FROM PARTICLES IN RANDOM POTENTIALS TO A NONLINEAR VLASOV-FOKKER-PLANCK EQUATION

JIN FENG

Abstract. We consider large time and infinite particle limit for system of particles with random potential functions. The randomness enters the potential through an external ergodic Markov process, modeling oscillating environment with good statistical averaging properties. At large time, the ergodic process in the potential converges to its equilibrium measure and an averaged (with respect to such measure) macroscopic equation for the whole system is derived.

From each individual particle's point of view, both law of large number and central limit theorem type of averaging are possible in this context. Models of this type are known as random evolutions. Instead of one particle, we focus on the collective behavior of infinite particles. We separately rescale potential functions (type one) which annihilates the equilibrium measure of the ergodic environment process, and the potential functions which may not annihilate such measure (type two). Appropriately rescaled to the macroscopic limit, type two potentials give a transport term while type one potentials give a nonlinear diffusion term. The resulting equation is a version of nonlinear Vlasov-Fokker-Planck equation. The solution of such equation is unique and we will verify this using a probabilistic particle representation method.

1. Introduction

We rigorously derive a version of nonlinear Vlasov-Fokker-Planck equation

(1.1)
$$\partial_t \rho(t,x,v) + v \cdot \nabla_x \rho(t,x,v) - (\rho_X * \nabla \bar{\Phi}_1 + \nabla \bar{\Psi}_1)(t,x) \cdot \nabla_v \rho(t,x,v) = a(\rho_X; x) \cdot D^2_{vv} \rho(t,x,v)$$

as multi-scale limit of infinite interacting particles with random potential. See Theorem 6.9.

In equation (1.1), $x, v \in R^d$, D^2_{vv} is the Hessian matrix where the derivatives are only taken with respect to v; a is some square matrix specified in (1.19) using potential functions defining the dynamic at the microscopic level; and by $M \cdot N$ for two $d \times d$ matrices $M = (m_{ij})_{d\times d}$ and $N = (n_{ij})_{d\times d}$, we mean

$$M \cdot N \equiv \sum_{i,j=1}^{d} m_{ij} n_{ji}. \tag{1.2}$$

The whole equation (1.1) is understood in the weak (Schwartz distributional) sense, where in particular $\rho(t,x,v)$ is understood as a probability measure in x, v-variables, and ρ_X is the x-marginal probability measure of ρ (i.e. $\rho_X(dx) = \rho(dx, R^d)$).

We prove that probability measure valued solution for (1.1) is unique. This is proved in Section 6 using a probabilistic method (Theorem 6.8) known as the particle representation method.

Date: Sept. 28, 2005.

I thank Thomas G. Kurtz for explaining the particle representation method to me.

1.1. **The microscopic model.** Let $x = (x_1, \dots, x_N)$ denote the position of N-particles, $N = 1, 2, \dots$. Each $x_i \in R^d$. Similarly, we also use v_i to denote the velocity of the i-th particle. We assume that each particle has unit mass. Let $\Phi : R^d \mapsto R$ be an even function, modeling pair-wise interaction potential between two particles. In the absence of external force, the N-particles $i = 1, 2, \dots, N$ follows Newton's law

$$\begin{aligned} \dot{x}_i(t) &= v_i(t) \\ \dot{v}_i(t) &= -\frac{1}{N}\sum_{j=1}^{N}(\nabla\Phi)\Big(x_i(t) - x_j(t)\Big). \end{aligned} \tag{1.3}$$

In this article, we consider a situation where the interaction potential is subject to influence by a random environment factor. Let S be a compact metric space modeling the state of a particle in the random environment. When particles i and j are in their respective "environmental" state $y_i, y_j \in S$, their interaction potential function is modeled using $\Phi(z; y_i, y_j) : (z, y_i, y_j) \in R^d \times S \times S \mapsto R$. Out of physical considerations, we assume that these functions are even in the z-variable

$$\Phi(z; y_1, y_2) = \Phi(z; y_2, y_1), \quad \Phi(z; y_1, y_2) = \Phi(-z; y_1, y_2). \tag{1.4}$$

