessian. In this section, we denote by $\pi^1, \pi^2 : \mathbb{R}^{2d} \to \mathbb{R}^d$ the coordinate projection maps of $\mathbb{R}^{2d}$ onto $\mathbb{R}^d$. Given a positive integer D, we denote by $C_b^3(\mathbb{R}^D)$ the set of $f \in C^3(\mathbb{R}^D)$ that have bounded second and third order derivatives.

We start by recalling a few facts about the Wasserstein tangent spaces $T_\mu\mathcal{P}_2(\mathbb{R}^d)$. Let $\xi_0 \in L^2_\mu(\mathbb{R}^d, \mathbb{R}^d)$. Note that $\xi_0 \in T_\mu\mathcal{P}_2(\mathbb{R}^d)$ if and only if

$$\lim_{r\to 0^+} \inf_{\phi} \sup_{(\nu,\gamma)} \left\{ \frac{\left| \int_{\mathbb{R}^{2d}} \left(\xi_0(q_1) - \nabla\phi(q_1)\right) \cdot (q_2 - q_1)\gamma(dq_1, dq_2)\right|}{\|\pi^2 - \pi^1\|_{L^2(\gamma)}} : 0 < \|\pi^2 - \pi^1\|_{L^2(\gamma)} \le r, \gamma \in \Gamma(\mu,\nu) \right\} = 0.$$

Here, the infimum is performed over the set $C_c^\infty(\mathbb{R}^d)$ or equivalently over the set $C_b^3(\mathbb{R}^d)$. The space $T_\mu\mathcal{P}_2(\mathbb{R}^d)$, being a Hilbert space, can be identified with the co–tangent space. The Wasserstein gradient of a function $\mathcal{U} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ at μ is the element of minimal norm in the subdifferential of $\mathcal{U}$ at μ and so it belongs to $T_\mu\mathcal{P}_2(\mathbb{R}^d)$.

We would like to propose an analogous characterization for all Wasserstein derivatives of order less than or equal to 2. We say that $f : \mathbb{R}^{2d} \to \mathbb{R}$ is symmetric if $f(q_1, q_2) = f(q_2, q_1)$ for all $q_1, q_2 \in \mathbb{R}^d$.

Given $\psi \in C_b^3(\mathbb{R}^{2d})$ which is symmetric and $\phi \in C_b^3(\mathbb{R}^d)$, we define

$$\mathcal{V}^\mu_{(\phi,\psi)}(\nu) := \int_{\mathbb{R}^d} \phi(q)(\nu - \mu)(dq) + \frac{1}{2}\int_{\mathbb{R}^{2d}} \psi(q_1, q_2)(\nu - \mu)(dq_1)(\nu - \mu)(dq_2)$$

for $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. The function $\mathcal{V}^\mu_{(\phi,\psi)}$ is twice differentiable in the sense of [37],

$$\nabla_\mu \mathcal{V}^\mu_{(\phi,\psi)}(\nu)(q_1) = \nabla\phi(q_1) + \int_{\mathbb{R}^d} \nabla_{q_1}\psi(q_1, b)(\nu - \mu)(db),$$

and so

$$\nabla_{q_1}\Big(\nabla_\mu \mathcal{V}^\mu_{(\phi,\psi)}(\nu)(q_1)\Big) = \nabla^2\phi(q_1) + \int_{\mathbb{R}^d} \nabla_{q_1q_1}\psi(q_1, b)(\nu - \mu)(db).$$

We conclude

$$\nabla_{\mu\mu}\mathcal{V}^\mu_{(\phi,\psi)}(\nu)(q) = \nabla_{q_2q_1}\psi(q_1, q_2).$$

Note that if $X, Y \in E$ are such that $X_\sharp\mathcal{L}^1_{(0,1)} = \mu$ and $Y_\sharp\mathcal{L}^1_{(0,1)} = \nu$, then

$$\mathcal{V}^\mu_{(\phi,\psi)}(\nu) = V^X_{(\phi,\psi)}(Y), \tag{5.3}$$

where we have set

$$V^X_{(\phi,\psi)}(Y) = \int_{(0,1)} \Big(\phi(Y(\omega)) - \phi(X(\omega))\Big)d\omega + \frac{1}{2}\int_{(0,1)^2} \Big(\psi\big(Y(\omega), Y(o)\big) + \psi\big(X(\omega), X(o)\big) - 2\psi\big(Y(\omega), X(o)\big)\Big)d\omega do.$$

Remark 5.1. Let $X \in E$, let $\phi \in C_b^3(\mathbb{R}^d)$, let $\psi \in C_b^3(\mathbb{R}^{2d})$ be symmetric, and set $V := V^X_{(\phi,\psi)}$.

(i) Note that if $X^*, Y, Y^* \in E$ are such that X and X^* have the same law and Y and Y^* have the same law, then (5.3) implies $V^X_{(\phi,\psi)}(Y) = V^{X^*}_{(\phi,\psi)}(Y^*)$.

(ii) The function V is Fréchet differentiable everywhere on E and for any $Y \in E$ we have

$$DV(Y)(\omega) = \nabla\phi\big(Y(\omega)\big) + \int_{(0,1)} \Big(\nabla_{q_1}\psi\big(Y(\omega), Y(o)\big) - \nabla_{q_1}\psi\big(Y(\omega), X(o)\big)\Big) do.$$

(iii) The subset of E where the function DV is Fréchet differentiable may not be E (cf. [37]) unless ϕ and ψ are polynomials of degree 2. However, DV is Gateaux differentiable everywhere on E and for any $Y \in E$ we have

$$\begin{aligned} D^2V(Y)(\zeta,\zeta) = &\int_{(0,1)} \nabla^2\phi(X)\zeta\cdot\zeta d\omega + \int_{(0,1)^2} \nabla_{q_2q_1}\psi\big(Y(\omega), Y(o)\big)\zeta(\omega)\cdot\zeta(o) d\omega do \\ &+ \int_{(0,1)^2} \Big(\nabla_{q_1q_1}\psi\big(Y(\omega), Y(o)\big) - \nabla_{q_1q_1}\psi\big(Y(\omega), X(o)\big)\Big)\zeta(\omega)\cdot\zeta(\omega) d\omega do \end{aligned}$$

for any $\zeta \in E$. In particular, $DV(X) = \nabla\phi \circ X$ and the operator $D^2V(X)$ is given by

$$D^2V(X)\,\zeta = \nabla^2\phi(X)\,\zeta + \int_{(0,1)} \nabla_{q_2q_1}\psi\big(X, X(\omega)\big)\zeta(\omega) d\omega$$

for $\zeta \in E$.

LEMMA 5.2. *Let $X, X^* \in E$, let $\phi \in C_b^3(\mathbb{R}^d)$, and let $\psi \in C_b^3(\mathbb{R}^{2d})$ be symmetric. If $X_\sharp\mathcal{L}^1_{(0,1)} = X^*_\sharp\mathcal{L}^1_{(0,1)}$, then $V^X := V^X_{(\phi,\psi)}$ is twice Fréchet differentiable at X if and only if $V^{X^*} := V^{X^*}_{(\phi,\psi)}$ is twice Fréchet differentiable at X^*.*

Proof. We only need to prove one direction of the lemma since the converse direction could be obtained by symmetry. Assume V^X is twice Fréchet differentiable at X. By assumption, there exists a function $\epsilon : \mathbb{R} \to \mathbb{R}$, monotone nondecreasing, continuous at 0, and such that $\epsilon(0) = 0$ and there exists a function $\epsilon_0 : E \to \mathbb{R}$ such that $|\epsilon_0(h)| \le \epsilon(|h|)$ and

$$\begin{aligned} V^X(X+h) = &\int_{(0,1)} \nabla\phi(X)\cdot h d\omega + \frac{1}{2}\int_{(0,1)} \nabla^2\phi(X)h\cdot h d\omega \\ &+ \frac{1}{2}\int_{(0,1)^2} \nabla_{q_2q_1}\psi\big(X(o), X(\omega)\big)h(\omega)\cdot h(o) d\omega do + |h|^2\epsilon_0(h). \end{aligned} \tag{5.4}$$

Since X and X^* have the same laws, it is well known that we can choose a sequence of Borel functions $S_n : [0,1] \to [0,1]$ which are one-to-one, onto, such that $S_n\,\sharp\mathcal{L}^1_{(0,1)} = \mathcal{L}^1_{(0,1)}$ and such that

$$\lim_n |X^* - X\circ S_n| = 0.$$

In light of Remark 5.1(i), (5.4) implies

$$\begin{aligned} &V^{X\circ S_n}(X\circ S_n + h) = V^X(X + h\circ S_n^{-1}) \\ &= \int_{(0,1)} \nabla\phi(X)\cdot h\circ S_n^{-1} d\omega + \frac{1}{2}\int_{(0,1)} \nabla^2\phi(X)h\circ S_n^{-1}\cdot h\circ S_n^{-1} d\omega \\ &\quad + \frac{1}{2}\int_{(0,1)^2} \nabla_{q_2q_1}\psi\big(X(o), X(\omega)\big)h\circ S_n^{-1}(\omega)\cdot h\circ S_n^{-1}(o) d\omega do + |h\circ S_n^{-1}|^2\epsilon_0(h\circ S_n^{-1}) \\ &= \int_{(0,1)} \nabla\phi(X\circ S_n)\cdot h d\omega + \frac{1}{2}\int_{(0,1)} \nabla^2\phi(X\circ S_n)h\cdot h d\omega \\ &\quad + \frac{1}{2}\int_{(0,1)^2} \nabla_{q_2q_1}\psi\big(X\circ S_n(o), X\circ S_n(\omega)\big)h(\omega)\cdot h(o) d\omega do + |h|^2\epsilon_0(h\circ S_n^{-1}). \end{aligned}$$

We use again Remark 5.1(i) to conclude that

$$V^{X^*}(X\circ S_n+h)=\int_{(0,1)}\nabla\phi(X\circ S_n)\cdot h d\omega+\frac{1}{2}\int_{(0,1)}\nabla^2\phi(X\circ S_n)h\cdot h d\omega$$
$$+\frac{1}{2}\int_{(0,1)^2}\nabla_{q_2q_1}\psi\big(X\circ S_n(o),X\circ S_n(\omega)\big)h(\omega)\cdot h(o)d\omega do+|h|^2\epsilon_0(h\circ S_n^{-1}).$$

We let n tend to ∞ to obtain

$$V^{X^*}(X^*+h)=\int_{(0,1)}\nabla\phi(X^*)\cdot h d\omega+\frac{1}{2}\int_{(0,1)}\nabla^2\phi(X^*)h\cdot h d\omega$$
$$+\frac{1}{2}\int_{(0,1)^2}\nabla_{q_2q_1}\psi\big(X^*(o),X^*(\omega)\big)h(\omega)\cdot h(o)d\omega do+|h|^2\lim_{n\to\infty}\epsilon_0(h\circ S_n^{-1}).$$

We use the fact that $\lim_{n\to\infty}|\epsilon_0(h\circ S_n^{-1})|\le\epsilon(|h|)$ to conclude that V^{X^*} is twice Fréchet differentiable at X^*. ∎

DEFINITION 5.3. *Let $\mu\in\mathcal{P}_2(\mathbb{R}^d)$, $\xi_0\in T_\mu\mathcal{P}_2(\mathbb{R}^d)$, let $A_1\in L^\infty_\mu(\mathbb{R}^d,\mathbb{R}^{d\times d})$ be symmetric μ–almost everywhere, and let $A_2\in L^\infty_{\mu\otimes\mu}(\mathbb{R}^{2d},\mathbb{R}^{d\times d})$ be such that $A_2(q_1,q_2)=A_2^T(q_2,q_1)$, $\mu\otimes\mu$–almost everywhere. We say that (ξ_0,A_1,A_2) belongs to $T^{*,2}_\mu\mathcal{P}_2(\mathbb{R}^d)$ if*

$$\lim_{r\to0^+}\inf_{(\phi,\psi)}\sup_{(\nu,\gamma)}\left\{\frac{|e_0(\mu,\gamma,\xi_0-\nabla\phi,A_1-\nabla^2\phi,A_2-\nabla_{q_2q_1}\psi)|}{\|\pi^2-\pi^1\|^2_{L^2(\gamma)}}\ :\ 0<\|\pi^2-\pi^1\|_{L^2(\gamma)}\le r,\gamma\in\Gamma(\mu,\nu)\right\}=0,$$

where the infimum is performed over the set of pairs (ϕ,ψ) such that $\phi\in C^3_b(\mathbb{R}^d)$, $\psi\in C^3_b(\mathbb{R}^{2d})$ is symmetric, and $V^X_{(\phi,\psi)}$ is twice Fréchet differentiable at X with law$(X)=\mu$.

Here we have set

$$e_0(\mu,\gamma,\xi_0,A_1,A_2):=\int_{\mathbb{R}^{2d}}\Big(\xi_0(q_1)+\frac{1}{2}A_1(q_1)(q_2-q_1)\Big)\cdot(q_2-q_1)\gamma(dq_1,q_2)$$
$$+\frac{1}{2}\int_{\mathbb{R}^{2d}\times\mathbb{R}^{2d}}A_2(q_1,z)(w-z)\cdot(q_2-q_1)\gamma(dq_1,q_2)\gamma(dz,dw).$$

We shall later use the expression

$$E(r,\mu,\xi_0,A_1,A_2):=\sup_\nu\sup_\gamma\left\{\frac{|e_0\big(\mu,\gamma,\xi_0,A_1,A_2\big)|}{\|\pi^2-\pi^1\|^2_{L^2(\gamma)}}\right\},$$

where the supremum is performed over the set of pairs (ν,γ) such that $0<W_2^2(\nu,\mu)\le r$ and $\gamma\in\Gamma(\mu,\nu)$.

5.2. Specific expression for superjets/subjets. For $\mu,\nu,\gamma,\xi_0,A_1,A_2$ as in Definition 5.3 and $t,s\in(0,T),a\in\mathbb{R}$, we set

$$e(\mathcal{U},s,t,a,\mu,\nu,\gamma,\xi_0,A_1,A_2):=\mathcal{U}(s,\nu)-\mathcal{U}(t,\mu)-a(s-t)-\int_{\mathbb{R}^{2d}}\xi_0(q_1)\cdot(q_2-q_1)\gamma(dq_1,dq_2)$$
$$-\frac{1}{2}\int_{\mathbb{R}^{2d}}A_1(q_1)(q_2-q_1)\cdot(q_2-q_1)\gamma(dx,dy)$$
$$-\frac{1}{2}\int_{\mathbb{R}^{2d}\times\mathbb{R}^{2d}}A_2(q_1,z)(w-z)\cdot(q_2-q_1)\gamma(dz,dw)\gamma(dq_1,dq_2).$$

Similarly, for

$$t,s\in(0,T),a\in\mathbb{R},\quad X,Y,\zeta_0\in E,\quad\overline{A}_1\in L^\infty\big((0,1),\mathbb{R}^{d\times d}\big),\quad\overline{A}_2\in L^\infty\big((0,1)^2,\mathbb{R}^{d\times d}\big),$$

we set

$$\overline{e}(U,s,t,a,X,Y,\zeta_0,\overline{A}_1,\overline{A}_2) := U(s,Y) - U(t,X) - a(s-t) - \int_{(0,1)} \zeta_0 \cdot (Y-X)d\omega$$
$$- \frac{1}{2}\int_{(0,1)} \overline{A}_1(Y-X)\cdot(Y-X)d\omega$$
$$- \frac{1}{2}\int_{(0,1)^2} \overline{A}_2(\omega,o)(Y(o)-X(o))\cdot(Y(\omega)-X(\omega))dod\omega.$$

If the functions $\mathcal{U}$ and U are independent of t, the right-hand sides of the above expressions do not have the $a(s-t)$ term and we will write $e(\mathcal{U},\mu,\nu,\gamma,\xi_0,A_1,A_2)$ and $\overline{e}(U,X,Y,\zeta_0,\overline{A}_1,\overline{A}_2)$.

DEFINITION 5.4. *Let $t\in(0,T)$ and $\mu\in\mathcal{P}_2(\mathbb{R}^d)$ and let $\mathcal{U}:[0,T)\times\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}$.*

(i) *We define the parabolic second order subjet of $\mathcal{U}$ at (t,μ) to be the set $\mathcal{P}^{2,-}\mathcal{U}(t,\mu)$, which consists of $(a,\xi_0,A_1,A_2)\in\mathbb{R}\times T^{*,2}_{\mu}\mathcal{P}_2(\mathbb{R}^d)$ such that*

$$\liminf_{(s,\nu)\to(t,\mu)}\ \inf_{\gamma\in\Gamma(\mu,\nu)} \frac{e(\mathcal{U},s,t,\mu,\nu,\gamma,\xi_0,A_1,A_2)}{|s-t|+W_2^2(\nu,\mu)} \geq 0.$$

(ii) *We define the parabolic second order superjet of $\mathcal{U}$ at (t,μ) to be the set $\mathcal{P}^{2,+}\mathcal{U}(t,\mu)$, of $(a,\xi_0,A_1,A_2)\in\mathbb{R}\times T^{*,2}_{\mu}\mathcal{P}_2(\mathbb{R}^d)$ such that*

$$\limsup_{(s,\nu)\to(t,\mu)}\ \inf_{\gamma\in\Gamma(\mu,\nu)} \frac{e(\mathcal{U},s,t,\mu,\nu,\gamma,\xi_0,A_1,A_2)}{|s-t|+W_2^2(\nu,\mu)} \leq 0.$$

We set

$$S_{(A_1,A_2)}(X)(h) = A_1(X)\,h + \int_{(0,1)} A_2\big(X,X(\omega)\big)h(\omega)d\omega \quad \text{for } h\in E.$$

Owing to the properties of A_1 and A_2 in Definition 5.3, $S_{(A_1,A_2)}(X)$ is a self-adjoint operator on E.

LEMMA 5.5. *Let $\mu\in\mathcal{P}_2(\mathbb{R}^d)$ and let $X\in E$ be such that $X_\sharp\mathcal{L}^1_{(0,1)}=\mu$.*

(i) *If $(a,\xi_0,A_1,A_2)\in\mathcal{P}^{2,-}\mathcal{U}(t,\mu)$, then*

$$\Big(a,\xi_0(X),S_{(A_1,A_2)}(X)\Big)\in\mathcal{P}^{2,-}U(t,X).$$

(ii) *If $(a,\xi_0,A_1,A_2)\in\mathcal{P}^{2,+}\mathcal{U}(t,\mu)$, then*

$$\Big(a,\xi_0(X),S_{(A_1,A_2)}(X)\Big)\in\mathcal{P}^{2,+}U(t,X).$$

Proof. It suffices to prove (i). Let us assume that $(a,\xi_0,A_1,A_2)\in\mathcal{P}^{2,-}\mathcal{U}(t,\mu)$. Let $Y\in E$ and set $Y_\sharp\mathcal{L}^1_{(0,1)}=\nu$. Choose first $\gamma\in\Gamma_0(\mu,\nu)$ and then choose $X^*,Y^*\in E$ such that

$$\gamma := (X^*\times Y^*)_\sharp\mathcal{L}^1_{(0,1)}.$$

Note

$$p := (X\times Y)_\sharp\mathcal{L}^1_{(0,1)}\in\Gamma(\mu,\nu)$$

and

$$\overline{e}(U,s,t,a,X,Y,\xi_0\circ X,A_1\circ X,A_2\circ(X\times X)) = U(s,Y)-U(t,X)-a(s-t) \\ - e_0(\mu,p,\xi_0,A_1,A_2).$$

Hence,

$$\overline{e}(U,s,t,a,X,Y,\xi_0\circ X,A_1\circ X,A_2\circ(X\times X)) = \mathcal{U}(s,\nu)-\mathcal{U}(t,\mu)-a(s-t) \\ - e_0(\mu,p,\xi_0,A_1,A_2), \tag{5.5}$$

where $A_2\circ(X\times X)$ denotes the function $A_2\circ(X\times X)(\omega,o)=A_2(X(\omega),X(o))$.