We only consider the case where the random environment is provided by a large external Markov process $Y(t) = (Y_1(t), \dots, Y_N(t)) \in S^N$ with generator $\mathbf{B}$. Generalizations to stationary non-Markovian process is possible but will not be pursued here. To free us from possible complications caused by boundary conditions in the x-variable, we also introduce an external potential $\Psi(x; y)$, acting on each individual particle at location x when the environment is in state y. The defining system of equations for the N-particles then becomes

$$\begin{aligned} \dot{x}_i(t) &= v_i(t) \\ \dot{v}_i(t) &= -\frac{1}{N}\sum_{j=1}^{N}(\nabla\Phi_n)\Big(x_i(t) - x_j(t); Y_i(nt), Y_j(nt)\Big) - (\nabla\Psi_n)\Big(x_i(t); Y_i(nt)\Big) \end{aligned} \tag{1.5}$$

where $i = 1, 2, \dots, N$. We adapt the following notational convention throughout this article:

$$\nabla\Phi(x; y_1, y_2) = \nabla_x\Phi(x; y_1, y_2), \quad \nabla\Psi(x; y) = \nabla_x\Psi(x; y). \tag{1.6}$$

The n in $Y_i(nt)$ can be viewed as a scaling parameter which measures the degree of separation of time scales between the environment process and the particle process. The above system of equations is a convienient way of modeling "random exchange of momentum" type phenomenon which is typical in plasma physics, statistical physics, among other applications.

To explain (1.5), we introduce a family of energy functions indexed by $\vec{y} = (y_1, y_2, \dots, y_N)$

$$\mathcal{E}_{N,\vec{y}} = \frac{1}{2}\sum_{i=1}^{N}|v_i|^2 + \sum_{i=1}^{N}\Psi_n(x_i; y_i) + N^{-1}\sum_{i,j=1}^{N}\Phi_n(x_i - x_j; y_i, y_j).$$

The dynamic given by (1.5) can be viewed as a mixture of two motions: First, a random hopping among the different energy surfaces dictated by the large Markov processes $(Y_1(nt), \dots, Y_N(nt))$; Second, the deterministic motion on the energy surface it is currently on, the energy function is constant along the trajectory of this motion.

We are interested in large time behavior of the system. Therefore, we introduce the following further rescaling: let the new position and velocity vectors for the i-th particle be

$x_i^{(n)}(t) = x_i(nt)$ and $v_i^{(n)}(t) = \dot{x}_i^{(n)}(t) = nv_i(nt)$. Then

$$\begin{aligned} \dot{x}_i^{(n)}(t) &= v_i^{(n)}(t) \\ \dot{v}_i^{(n)}(t) &= n\Big\{-\frac{1}{N}\sum_{j=1}^{N}(\nabla\Phi_n)\Big(x_i^{(n)}(t)-x_j^{(n)}(t);Y_i(n^2t),Y_j(n^2t)\Big) \\ &\qquad -\nabla\Psi_n(x_i^{(n)}(t);Y_i(n^2t))\Big\}. \end{aligned} \tag{1.7}$$

Under appropriate smoothness and growth conditions on Ψ_n, Φ_n and regularities on Y, the above equation has a unique solution.

1.2. **Structural assumptions.** To highlight ideas, we only consider the setting where all Y_is are independent Markov process with weak infinitesimal generator B in $C_b(S)$ in the following sense. For bounded measurable function f, let

$$S(t)f(y) = E[f(Y_i(t))|Y_i(0) = y].$$

By the Markov property of Y_i, we have

$$S(t+r) = S(t)S(r), \quad t, r \geq 0.$$

We assume that $S(t) : C_b(S) \mapsto C_b(S)$. We define the domain $D(B)$ of B to be functions such that

$$\limsup_{t\to 0}\sup_{y\in S} t^{-1}|S(t)f(y) - f(y)| < +\infty$$

and that the limit

$$Bf(y) = \lim_{t\to 0+} t^{-1}(S(t)f(y) - f(y))$$

exists. We define Bf as the above limit.