Fix for a moment a symmetric function $\psi\in C_b^3(\mathbb{R}^{2d})$ and $\phi\in C_b^3(\mathbb{R}^d)$ such that $V^X_{(\phi,\psi)}$ is twice Fréchet differentiable at X. Recall that by (5.4)

$$V^X_{(\phi,\psi)}(Y)-e_0\Big(\mu,p,\nabla\phi,\nabla^2\phi,\nabla_{q_2q_1}\psi\Big)=o\big(\|Y-X\|^2\big). \tag{5.6}$$

Since the first marginal of γ is μ, X and X^* have the same laws, and so, by Lemma 5.2, $V^{X^*}_{(\phi,\psi)}$ is twice Fréchet differentiable at X^*, and so,

$$V^{X^*}_{(\phi,\psi)}(Y^*)-e_0\Big(\mu,\gamma,\nabla\phi,\nabla^2\phi,\nabla_{q_2q_1}\psi\Big)=o\big(\|Y^*-X^*\|^2\big). \tag{5.7}$$

Using (5.5), we have the decomposition

$$\begin{aligned}&\overline{e}\Big(U,s,t,a,X,Y,\xi_0\circ X,A_1\circ X,A_2\circ(X\times X)\Big)\\ &\quad=\mathcal{U}(s,\nu)-\mathcal{U}(t,\mu)-a(s-t)-e_0\Big(\mu,p,\xi_0-\nabla\phi,A_1-\nabla^2\phi,A_2-\nabla_{q_2q_1}\psi\Big)\\ &\qquad-e_0\Big(\mu,p,\nabla\phi,\nabla^2\phi,\nabla_{q_2q_1}\psi\Big).\end{aligned}$$

Thus,

$$\begin{aligned}&\overline{e}\Big(U,s,t,a,X,Y,\xi_0\circ X,A_1\circ X,A_2\circ(X\times X)\Big)\\ &=\mathcal{U}(s,\nu)-\mathcal{U}(t,\mu)-a(s-t)-e_0\Big(\mu,p,\xi_0-\nabla\phi,A_1-\nabla^2\phi,A_2-\nabla_{q_2q_1}\psi\Big)\\ &\quad-e_0\Big(\mu,p,\nabla\phi,\nabla^2\phi,\nabla_{q_2q_1}\psi\Big)-e_0\Big(\mu,\gamma,\xi_0,A_1,A_2\Big)\\ &\quad+e_0\Big(\mu,\gamma,\xi_0-\nabla\phi,A_1-\nabla^2\phi,A_2-\nabla_{q_2q_1}\psi\Big)\\ &\quad+e_0\Big(\mu,\gamma,\nabla\phi,\nabla^2\phi,\nabla_{q_2q_1}\psi\Big).\end{aligned}$$

Rearranging, we obtain

$$\begin{aligned}\overline{e}\Big(U,s,t,a,X,Y,\xi_0\circ X,A_1\circ X,A_2\circ(X\times X)\Big)&=\mathcal{U}(s,\nu)-\mathcal{U}(t,\mu)-a(s-t)-e_0\Big(\mu,\gamma,\xi_0,A_1,A_2\Big)\\ &\quad-e_0\Big(\mu,p,\xi_0-\nabla\phi,A_1-\nabla^2\phi,A_2-\nabla_{q_2q_1}\psi\Big)\\ &\quad-e_0\Big(\mu,p,\nabla\phi,\nabla^2\phi,\nabla_{q_2q_1}\psi\Big)\\ &\quad+e_0\Big(\mu,\gamma,\nabla\phi,\nabla^2\phi,\nabla_{q_2q_1}\psi\Big)\\ &\quad+e_0\Big(\mu,\gamma,\xi_0-\nabla\phi,A_1-\nabla^2\phi,A_2-\nabla_{q_2q_1}\psi\Big).\end{aligned}$$

Using Remark 5.1(i) we conclude that

$$\begin{aligned}\overline{e}\Big(U,s,t,a,X,Y,\xi_0\circ X,A_1\circ X,A_2\circ(X\times X)\Big) &= \mathcal{U}(s,\nu)-\mathcal{U}(t,\mu)-a(s-t)-e_0\Big(\mu,\gamma,\xi_0,A_1,A_2\Big)\\ &\quad -e_0\Big(\mu,p,\xi_0-\nabla\phi,A_1-\nabla^2\phi,A_2-\nabla_{q_2q_1}\psi\Big)\\ &\quad V^{X^*}_{(\phi,\psi)}(Y^*)-e_0\Big(\mu,p,\nabla\phi,\nabla^2\phi,\nabla_{q_2q_1}\psi\Big)\\ &\quad +e_0\Big(\mu,\gamma,\nabla\phi,\nabla^2\phi,\nabla_{q_2q_1}\psi\Big)-V^X_{(\phi,\psi)}(Y)\\ &\quad +e_0\Big(\mu,\gamma,\xi_0-\nabla\phi,A_1-\nabla^2\phi,A_2-\nabla_{q_2q_1}\psi\Big).\end{aligned}$$

We first use the fact that $(a,\xi_0,A_1,A_2)\in\mathcal{P}^-\mathcal{U}(t,\mu)$, second use (5.6) and (5.7), and third use the fact that $(\xi_0,A_1,A_2)\in T^{*,2}_{\mu}\mathcal{P}_2(\mathbb{R}^d)$ to conclude that

$$\begin{aligned}&\overline{e}\Big(U,s,t,a,X,Y,\xi_0\circ X,A_1\circ X,A_2\circ(X\times X)\Big)\\ &\ge o\big(W_2^2(\nu,\mu)\big)+o(|s-t|)-o\big(\|Y-X\|^2\big)-o\big(\|Y^*-X^*\|^2\big)\\ &\quad -E\Big(\|\pi^1-\pi^2\|_{L^2(\gamma)},\xi_0-\nabla\phi,A_1-\nabla^2\phi,A_2-\nabla_{q_2q_1}\psi\Big)\|\pi^1-\pi^2\|^2_{L^2(\gamma)}\\ &\quad -E\Big(\|\pi^1-\pi^2\|_{L^2(p)},\xi_0-\nabla\phi,A_1-\nabla^2\phi,A_2-\nabla_{q_2q_1}\psi\Big)\|\pi^1-\pi^2\|^2_{L^2(p)}.\end{aligned}$$

Since

$$\|Y-X\|=\|\pi^1-\pi^2\|^2_{L^2(p)}\ge\|Y^*-X^*\|=\|\pi^1-\pi^2\|^2_{L^2(\gamma)}$$

we conclude that for any $r>0$

$$\begin{aligned}\liminf_{(s,Y)\to(t,X)}\frac{\overline{e}(U,s,t,a,X,Y,\xi_0\circ X,A_1\circ X,A_2\circ(X\times X))}{|s-t|+\|Y-X\|^2} &\ge -E\Big(r,\xi_0-\nabla\phi,A_1-\nabla^2\phi,A_2-\nabla_{q_2q_1}\psi\Big)\\ &\quad -E\Big(r,\xi_0-\nabla\phi,A_1-\nabla^2\phi,A_2-\nabla_{q_2q_1}\psi\Big).\end{aligned}\tag{5.8}$$

Maximizing $-E$, which means minimizing E over (r,ϕ,ψ) and using the fact that $(\xi_0,A_1,A_2)\in T^{*,2}_{\mu}\mathcal{P}_2(\mathbb{R}^d)$ we conclude

$$\liminf_{(s,Y)\to(t,X)}\frac{\overline{e}(U,s,t,a,X,Y,\xi_0\circ X,A_1\circ X,A_2\circ(X\times X))}{|s-t|+\|Y-X\|^2}\ge 0,$$

which proves (i). ∎

Definition 5.6. *Suppose $\kappa>0$. An upper semicontinuous function $\mathcal{U}:[0,T)\times\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}$ is an* intrinsic *viscosity subsolution of* (1.1) *on the Wasserstein space if $\mathcal{U}(0,\cdot)\le\mathcal{U}_0$ on $\mathcal{P}_2(\mathbb{R}^d)$ and*

$$a-\kappa\bigg(\int_{\mathbb{R}^d}\mathrm{Tr}(A_1(q))\mu(dq)+\int_{\mathbb{R}^{2d}}\mathrm{Tr}(A_2(q_1,q_2))\mu(dq_1)\mu(dq_2)\bigg)+\mathcal{H}(\mu,\mu,\xi_0)+\mathcal{F}(\mu)\le 0$$

for all $(t,\mu)\in(0,T)\times\mathcal{P}_2(\mathbb{R}^d)$ and $(a,\xi_0,A_1,A_2)\in\mathcal{P}^{2,+}\mathcal{U}(t,\mu)$.

A lower semicontinuous function $\mathcal{U}:[0,T)\times\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}$ is an intrinsic *viscosity supersolution of* (1.1) *on the Wasserstein space if $\mathcal{U}(0,\cdot)\ge\mathcal{U}_0$ on $\mathcal{P}_2(\mathbb{R}^d)$ and*

$$a-\kappa\bigg(\int_{\mathbb{R}^d}\mathrm{Tr}(A_1(q))\mu(dq)+\int_{\mathbb{R}^{2d}}\mathrm{Tr}(A_2(q_1,q_2))\mu(dq_1)\mu(dq_2)\bigg)+\mathcal{H}(\mu,\mu,\xi_0)+\mathcal{F}(\mu)\ge 0\tag{5.9}$$

for all $(t,\mu)\in(0,T)\times\mathcal{P}_2(\mathbb{R}^d)$ and $(a,\xi_0,A_1,A_2)\in\mathcal{P}^{2,-}\mathcal{U}(t,\mu)$.

If $\mathcal{U}$ is both an intrinsic *viscosity subsolution and an* intrinsic *viscosity supersolution of* (1.1) *on the Wasserstein space, we say it is an* intrinsic *viscosity solution of* (1.1) *on the Wasserstein space.*

THEOREM 5.7. *Let* $\mathcal{U} : [0,T) \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$.

(i) *If* $\mathcal{U}$ *is an* L*-viscosity subsolution of* (1.1) *on the Wasserstein space, then it is an* intrinsic *viscosity subsolution of* (1.1).