Similarly, we define generator $\mathbf{B}$ for $Y = (Y_1, \dots, Y_N)$. The relation between B and $\mathbf{B}$ is

$$\mathbf{B}\varphi(y_1, \dots, y_N) = \sum_{k=1}^{m} B\varphi(y_1, \dots, y_{i-1}, \cdot, y_{i+1}, \dots, y_N)(y_i).$$

whenever $\varphi(y_1, \dots, y_{i-1}, \cdot, y_{i+1}, \dots, y_N) \in D(B)$ and $B\varphi(y_1, \dots, y_{i-1}, \cdot, y_{i+1}, \dots, y_N)(y_i) \in C_b(S^N)$.

We assume that Y_i has the following ergodic properties.

Condition 1.1. *Let Y be the Markov process on a compact metric space S with weak infinitesimal generator B.*

1. *Y has a unique stationary probability measure $\pi_0(dy)$ in the sense that*

$$\lim_{t\to+\infty} S(t)f(x) = \lim_{t\to+\infty} E[f(Y(t))|Y(0) = x] = \int f(y)\pi_0(dy) \equiv P_{\pi_0}f, \quad \forall x \in S.$$

2. *for each bounded measurable function h, there exists constant $C_h > 0$ such that*

$$|S(t)h(x) - \pi_0 h(x)| \leq C_h(1+t^2)^{-1}.$$

Typical examples satisfying the above requirements are random walks (continuous time) with a communication condition (in the sense that there is positive probability to reach each point from any other point), Brownian motions on compact manifold, etc, etc. By the second part of Condition 1.1, the measure

$$\nu(y, dz) = \int_0^\infty (P(Y(t) \in dz | Y(0) = y) - \pi_0(dz))dt$$

is well defined. Let $\varphi \in B(S)$. We will write

$$(P_\nu \varphi)(y) = \int \varphi(z)\nu(y, dz).$$

Then using the semigroup property, it follows that for φ satisfying $P_{\pi_0}\varphi = 0$,

$$BP_\nu\varphi = -\varphi. \tag{1.8}$$

For general $\varphi \in C_b(S)$, since $BP_{\pi_0}\varphi = 0$ and $P_{\pi_0}(\varphi - P_{\pi_0}\varphi) = 0$, consequently

$$BP_\nu\varphi(y) + \varphi(y) = P_{\pi_0}\varphi. \tag{1.9}$$

Similarly, we can state a multi-variate version of the above result using $\mathbf{B}$. Let $\varphi = \varphi(y_1, \ldots, y_m) \in C_b(S^m)$, define

$$(P_{\nu\otimes\ldots\otimes\nu}\varphi)(y_1, \ldots, y_m) = \int_S \ldots \int_S \varphi(z_1, \ldots, z_m)\nu(y_1, dz_1)\ldots\nu(y_m, dz_m).$$

Then

$$\begin{aligned} (1.10) \qquad & \mathbf{B}P_{\nu\otimes\ldots\otimes\nu}\varphi(y_1, \ldots, y_m) + \varphi(y_1, \ldots, y_m) \\ = \; & B_{y_1}P_{\nu\otimes\ldots\otimes\nu}\varphi(y_1, \ldots, y_m) + \ldots + B_{y_m}P_{\nu\otimes\ldots\otimes\nu}\varphi(y_1, \ldots, y_m) + \varphi(y_1, \ldots, y_m) \\ = \; & P_{\pi_0\otimes\ldots\otimes\pi_0}\varphi. \end{aligned}$$

The following notation will be useful. For a smooth function f on R^d, we denote

$$|D^k f(x)| = \sum_{i_1+\ldots+i_d=k} |\frac{\partial^k}{\partial x_1^{i_1}\ldots\partial x_d^{i_d}} f(x)|, \quad x \in R^d.$$

Without pursuing generality, we assume the following working conditions on potential functions throughout the article. We also assume a scaling relation between n and N.

Condition 1.2.

1. *Φ_n and Ψ_n satisfy*

$$\Phi_n(x; y_1, y_2) = n^{-1}\Phi_1(x; y_1, y_2) + \Phi_2(x; y_1, y_2), \quad \Psi_n(x; y) = n^{-1}\Psi_1(x; y) + \Psi_2(x; y) \tag{1.11}$$

where

$$\int_{S\times S} \nabla\Phi_2(x; y_1, y_2)\pi_0(dy_1)\pi_0(dy_2) = 0, \quad \int_S \nabla\Psi_2(x, y)\pi_0(dy) = 0 \tag{1.12}$$

(recall (1.6) for the notation $\nabla\Phi_2, \nabla\Psi_2$); and

2. *$\nabla\Phi_1(z; y_1, y_2), \nabla\Phi_2(z; y_1, y_2) \in C_b(R^d \times S \times S)$ and*

$$\sup_{z\in R^d, y_1, y_2\in S} (|D^1\Phi_2(z, y_1, y_2)| + |D^2\Phi_2(z, y_1, y_2)|) < +\infty,$$

where the $D^k, k = 1, 2$ applies to the z-variable; and

3. $\Phi_1, \Psi_1 \geq 0$, $\nabla \Psi_2(z;y) \in C_b(R^d \times S)$; *and for each* $x \in R^d$, $\nabla \Psi_1(x;\cdot) \in C_b(S)$, $\bar{\Psi}_1 \in C(R^d)$, *and*

$$\lim_{M \to +\infty} \inf_{|z|>M} \bar{\Psi}_1(z) = +\infty.$$

See (1.14) for definition of $\bar{\Psi}_1$.

4. *The number of particles* N *and the multiple time scale parameter* n *satisfy*

$$\lim_{n \to +\infty} nN^{-1} = 0. \tag{1.13}$$

From now on, we write

$$\bar{\Phi}_1(x) = \int_S \int_S \Phi_1(x; y_1, y_2)\pi_0(dy_1)\pi_0(dy_2), \quad \bar{\Psi}_1(x) = \int_{y \in S} \Psi_1(x, y)\pi_0(dy). \tag{1.14}$$

1.3. **Main result.** The symmetry among particle labels suggests that we can identify the $(x_i(\cdot), v_i(\cdot), Y_i(n^2\cdot))$-system with its empirical measure without any lose of generality:

$$\gamma_n(t, dx, dv, dy) \equiv \frac{1}{N}\sum_{i=1}^N \delta_{\{x_i^{(n)}(t), v_i^{(n)}(t), Y_i(n^2t)\}}(dx, dv, dy). \tag{1.15}$$

Let N be fixed and $n \to +\infty$, the $Y = (Y_1, \ldots, Y_N)$ dependence in (1.7) can be averaged out in at least two ways – law of large number and central limit theorem. In the central limit scaling, it is known that $(x_1^{(n)}, \cdots, x_N^{(n)}; v_1^{(n)}, \cdots, v_N^{(n)})$ converges to a R^{2d}-valued stochastic diffusion given by Ito's equation. The law of large number is simply an averaged ODE with the Y_i variables replaced by integration with respect to their stationary measures. Such problems are called random evolution, see Chapter 12.4 of Ethier and Kurtz [3]. Also see the end of that chapter for a review of history for random evolution problems. On the other hand, if we directly look at these weakly interacting stochastic diffusion equations (we can view these as the limiting case of $n = +\infty$ with N finite but fixed), and then pass N to infinity for the empirical measures (1.15), it is known that we will end up with some deterministic partial differential equation describing transport plus diffusion behaviors. Such limiting procedure is known as the McKean-Vlasov limit (e.g. Dawson and Gartner [2]).

The basic message of this article is that (1.1) can be viewed as a limit of (1.7) when both n and N go to infinity at the same time, with appropriate speed. We provide rigorous justification of such view through an infinite dimensional generalization of the techniques used for random evolution problems as discussed in Ethier and Kurtz [3] using martingale problem method.