(ii) *If* $\mathcal{U}$ *is an* L*-viscosity supersolution of* (1.1) *on the Wasserstein space, then it is an* intrinsic *viscosity supersolution of* (1.1).

(iii) *If* $\mathcal{U}$ *is an* L*-viscosity solution of* (1.1) *on the Wasserstein space, then it is an* intrinsic *viscosity solution of* (1.1).

Proof. It suffices to prove (ii). Assume $\mathcal{U}$ is an L-viscosity supersolution of (1.1) on the Wasserstein space.

Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $t \in [0,T)$. Choose $X \in E$ such that $X_\sharp \mathcal{L}_1 = \mu$. We have $\mathcal{U}(0,\mu) = U(0,X) \geq U_0(X) = \mathcal{U}_0(\mu)$.

In order to show that $\mathcal{U}$ is upper semicontinuous at (t,μ), we choose an arbitrary sequence $(\mu_n)_n \subset \mathcal{P}_2(\mathbb{R}^d)$ converging to μ and an arbitrary sequence $(t_n)_n \subset [0,T)$ converging to t. Let $(X_n)_n \subset E$ such that $X_{n\sharp}\mathcal{L}_1 = \mu_n$ and $(X_n)_n$ converges to X. We have

$$\lim_{n\to\infty} \mathcal{U}(t_n,\mu_n) = \lim_{n\to\infty} U(t_n,X_n) \geq U(t,X) = \mathcal{U}(t,\mu).$$

Thus, $\mathcal{U}$ is lower semicontinuous at (t,μ).

Let now $t > 0$ and $(a,\xi_0,A_1,A_2) \in \mathcal{P}^{2,-}\mathcal{U}(t,\mu)$. We would like to show that (5.9) holds. Let $X \in E$ be such that $X_\sharp \mathcal{L}^1_{(0,1)} = \mu$. By Lemma 5.5,

$$\Big(a, \xi_0(X), S_{(A_1,A_2)}\Big) \in \mathcal{P}^- U(t,X).$$

Since U is an L-viscosity supersolution of (1.1) on the Wasserstein space, we use Proposition 2.3 to infer

$$a - \kappa\left(\sum_{k=1}^d \int_{(0,1)} A_1(X(\omega))e_k \cdot e_k d\omega + \int_{(0,1)^2} A_2(X(\omega),X(o))e_k \cdot e_k d\omega do\right) + \tilde{H}\big(X, X_\sharp\mathcal{L}_1, \xi_0(X)\big) + F(X) \geq 0.$$

This gives (5.9). ∎

Remark 5.8. Let $U : [0,T) \times E \to \mathbb{R}$ and let $\mathcal{U} : [0,T) \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ be such that $\mathcal{U}(t,\mu) = U(t,X)$ whenever $X \in E$ is the law of μ. In [43], it was proved that if U is a viscosity solution of the first order equation

$$\begin{cases} \partial_t U + \tilde{H}(X, X_\sharp\mathcal{L}_1, DU) + F(X) = 0 & \text{in } (0,T) \times E, \\ U(0,X) = U_0(X) & \text{on } E, \end{cases} \tag{5.10}$$

then $\mathcal{U}$ is an intrinsic viscosity solution of the first order equation

$$\begin{cases} \partial_t \mathcal{U} + \mathcal{H}(\mu,\mu,\nabla_\mu \mathcal{U}) + \mathcal{F}(\mu) = 0 & \text{in } (0,T) \times \mathcal{P}_2(\mathbb{R}^d), \\ \mathcal{U}(0,\mu) = \mathcal{U}_0(\mu) & \text{on } \mathcal{P}_2(\mathbb{R}^d), \end{cases} \tag{5.11}$$

according to the definition proposed in [43]. Therefore, Theorem 5.7 is an extension of the results of [43] from the case $\kappa = 0$ to the case $\kappa > 0$.

6. First order convex HJB equations and value functions. In this section we show that if $\kappa = 0$ and H does not depend on μ and is convex in the gradient variable, then the solutions u_n of (1.2), which are value functions of optimal control problems for n-particle systems, converge to the value function of a variational problem in $\mathcal{P}_2(\mathbb{R}^d)$. Thus we obtain a representation formula for the solution of (1.1).

Hypothesis 6.1. The function $H = H(x,p)$, in addition to satisfying Hypotheses 2.1 in the x and p variables, is convex in the p variable and

$$H(x,p) \geq C_1 + C_2|p|^2 \quad \forall\, x,p \in \mathbb{R}^d \tag{6.1}$$

for some constants C_1, C_2, where $C_2 > 0$.

We define $L(x,v)$ to be the Legendre transform of $H(x,p)$, that is,

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

This implies, by (2.3) and (6.1),

$$C_3 + C_4|v|^2 \leq L(x,v) \leq C_5 + C_6|v|^2 \quad \text{for some } C_3, C_4, C_5, C_6 \text{ with } C_4, C_6 > 0. \tag{6.2}$$

Given $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $\xi \in L^2_\mu(\mathbb{R}^d;\mathbb{R}^d)$, we define

$$\mathcal{L}(\mu,\xi) := \int_{\mathbb{R}^d} L(x,\xi(x))\mu(dx) - \mathcal{F}(\mu).$$

For $0 \leq t \leq T$, we define the action

$$\bar{\mathcal{A}}(\sigma,v) := \int_0^t \mathcal{L}(\sigma_\tau, v_\tau)d\tau + \mathcal{U}_0(\sigma_0).$$

Let $u_n : [0,T]\times(\mathbb{R}^d)^n$ be, as before, the viscosity solution to (1.2), for $n = 1,\dots$.

For $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, let

$$\bar{\mathcal{U}}(t,\mu) := \inf_{(\sigma,v)} \left\{ \bar{\mathcal{A}}(\sigma,v) \;\middle|\; \sigma_t = \mu \right\} \tag{6.3}$$

with the infimum taken over all the pairs (σ,v), where[1] $\sigma = \sigma_\tau \in AC^2(0,t;\mathcal{P}_2(\mathbb{R}^d))$, $v = v_\tau$ is a velocity vector field for σ_τ, and $\sigma_t = \mu$. Here $AC^2(0,t;\mathcal{P}_2(\mathbb{R}^d))$ is the space of absolutely continuous curves in $\mathcal{P}_2(\mathbb{R}^d)$ with square-integrable metric derivative; see [3, Definition 1.1.1]. Define

$$\bar{u}_n(t,\mathbf{x}) = \bar{\mathcal{U}}\left(t, \frac{1}{n}\sum_{j=1}^n \delta_{x_j}\right), \quad \mathbf{x} = (x_1,\dots,x_n) \in (\mathbb{R}^d)^n.$$

We want to investigate the asymptotic relationship between $\bar{u}_n$ and u_n.

Set

$$f(\mathbf{x}) = \mathcal{F}\left(\frac{1}{n}\sum_{j=1}^n \delta_{x_j}\right), \quad u_0(\mathbf{x}) = \mathcal{U}_0\left(\frac{1}{n}\sum_{j=1}^n \delta_{x_j}\right),$$

and

$$l_n(\mathbf{x},\mathbf{v}) = -f(\mathbf{x}) + \frac{1}{n}\sum_{j=1}^n L(x_j,v_j).$$

Define

$$\mathcal{C}_n(t,\mathbf{x}) := \{x(\cdot) \in AC^2(0,t;(\mathbb{R}^d)^n) \;\big|\; x(t) = \mathbf{x}\}.$$

[1] We use the subindex notation σ_τ or $\sigma(\tau)$ interchangeably to mean the value of the path at time τ.

With the conditions on H listed in Hypothesis 2.1, the solution u_n to (1.2) has the value function representation

$$u_n(t,\mathbf{x}) = \inf_{\xi}\left\{\int_0^t l_n(\xi(\tau),\dot{\xi}(\tau))d\tau + u_0(\xi(0)) \;\middle|\; \xi(\cdot)\in\mathcal{C}_n(t,\mathbf{x})\right\}. \tag{6.4}$$

Denote by $\mathcal{A}$ the functional that is minimized in (6.4), i.e.,

$$\mathcal{A}(\xi(\cdot)) := \int_0^t l_n(\xi(\tau),\dot{\xi}(\tau))d\tau + u_0(\xi(0)).$$

Observe that when $\sigma_\tau = \frac{1}{n}\sum_{j=1}^n \delta_{x_j(\tau)}$ for $x(\cdot)\in AC^2(0,t;(\mathbb{R}^d)^n)$, then

$$v_\tau(x) = \sum_{j=1}^n \mathbf{1}_{x_j(\tau)}\dot{x}_j(\tau) \quad \text{for a.e. } \tau\in(0,t) \qquad \text{and} \qquad \bar{\mathcal{A}}(\sigma,v) = \mathcal{A}(x(\cdot)).$$

We will make use of the following lemma.

LEMMA 6.2. *Let $\mu\in\mathcal{P}_2(\mathbb{R}^d)$, and let $\sigma\in AC^2(0,t;\mathcal{P}_2(\mathbb{R}^d))$ be a path of velocity w such that $\sigma_t=\mu$. There exist sequences $\{y^m\}_{m=1}^\infty$, $y^m\in(\mathbb{R}^d)^m$, $\{\sigma^m\}_{m=1}^\infty$, $\sigma^m\in\mathcal{C}_m(t,y^m)$ with corresponding velocity vector fields w^m, and $\{r_m\}_1^\infty$, $r_m\searrow 0$, such that*

$$\sup_{0\le\tau\le t} d_2(\sigma_\tau,\sigma_\tau^m)\le r_m, \tag{6.5}$$

$$\bar{\mathcal{A}}(\sigma^m,w^m)\le\bar{\mathcal{A}}(\sigma,w)+r_m. \tag{6.6}$$

Proof. We are first going to prove the existence of such sequences as in the statement, for which

$$\int_0^t\int_{\mathbb{R}^d} L(x,w_\tau^m(x))\sigma_\tau^m(dx)d\tau \le \int_0^t\int_{\mathbb{R}^d} L(x,w_\tau(x))\sigma_\tau(dx)d\tau + r_m. \tag{6.7}$$

Step 1. We start with a standard mollification procedure by setting

$$\eta(x) := \frac{1}{(4\pi)^{d/2}}\exp(-|x|^2/4), \qquad \eta^\varepsilon(x) := \frac{1}{\varepsilon^d}\eta(x/\varepsilon), \qquad \sigma_\tau^\varepsilon = \sigma_\tau * \eta^\varepsilon,$$

$$w_\tau^\varepsilon = \frac{w_\tau\sigma_\tau * \eta^\varepsilon}{\sigma_\tau^\varepsilon}, \quad j^\varepsilon(x,y) := \frac{\eta^\varepsilon(x-y)}{\int_{\mathbb{R}^d}\eta^\varepsilon(x-y)\sigma_\tau(dy)}.$$

By Lemma 7.1.10 of [3],

$$d_2^2(\sigma_\tau^\varepsilon,\sigma_\tau)\le\varepsilon\int_{\mathbb{R}^d}|x|^2\eta(x)dx \qquad \forall\quad 0\le\tau\le t. \tag{6.8}$$

Let us now prove that

$$\limsup_{\varepsilon\to 0}\int_0^t\int_{\mathbb{R}^d} L(x,w_\tau^\varepsilon)\sigma_\tau^\varepsilon(dx)d\tau \le \int_0^t\int_{\mathbb{R}^d} L(x,w_\tau(x))\sigma_\tau(dx)d\tau. \tag{6.9}$$

Note for any arbitrary fixed τ,

$$w_\tau^\varepsilon(x) = \int_{\mathbb{R}^d} w_\tau(y)j^\varepsilon(x,y)\sigma_\tau(dy).$$

Since for every $x \in \mathbb{R}^d$, $\int_{\mathbb{R}^d} j^\varepsilon(x,y)\sigma_\tau(dy) = 1$ and $L(x,\cdot)$ is convex, we use Jensen's inequality to infer

(6.10)
$$L(\cdot, w_\tau^\varepsilon) = L\left(\cdot, \int_{\mathbb{R}^d} w_\tau(y) j^\varepsilon(\cdot,y)\sigma_\tau(dy)\right) \le \int_{\mathbb{R}^d} L(\cdot, w_\tau(y)) j^\varepsilon(\cdot,y)\sigma_\tau(dy) < \infty.$$

We have obtained the finiteness of the expression at the right-hand side of (6.10) since

$$|L(\cdot, w_\tau(y))| \le C(1+|w_\tau(y)|^2), \quad j^\varepsilon \in L^\infty(\mathbb{R}^{2d}), \quad w_\tau \in L^2(\sigma_\tau).$$

Observe that the function $x \mapsto \int_{\mathbb{R}^d} L(x, w_\tau(y)) j^\varepsilon(x,y)\sigma_\tau(dy)$ belongs to $L^1(\sigma_\tau^\varepsilon)$. Indeed,

$$\begin{aligned}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |L(x,w_\tau(y))| j^\varepsilon(x,y)\sigma_\tau(dy)\sigma_\tau^\varepsilon(dx) &\le \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} C(1+|w_\tau(y)|^2) j^\varepsilon(x,y)\sigma_\tau(dy)\sigma_\tau^\varepsilon(dx)\\ &= C + \int_{\mathbb{R}^d} |w_\tau(y)|^2 \left(\int_{\mathbb{R}^d} \eta^\varepsilon(x-y)dx\right)\sigma_\tau(dy)\\ &\le C(1+\|w_\tau\|^2_{L^2(\sigma_\tau)}).\end{aligned}$$

Similarly,

$$\begin{aligned}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} L(x,w_\tau(y)) j^\varepsilon(x,y)\sigma_\tau(dy)\sigma_\tau^\varepsilon(dx) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} L(x,w_\tau(y)) j^\varepsilon(x,y)\sigma_\tau^\varepsilon(dx)\sigma_\tau(dy)\\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} L(x,w_\tau(y))\eta^\varepsilon(x-y)dx\sigma_\tau(dy).\end{aligned}$$

Thus, integrating (6.10) on both sides with respect to σ_τ^ε, we get

$$\int_{\mathbb{R}^d} L(x, w_\tau^\varepsilon(x))\sigma_\tau^\varepsilon(dx) \le \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} L(x,w_\tau(y))\eta^\varepsilon(x-y)dx\sigma_\tau(dy). \tag{6.11}$$

Classic arguments show that

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} L(x,w_\tau(y))\eta^\varepsilon(x-y)dx\sigma_\tau(dy) = \int_{\mathbb{R}^d} L(y,w_\tau(y))\sigma_\tau(dy).$$

From this, together with (6.11), it follows that

$$\limsup_{\varepsilon\to 0}\int_{\mathbb{R}^d} L(x,w_\tau^\varepsilon)\sigma_\tau^\varepsilon(dx) \le \int_{\mathbb{R}^d} L(x,w_\tau(x))\sigma_\tau(dx).$$

An application of Fatou's lemma now yields (6.9).

Step 2. Notice that the constructed σ_τ^ε solve the continuity equation

$$\partial_\tau \sigma_\tau^\varepsilon + \operatorname{div}(w_\tau^\varepsilon \sigma_\tau^\varepsilon) = 0 \qquad \text{in } (0,t)\times\mathbb{R}^d,$$

because

$$\operatorname{div}(w_\tau^\varepsilon\sigma_\tau^\varepsilon) = \operatorname{div}((w_\tau\sigma_\tau)*\eta^\varepsilon) = (\operatorname{div}(w_\tau\sigma_\tau))*\eta^\varepsilon \quad \text{and} \quad \partial_\tau\sigma_\tau^\varepsilon = (\partial_\tau\sigma_\tau)*\eta^\varepsilon.$$

Since $w^\varepsilon_\tau \sigma^\varepsilon_\tau$ is smooth, for arbitrary $\delta > 0$ we can find $\sigma^{\varepsilon,N}_\tau \in \mathcal{C}_N(t, y^N)$ for some $N \in \mathbb{N}$ and $y^N \in (\mathbb{R}^d)^N$, satisfying

$$\sup_{0\le\tau\le t} d_2(\sigma^\varepsilon_\tau, \sigma^{\varepsilon,N}_\tau) \le \delta,$$
$$\int_0^t \int_{\mathbb{R}^d} L(x, w^{\varepsilon,N}_\tau(x))\sigma^{\varepsilon,N}_\tau(dx)d\tau \le \int_0^t \int_{\mathbb{R}^d} L(x, w^\varepsilon_\tau(x))\sigma^\varepsilon_\tau(dx)d\tau + \delta.$$

It is clear that combining the latter inequalities, together with (6.9) and (6.8), gives the desired sequence r_m such that inequalities (6.7) and (6.5) hold.

To finish the proof, note that by (6.5) and the uniform continuity of $\mathcal{F}, \mathcal{U}_0$, there exists a sequence $s_m \searrow 0$ such that

$$\mathcal{U}_0(\sigma^m_0) \le \mathcal{U}_0(\sigma_0) + s_m, \quad -\int_0^t \mathcal{F}(\sigma^m_\tau)d\tau \le -\int_0^t \mathcal{F}(\sigma_\tau)d\tau + s_m.$$

Denoting $r_m + 2s_m$ still by r_m, we obtain (6.6). ∎

LEMMA 6.3. *For any $0 \le t \le T$, the value function $\bar{\mathcal{U}}(t,\cdot)$ is lower semicontinuous on $\mathcal{P}_2(\mathbb{R}^d)$.*

Proof. Let $\mu^n \to \mu$ be such that

$$\lim_{n\to\infty} \bar{\mathcal{U}}(t, \mu^n) = \liminf_{\nu\to\mu} \bar{\mathcal{U}}(t, \nu).$$

Let $\varepsilon > 0$ and let $\sigma^n \in AC^2(0, t; \mathcal{P}_2(\mathbb{R}^d))$ be paths of velocity v^n such that $\sigma^n_t = \mu^n$ and

$$\int_0^t \mathcal{L}(\sigma^n_\tau, v^n_\tau)d\tau + \mathcal{U}_0(\mu^n) < \bar{\mathcal{U}}(t, \mu^n) + \varepsilon.$$

It follows from (6.2) that

$$\int_0^t \|v^n_\tau\|^2_{L^2(\sigma^n_\tau)} d\tau < C \tag{6.12}$$

for some C independent of n. Therefore, by Proposition 7.1 in the appendix, we have the existence of a subsequence (still denoted by σ^n) and $\sigma \in AC^2(0, t; \mathcal{P}_2(\mathbb{R}^d))$, with $\sigma_t = \mu$, such that for every $s \in [0, t]$, σ^n_s converges narrowly to σ_s. Denote the product measures on $\mathbb{R}^d \times [0, t]$ by $\sigma^n_s ds$. These converge narrowly to $\sigma_s ds$. Furthermore, denote by $v^n_s \sigma^n_s ds$ the vector measure whose density with respect to $\sigma^n_s ds$ is the time-dependent vector field $v^n_s = v^n(s, x)$.

We then obtain from (6.12) that there exists a subsequence of (σ^n, v^n), still indexed by n, such that $\sigma^n_s ds$ converge narrowly to $\sigma_s ds$ while $v^n_s \sigma^n_s ds$ converge narrowly to a vector measure w on $\mathbb{R}^d \times [0, t]$.

Let $\varphi \in C^1_c((0, t) \times \mathbb{R}^d)$. Then

$$0 = \lim_{n\to\infty} \left(\int_0^t \int_{\mathbb{R}^d} \partial_s \varphi(s, x)\sigma^n_s(dx)ds + \int_0^t \int_{\mathbb{R}^d} D\varphi(s, x) \cdot v^n_s(x)\sigma^n_s(dx)ds \right)$$
$$= \int_0^t \int_{\mathbb{R}^d} \varphi(s, x)\sigma_s(dx)ds + \int_{[0,t]\times\mathbb{R}^d} D\varphi(s, x) \cdot w(dx, ds). \tag{6.13}$$

By Proposition 7.2 in the appendix, $w \ll \sigma_s ds$, so there is an $L^1(\sigma_s ds)$ vector field $v(s,x)$ such that $w = v_s \sigma_s ds$, and, by the same proposition,

$$\int_0^t \int_{\mathbb{R}^d} L(x, v_\tau)\sigma_\tau(dx)d\tau \le \liminf_{n\to\infty} \int_0^t \int_{\mathbb{R}^d} L(x, v_\tau^n)\sigma_\tau^n(dx)d\tau.$$

However, by (6.2) and (6.12), we actually obtain

$$\int_0^t \|v_\tau\|^2_{L^2(\sigma_\tau)} d\tau < C$$

for some constant C which, together with (6.13), means that v is a velocity vector field for σ. Therefore, since $\mathcal{F}$ is narrowly continuous, it follows that

$$\bar{\mathcal{U}}(t,\mu) \le \liminf_{\nu\to\mu} \bar{\mathcal{U}}(t,\nu). \qquad \square$$

THEOREM 6.4. *Given $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $0 \le t \le T$, there exists a sequence $\{\mathbf{x}(n)\}_{n=1}^\infty$, $\mathbf{x}(n) \in (\mathbb{R}^d)^n$, such that $d_2(\frac{1}{n}\sum_{j=1}^n \delta_{x_j(n)}, \mu) \to 0$ as $n \to 0$ and*

$$\bar{\mathcal{U}}(t,\mu) = \lim_{n\to\infty} \inf_{\xi(\cdot)\in\mathcal{C}_n(t,\mathbf{x}(n))} \left\{ \int_0^t l_n(\xi(\tau), \dot{\xi}(\tau))d\tau + u_0(\xi(0)) \;\middle|\; \xi(t) = \mathbf{x}(n) \right\},$$

i.e.,

$$\lim_{n\to\infty} u_n(t, \mathbf{x}(n)) = \bar{\mathcal{U}}(t,\mu).$$

In particular, $\bar{\mathcal{U}} = \mathcal{U}$ from Theorem 1.2 *and $\bar{\mathcal{U}}$ is continuous and satisfies the continuity estimate* (4.1).

Proof. Let $\{\sigma_k, v_k\}_{k=1}^\infty$ be a minimizing sequence of paths and velocities for $\mathcal{U}(t,\mu)$ such that

$$\bar{\mathcal{A}}(\sigma_k, v_k) \le \bar{\mathcal{U}}(t,\mu) + 1/k. \tag{6.14}$$

By Lemma 6.2, for each $k \in \mathbb{N}$ there exists a sequence $\{\sigma_k^m, v_k^m\}_{m=1}^\infty$, with the mth term in $\mathcal{C}_m(t, \sigma_k^m(t))$, such that

(6.15)
$$u_m(t, \sigma_k^m(t)) \le \mathcal{A}(\sigma_k^m) \le \bar{\mathcal{A}}(\sigma_k, v_k) + 1/m \quad \text{and} \quad \sigma_k^m(t) \underset{m\to\infty}{\longrightarrow} \sigma_k(t) = \mu \text{ in } d_2.$$

Then,

$$u_m(t, \sigma_k^m(t)) \le \bar{\mathcal{U}}(t,\mu) + 1/k + 1/m;$$

consequently,

$$\limsup_{m\to\infty} u_m(t, \sigma_m^m(t)) \le \bar{\mathcal{U}}(t,\mu).$$

Hence, since $d_2(\sigma_m^m(t), \mu) \underset{m\to\infty}{\longrightarrow} 0$, this, together with the lower semicontinuity of $\bar{\mathcal{U}}(t,\cdot)$ proved in Lemma 6.3, gives

$$\lim_{m\to\infty} u_m(t, \sigma_m^m(t)) = \bar{\mathcal{U}}(t,\mu).$$

Putting $\mathbf{x}(n) := \sigma_n^n(t)$, $n = 1, \ldots$, proves the statement. $\square$

7. Appendix.

PROPOSITION 7.1. *Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and let $(\sigma^n, v^n)_{n=1}^{\infty}$ be a sequence such that for each $n \in \mathbb{N}$, $\sigma^n \in AC^2(0,t;\mathcal{P}_2(\mathbb{R}^d))$ and v^n is a velocity vector field for σ^n. If*

$$\lim_{n\to\infty} d_2(\sigma_t^n, \mu) = 0 \quad \text{and} \quad \int_0^t \|v_\tau^n\|^2_{L^2(\sigma_\tau^n)} d\tau < C \quad \forall n \in \mathbb{N},$$

then there exists a subsequence $(\sigma^{n_k})_{k=1}^{\infty}$ and $\sigma \in AC^2(0,t;\mathcal{P}_2(\mathbb{R}^d))$, with $\sigma_t = \mu$, such that for a.e. $\tau \in [0,t]$, $\sigma_\tau^{n_k} \to \sigma_\tau$ narrowly.

Proof. Note that if $0 \le s_1 < s_2 \le t$, by Hölder's inequality we get

$$d_2(\sigma_{s_1}^n, \sigma_{s_2}^n) \le \int_{s_1}^{s_2} \|v_\tau^n\|_{L^2(\sigma_\tau^n)} d\tau \le \sqrt{C}\sqrt{s_2 - s_1}. \tag{7.1}$$

In particular, $(\sigma^n)_n$ is bounded and equicontinuous in $\mathcal{P}_2(\mathbb{R}^d)$. We apply the refined version of the Ascoli–Arzelà theorem in Proposition 3.3.1 of [3] to conclude the proof. ∎

PROPOSITION 7.2. *Let L be as in section 6. Consider a sequence $\{\nu_n\}_1^{\infty} \cup \{\nu\}$ of finite, positive Borel measures on $[0,T] \times \mathbb{R}^d$ that converges narrowly to ν. Suppose we have a sequence $g_n : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ of vector fields such that*

$$\int_{[0,T]\times\mathbb{R}^d} |g_n(t,x)|^2 \nu_n(dt,dx) < \infty \tag{7.2}$$

and $(g_n\nu_n)_n$ converges narrowly to a vector-valued Borel measure λ on $[0,T] \times \mathbb{R}^d$. Then,

(i) *there exists a Borel vector field $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ such that $\lambda = v\nu$;*
(ii) *we have*

$$\int_{[0,T]\times\mathbb{R}^d} L\big(x, v(t,x)\big)\nu(dt,dx) \le \liminf_{n\to\infty} \int_{[0,T]\times\mathbb{R}^d} L(x, g_n(t,x))\nu_n(dt,dx).$$

Proof. We define on $[0,T] \times \mathbb{R}^{2d}$ the measures f_n by

$$\int_{[0,T]\times\mathbb{R}^{2d}} \Phi(t,x,w) f_n(dx,dw) = \int_{[0,T]\times\mathbb{R}^d} \Phi(t,x,g_n(t,x))\nu_n(t,dx) \tag{7.3}$$

for $\Phi \in C_b([0,T] \times \mathbb{R}^{2d})$.

(i) We use (7.2) and the fact that $(\nu_n)_n$ is precompact for the narrow convergence topology to conclude that $(f_n)_n$ is precompact for the narrow convergence topology. Therefore, without loss of generality, we may assume that $(f_n)_n$ converges narrowly to some Borel measure on $[0,T] \times \mathbb{R}^{2d}$ which we denote by f_∞. When Φ depends only on the (t,x) variables, passing to the limit in (7.3), we conclude that the first marginal of f_∞ is ν. Hence, there exists a Borel map $(t,x) \to f_\infty^{(t,x)}$ of probability measures (cf. [3, subsection 5.3]) such that we have the disintegration

$$\int_{[0,T]\times\mathbb{R}^{2d}} \Phi(t,x,w) f_\infty(dt,dx,dw) = \int_{[0,T]\times\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \Phi(t,x,w) f_\infty^{(t,x)}(dw) \right) \nu(dt,dx)$$

for all $\Phi \in C_b([0,T] \times \mathbb{R}^{2d})$.

Let $\varphi \in C([0,T]\times\mathbb{R}^d,\mathbb{R}^d)$ be a bounded function. Although $(t,x,w)\in[0,T]\times\mathbb{R}^{2d}\mapsto\varphi(t,x)\cdot w$ is not bounded, (7.2) allows us to assert that (setting $z=(t,x)$)

$$\begin{aligned}\int_{[0,T]\times\mathbb{R}^{2d}}\varphi(z)\cdot w f_\infty(dz,dw) &= \lim_{n\to\infty}\int_{[0,T]\times\mathbb{R}^{2d}}\varphi(z)\cdot w f_n(dz,dw)\\ &= \lim_{n\to\infty}\int_{[0,T]\times\mathbb{R}^{d}}\varphi(z)\cdot g_n(z)\nu_n(dz).\end{aligned}$$

We now use the fact that λ is a point of accumulation of $(g_n\nu_n)$ to conclude that

$$\int_{[0,T]\times\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\varphi(t,x)\cdot w f_\infty^{(t,x)}(dw)\right)\nu(dt,dx)=\int_{[0,T]\times\mathbb{R}^d}\varphi(t,x)\cdot\lambda(dt,dx).$$

We conclude the proof of (i) by setting $v(t,x):=\int_{\mathbb{R}^d} w f_\infty^{(t,x)}(dw)$.

(ii) Since L is bounded below by the hypotheses, we may suppose without loss of generality that $L\ge 0$. For each $r>0$ let $\Phi_r\in C(\mathbb{R}^{2d})$ be a function which is identically 1 on the ball of radius r, is zero outside of the ball of radius $r+1$, but remains between 0 and 1 everywhere. We have

$$\begin{aligned}\int_{[0,T]\times\mathbb{R}^{2d}}L(x,w)\Phi_r(x,w)f_\infty(dt,dx,dw) &= \liminf_{n\to\infty}\int_{[0,T]\times\mathbb{R}^{2d}}L(x,w)\Phi_r(x,w)f_n(dt,dx,dw)\\ &\le \liminf_{n\to\infty}\int_{[0,T]\times\mathbb{R}^{2d}}L(x,w)f_n(dt,dx,dw)\end{aligned}$$

and so, letting $r\to\infty$, we conclude

$$\int_{[0,T]\times\mathbb{R}^{2d}}L(x,w)f_\infty(dt,dx,dw)\le\liminf_{n\to\infty}\int_{[0,T]\times\mathbb{R}^{2d}}L(x,w)f_n(dt,dx,dw).$$

Thus,

$$\int_{[0,T]\times\mathbb{R}^d}\Big(\int_{\mathbb{R}^d}L(x,w)f_\infty^{(t,x)}(dw)\Big)\nu(dt,dx)\le\lim_{n\to\infty}\int_{[0,T]\times\mathbb{R}^{2d}}L(x,w)f_n(dt,dx,dw).$$

Since $L(x,\cdot)$ is convex, we apply Jensen's inequality and use the fact that $f_\infty^{(t,x)}$ is a Borel probability measure to conclude the proof. ∎

We conclude the appendix with a proof of comparison for viscosity solutions of a class of equations that includes (1.3). Let W be a real separable Hilbert space. We assume the following hypothesis.

Hypothesis 7.3.

(i) The function $\hat H: W\times W\to\mathbb{R}$ satisfies

$$|\hat H(X,P)-\hat H(X,Q)|\le C(1+|P|+|Q|)|P-Q|\quad\forall\ P,Q,X\in W \tag{7.4}$$

and

$$|\hat H(X,P)-\hat H(Y,P)|\le\sigma\big(|X-Y|(1+|P|)\big)\quad\forall\ P,X,Y\in W \tag{7.5}$$

for some modulus of continuity σ.

(ii) The function $U_0: W\to\mathbb{R}$ is such that

$$|U_0(X)-U_0(Y)|\le m_1\left(|X-Y|\right)\quad\forall\ X,Y\in W \tag{7.6}$$

for some modulus of continuity m_1.

We note that if $\hat{H}(X,P) = \tilde{H}(X, \mathrm{law}(X), P) + F(X)$ and Hypothesis 2.1 is satisfied, then Hypothesis 7.3 holds. Thus comparison for viscosity solutions of (1.3) follows from the more general theorem below.

THEOREM 7.4. *Let Hypothesis* 7.3 *be satisfied and let* $\kappa \geq 0$. *Let* u *be a viscosity subsolution of*

$$\begin{cases} \partial_t u - \kappa \sum_{k=1}^{d} \langle D^2 u e_k, e_k \rangle + \hat{H}(X, Du) = 0 \quad \text{in } (0,T) \times W, \\ u(0,X) = U_0(X) \quad \text{on } W, \end{cases} \tag{7.7}$$

let v *be a viscosity supersolution of* (7.7), *and suppose that there exists* $M \geq 0$ *such that*

$$\sup_{(t,X) \in [0,T) \times W} u(t,X) \leq M, \qquad \sup_{(t,X) \in [0,T) \times W} -v(t,X) \leq M.$$

Then $u \leq v$ *on* $[0,T) \times W$.

Proof. The proof is similar to the proof of Theorem 3.3. The main difference is that we have to use a Hilbert space version of the maximum principle for semicontinuous functions, Theorem 3.2 of [27], instead of Theorem 8.3 of [26]. For $\delta > 0$, let φ_δ be the function from Lemma 3.2 applied to the modulus $\sigma_1(s) = (1+T)\sigma(s) + m_1(s) + (2M+1)s$. In particular we have

$$\varphi_\delta(1) \geq 2M+1, \quad \varphi_\delta(s) \geq m_1(s). \tag{7.8}$$

If $u \not\leq v$, then there is $\nu > 0$ such that

$$\sup_{(t,X) \in [0,T) \times E} (u - v) \geq \nu. \tag{7.9}$$

We will show that this leads to a contradiction.

Let $\{\eta_1, \eta_2, \ldots\}$ be an orthonormal basis of W. For $N = 1, 2, \ldots$, we denote by P_N the orthogonal projection in W onto $\mathrm{span}\{\eta_1, \ldots, \eta_N\}$, and we set $Q_N = I - P_N$. Denote $h(X) := (1 + |X|^2)^{1/2}$. If (7.9) is true, then for sufficiently small $\mu, \gamma, \alpha > 0$

$$\sup_{X,Y \in W, t \in [0,T]} \left(u(t,X) - u_n(t,Y) - \frac{\mu}{T-t} - \varphi_\delta((\gamma + |X-Y|^2)^{\frac{1}{2}})(1+t) - \alpha(h(X) + h(Y)) \right) > 0.$$

We also notice that the expression above goes to $-\infty$ as $|X| + |Y| \to +\infty$. Therefore, by the perturbed optimization result of Ekeland and Lebourg (see, for instance, [29, Theorem 3.25]), for every $n \geq 1$ there exist $a_n \in \mathbb{R}, p_n, q_n \in W$ such that $|a_n| + |p_n| + |q_n| < \frac{1}{n}$ and

$$\begin{aligned} u(t,X) - u_n(t,Y) - \frac{\mu}{T-t} - \varphi_\delta((\gamma + |X-Y|^2)^{\frac{1}{2}})(1+t) - \alpha(h(X) + h(Y)) \\ + a_n t + \langle p_n, X \rangle + \langle q_n, Y \rangle \end{aligned}$$

attains a strict maximum at some point $(\bar{t}, \bar{X}, \bar{Y})$. By the construction of φ_δ we have $0 < \bar{t} < T$ and $|\bar{X} - \bar{Y}| < 1$. It now follows from Theorem 3.2 of [27], together with Remarks 2.3 and 3.1 there, that for every $N \geq 1$ there exist $b_1, b_2 \in \mathbb{R}, S_N, R_N \in S(W)$ and $C > 0$ independent of N such that $S_N = P_N S_N P_N, R_N = P_N R_N P_N, S_N \leq R_N$ and such that, denoting $\bar{s} = (\gamma + |\bar{X} - \bar{Y}|^2)^{\frac{1}{2}}$,

$$\left(b_1, \varphi_\delta'(\bar{s}) \frac{\bar{X} - \bar{Y}}{\bar{s}} (1 + \bar{t}) + \alpha Dh(\bar{X}) - p_n, S_N + CQ_N + \alpha D^2 h(\bar{X}) \right) \in \overline{\mathcal{P}}^{2,+} u(\bar{t}, \bar{X}),$$

$$\left(b_2, \varphi_\delta'(\bar{s})\frac{\bar{X}-\bar{Y}}{\bar{s}}(1+\bar{t}) - \alpha Dh(\bar{X}) + p_n, R_N - CQ_N - \alpha D^2 h(\bar{X})\right) \in \overline{\mathcal{P}}^{2,-} v(\bar{t}, \bar{Y}),$$

$$b_1 - b_2 = \varphi_\delta(\bar{s}) + \frac{\mu}{(T-\bar{t})^2} - a_n.$$

Using the definition of viscosity subsolution we now have

$$b_1 - \kappa \sum_{k=1}^{d} \langle (S_N + CQ_N + \alpha D^2 h(\bar{X}))e_k, e_k\rangle + \hat{H}\left(\bar{X}, \varphi_\delta'(\bar{s})\frac{\bar{X}-\bar{Y}}{\bar{s}}(1+\bar{t}) + \alpha Dh(\bar{X}) - p_n\right) \le 0,$$

which implies, by (7.4),

$$\begin{aligned} b_1 - \kappa \sum_{k=1}^{d} \langle S_N e_k, e_k\rangle + \hat{H}\left(\bar{X}, \varphi_\delta'(\bar{s})\frac{\bar{X}-\bar{Y}}{\bar{s}}(1+\bar{t})\right) \\ \le \sigma_2\left(\frac{1}{N}\right) + \sigma_3\left(\frac{1}{n}\right) + \sigma_4(\alpha) \end{aligned} \tag{7.10}$$

for some moduli $\sigma_2, \sigma_3, \sigma_4$. Similarly we have

$$\begin{aligned} b_2 - \kappa \sum_{k=1}^{d} \langle R_N e_k, e_k\rangle + \hat{H}\left(\bar{Y}, \varphi_\delta'(\bar{s})\frac{\bar{X}-\bar{Y}}{\bar{s}}(1+\bar{t})\right) \\ \ge \sigma_2\left(\frac{1}{N}\right) + \sigma_3\left(\frac{1}{n}\right) + \sigma_4(\alpha). \end{aligned} \tag{7.11}$$

Subtracting (7.11) from (7.10) and using $S_N \le R_N$, (7.5), we obtain

$$\begin{aligned} \sigma_2\left(\frac{1}{N}\right) + \sigma_3\left(\frac{1}{n}\right) + \sigma_4(\alpha) &\ge \varphi_\delta(\bar{s}) + \frac{\mu}{(T-\bar{t})^2} \\ &\quad + \hat{H}\left(\bar{X}, \varphi_\delta'(\bar{s})\frac{\bar{X}-\bar{Y}}{\bar{s}}(1+\bar{t})\right) - \hat{H}\left(\bar{Y}, \varphi_\delta'(\bar{s})\frac{\bar{X}-\bar{Y}}{\bar{s}}(1+\bar{t})\right) \\ &\ge \varphi_\delta(\bar{s}) + \frac{\mu}{(T-\bar{t})^2} - \sigma\left(\bar{s}(1+\varphi_\delta'(\bar{s})(1+T))\right) \\ &\ge \frac{\mu}{(T-\bar{t})^2} + \varphi_\delta(\bar{s}) - \sigma_1(\varphi_\delta'(\bar{s})\bar{s} + \bar{s}) \ge \frac{\mu}{T^2}, \end{aligned} \tag{7.12}$$

where we have used the definition of σ_1 and Lemma 3.2 to justify the last two inequalities. Inequality (7.12) yields a contradiction after we send $N \to +\infty$, then $n \to +\infty$, and finally $\alpha \to 0$. ∎

Acknowledgment. The authors would like to thank A. Mészáros for his input on the first draft of the manuscript.

REFERENCES

[1] L. Ambrosio and J. Feng, *On a class of first order Hamilton-Jacobi equations in metric spaces*, J. Differential Equations, 256 (2014), pp. 2194–2245.

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Math. Monogr., Oxford University Press, New York, 2000.

[3] L. Ambrosio, N. Gigli, and G. Savare, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed., Birkäuser, Basel, 2008.
[4] E. Bandini, A. Cosso, M. Fuhrman, and H. Pham, *Backward SDEs for optimal control of partially observed path-dependent stochastic systems: A control randomization approach*, Ann. Appl. Probab., 28 (2018), pp. 1634–1678.
[5] E. Bandini, A. Cosso, M. Fuhrman, and H. Pham, *Randomized filtering and Bellman equation in Wasserstein space for partial observation control problem*, Stoch. Process. Appl., 129 (2019), pp. 674–711.
[6] A. Bensoussan, J. Frehse, and S. C. P. Yam, *Mean field games and mean field type control theory*, Springer Briefs Math., Springer, New York, 2013.
[7] A. Bensoussan, J. Frehse, and S. C. P. Yam, *The master equation in mean field theory*, J. Math. Pures Appl. (9), 103 (2015), pp. 1441–1474.
[8] A. Bensoussan, J. Frehse, and S. C. P. Yam, *On the interpretation of the Master Equation*, Stochastic Process. Appl., 127 (2017), pp. 2093–2137.
[9] A. Bensoussan, P. J. Graber, and S. C. P. Yam, *Control on Hilbert Spaces and Application to Mean Field Type Control Theory*, preprint, https://arxiv.org/abs/2005.10770, 2020.
[10] A. Bensoussan and S. C. P. Yam, *Control problem on space of random variables and master equation*, ESAIM Control Optim. Calc. Var., 25 (2019), Art. 10, 36 pp.
[11] U. Bessi, *Existence of solutions of the master equation in the smooth case*, SIAM J. Math. Anal., 48 (2016), pp. 204–228.
[12] U. Bessi, *The stochastic value function on metric measure spaces*, Discrete Contin. Dyn. Syst., 37 (2017), pp. 1819–1839.
[13] U. Bessi, *Hamilton-Jacobi in metric spaces with a homological term*, Comm. Partial Differential Equations, 45 (2020), no. 7, pp. 776–819.
[14] R. Buckdahn, J. Li, S. Peng, and C. Rainer, *Mean-field stochastic differential equations and associated PDEs*, Ann. Probab., 45 (2017), pp. 824–878.
[15] P. Cardialaguet, *Notes on Mean-Field Games*, https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf, 2013.
[16] P. Cardaliaguet, *The convergence problem in mean field games with local coupling*, Appl. Math. Optim., 76 (2017), pp. 177–215.
[17] P. Cardaliaguet, M. Cirant, and A. Porretta, *Splitting Methods and Short Time Existence for the Master Equations in Mean Field Games*, preprint, https://arxiv.org/abs/2001.10406, 2020.
[18] P. Cardaliaguet, F. Delarue, J. M. Lasry, and P. L. Lions, *The Master Equation and the Convergence Problem in Mean Field Games*, Ann. of Math. Stud. 201, Princeton University Press, Princeton, NJ, 2019.
[19] P. Cardaliaguet and A. Porretta, *Long time behavior of the master equation in mean field game theory*, Anal. PDE, 12 (2019), pp. 1397–1453.
[20] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications* I: *Mean Field FBSDEs, Control, and Games*, Probab. Theory Stoch. Model. 83. Springer, Cham, 2018.
[21] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications* II: *Mean Field Games with Common Noise and Master Equations*, Probab. Theory Stoch. Model. 84. Springer, Cham, 2018.
[22] J. F. Chassagneux, D. Crisan, and F. Delarue, *A Probabilistic Approach to Classical Solutions of the Master Equation for Large Population Equilibria*, preprint, https://arxiv.org/abs/1411.3009, 2014.
[23] J. F. Chassagneux, L. Szpruch, and A. Tse, *Weak Quantitative Propagation of Chaos via Differential Calculus on the Space of Measures*, preprint, https://arxiv.org/abs/1901.02556v1, 2019.
[24] Y. T. Chow and W. Gangbo, *A partial Laplacian as an infinitesimal generator on the Wasserstein space*, J. Differential Equations, 267 (2019), pp. 6065–6117.
[25] A. Cosso and H. Pham, *Zero-sum stochastic differential games of generalized McKean-Vlasov type*, J. Math. Pures Appl. (9), 129 (2019), pp. 180–212.
[26] 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., 27 (1992), pp. 1–67.
[27] M. G. Crandall, M. Kocan, and A. Święch, *On partial sup-convolutions, a lemma of P.L. Lions and viscosity solutions in Hilbert spaces*, Adv. Math. Sci. Appl., 3 (1993/94), pp. 1–15.
[28] F. Delarue, D. Lacker, and K. Ramanan, *From the master equation to mean field game limit theory: A central limit theorem*, Electron. J. Probab., 24 (2019), 51.

[29] G. Fabbri, F. Gozzi, and A. Święch, *Stochastic Optimal Control in Infinite Dimension: Dynamic Programming and HJB Equations*, Probab. Theory Stochastic Model. 82, Springer, Cham, 2017.
[30] J. Feng and M. Katsoulakis, *A comparison principle for Hamilton-Jacobi equations related to controlled gradient flows in infinite dimensions*, Arch. Ration. Mech. Anal., 192 (2009), pp. 275–310.
[31] J. Feng and T. G. Kurtz, *Large Deviations for Stochastic Processes*, Math. Surveys Monographs 131, Providence, RI, 2006.
[32] J. Feng, T. Mikami, and J. Zimmer, *A Hamilton-Jacobi PDE Associated with Hydrodynamic Fluctuations from a Nonlinear Diffusion Equation*, preprint, https://arxiv.org/abs/1903.00052, 2019.
[33] J. Feng and T. Nguyen, *Hamilton-Jacobi equations in space of measures associated with a system of conservation laws*, J. Math. Pures Appl. (9), 97 (2012), pp. 318–390.
[34] J. Feng and A. Święch, *Optimal control for a mixed flow of Hamiltonian and gradient type in space of probability measures*, Trans. Amer. Math. Soc., 365 (2013), pp. 3987–4039.
[35] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, 2nd ed., Stochastic Model. Appl. Probab. 25, Springer, New York, 2006.
[36] W. Gangbo, *On Some Analytical Aspects of Mean Field Games*, https://math.berkeley.edu/~wgangbo/math_278/MFG-1-18.pdf, 2019.
[37] W. Gangbo and A. Mészáros, *Global Well-Posedness of Master Equations for Deterministic Displacement Convex Potential Mean Field Games*, preprint, https://arxiv.org/abs/2004.01660, 2020.
[38] W. Gangbo, T. Nguyen, and A. Tudorascu, *Hamilton-Jacobi equations in the Wasserstein space*, Methods Appl. Anal., 15 (2008), pp. 155–183.
[39] W. Gangbo and A. Święch, *Optimal transport and large number of particles*, Discrete Contin. Dyn. Syst., 34 (2014), pp. 1397–1441.
[40] W. Gangbo and A. Święch, *Metric viscosity solutions of Hamilton-Jacobi equations depending on local slopes*, Calc. Var. Partial Differential Equations, 54 (2015), pp. 1183–1218.
[41] W. Gangbo and A. Święch, *Existence of a solution to an equation arising from the theory of mean field games*, J. Differential Equations 259 (2015), pp. 6573–6643.
[42] W. Gangbo and A. Tudorascu, *Weak KAM theory on the Wasserstein torus with multi–dimensional underlying space*, Comm. Pure Appl. Math., 67 (2014), pp. 408–463.
[43] W. Gangbo and A. Tudorascu, *On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations*, J. Math. Pures Appl. (9), 125 (2019), pp. 119–174.
[44] Y. Giga, N. Hamamuki, and A. Nakayasu, *Eikonal equations in metric spaces*, Trans. Amer. Math. Soc., 367 (2015), pp. 49–66.
[45] R. Hynd and H. K. Kim, *Infinite horizon value functions in the Wasserstein spaces*, J. Differential Equations, 258 (2015), pp. 1933–1966.
[46] R. Hynd and H. K. Kim, *Value functions in the Wasserstein spaces: Finite time horizons*, J. Funct. Anal., 269 (2015), pp. 968–997.
[47] C. Jimenez and M. Quincampoix, *Hamilton Jacobi Isaacs equations for differential games with asymmetric information on probabilistic initial condition*, J. Math. Anal. Appl., 457 (2018), pp. 1422–1451.
[48] R. C. Kraaij, *A General Convergence Result for Viscosity Solutions of Hamilton-Jacobi Equations and Non-Linear Semigroups*, preprint, https://arxiv.org/abs/1903.04196, 2019.
[49] R. C. Kraaij, *The Exponential Resolvent of a Markov Process and Large Deviations for Markov Processes via Hamilton-Jacobi Equations*, preprint, https://arxiv.org/abs/1910.03946, 2019.
[50] P. L. Lions, *Mean Field Games*, Cours au Collège de France, 2007–2011.
[51] Q. Liu and A. Nakayasu, *Convexity preserving properties for Hamilton-Jacobi equations in geodesic spaces*, Discrete Contin. Dyn. Syst., 39 (2019), pp. 157–183.
[52] S. Mayorga, *On a Classical Solution to the Master Equation of a First Order Mean Field Game*, Ph.D. thesis, Georgia Institute of Technology, 2019.
[53] S. Mayorga, *Short time solution to the master equation of a first order mean field game system*, J. Differential Equations, 268 (2020), pp. 6251–6318.
[54] C. Mou and J. Zhang, *Wellposedness of Second Order Master Equations for Mean Field Games with Nonsmooth Data*, preprint, https://arxiv.org/abs/1903.09907, 2020.
[55] A. Nakayasu, *Metric viscosity solutions for Hamilton-Jacobi equations of evolution type*, Adv. Math. Sci. Appl., 24 (2014), pp. 333–351.
[56] A. Nakayasu and T. Namba, *Stability properties and large time behavior of viscosity solutions of Hamilton-Jacobi equations on metric spaces*, Nonlinearity, 31 (2018), pp. 5147–5161.

[57] H. Pham and X. Wei, *Dynamic programming for optimal control of stochastic McKean-Vlasov dynamics*, SIAM J. Control Optim., 55 (2017), pp. 1069–1101.
[58] H. Pham and X. Wei, *Bellman equation and viscosity solutions for mean-field stochastic control problem*, ESAIM Control Optim. Calc. Var., 24 (2018), pp. 437–461.
[59] C. Wu and J. Zhang, *Viscosity solutions to parabolic master equations and McKean-Vlasov SDEs with closed-loop controls*, Ann. Appl. Probab., 30 (2020), pp. 936–986.