In the above discussed limit, we expect the dy component in γ_n to be averaged out. Therefore, we introduce notation ρ_n to denote the (X, V)-marginal probability measure of γ_n:

$$\rho_n(t, dx, dv) = \gamma_{n,X,V}(t, dx, dv) = \gamma_n(t, dx, dv, S). \tag{1.16}$$

Similarly, we denote the X-marginal of $\rho_n(t, dx, dv)$ by

$$\rho_{n,X}(t, dx) \equiv \rho_n(t, dx, R^d). \tag{1.17}$$

More generally, for any probability measure $\gamma \in \mathcal{P}(R^d \times R^d \times S)$, we denote $\rho(dx, dv) = \gamma(dx, dv, S)$ and $\rho_X(dx) = \rho(dx, R^d)$. In other words, ρ will always be used as the x, v-marginal of corresponding measure γ which has three arguments x, v, y.

In Theorem 4.4, we show that $\{\rho_n(\cdot) : n = 1, 2, \dots\}$ is a tight sequence as stochastic processes with trajectories in the space $C_{\mathcal{P}(R^d \times R^d)}[0, +\infty)$. In Theorem 5.1, we conclude that any limit point of the above sequence is a solution to the stochastic Vlasov-Fokker-Planck equation (1.1). Finally, with additional smoothness hypothesis on $\Phi_i, \Psi_i, i = 1, 2$, in Theorem 6.9, we show that solution to (1.1) is unique, hence ρ_n converges to the solution of (1.1).

We specify the coefficients in (1.1) next: Recall the definition of $\bar{\Phi}_1$ and $\bar{\Psi}_1$ in (1.14), we let

$$\bar{\bar{\Phi}}_2(x; y) = \int_{\bar{y} \in S} \int_{z \in S} \Phi_2(x; y, z) \nu(\bar{y}, dz) \pi_0(d\bar{y}); \tag{1.18}$$

and define the $d \times d$ square matrix $a(\rho, x) = (a_{ij}(\rho, x))_{i,j=1,\dots,d}$ to be

$$\begin{aligned} a_{ij}(\rho, x) &= a_{ij}(\rho_X, x) \\ &= -\frac{1}{2} P_{\pi_0} \Big\{ \Big(P_\nu(\rho_X * \nabla^{(i)} \bar{\bar{\Phi}}_2(x; y) + \nabla^{(i)} \Psi_2(x; y)) \Big) \\ &\qquad B_y \Big(P_\nu(\rho_X * \nabla^{(j)} \bar{\bar{\Phi}}_2(x; y) + \nabla^{(j)} \Psi_2(x; y)) \Big) \\ &\qquad + \Big(P_\nu(\rho_X * \nabla^{(j)} \bar{\bar{\Phi}}_2(x; y) + \nabla^{(j)} \Psi_2(x; y)) \Big) \\ &\qquad B_y \Big(P_\nu(\rho_X * \nabla^{(i)} \bar{\bar{\Phi}}_2(x; y) + \nabla^{(i)} \Psi_2(x; y)) \Big) \Big\} \\ &= \mathcal{E}(P_\nu(\rho_X * \nabla^{(i)} \bar{\bar{\Phi}}_2(x; \cdot) + \nabla^{(i)} \Psi_2(x; \cdot))), \\ &\qquad P_\nu(\rho_X * \nabla^{(j)} \bar{\bar{\Phi}}_2(x; \cdot) + \nabla^{(j)} \Psi_2(x; \cdot)))) \end{aligned} \tag{1.19}$$

where $\mathcal{E}$ is the Dirichlet form associated with B:

$$\mathcal{E}(f, g) = \frac{1}{2} \int (f B g + g B f) d\pi_0, \quad f, g \in D(B).$$

By symmetry property of $\mathcal{E}$, $a(\rho; x)$ is a non-negative definite matrix.

Our program for the proof is described as follows. First, $\gamma_n(\cdot)$ is a measure valued Markov process. We identify its infinitesimal generator A_n through the associated martingale problem (in the sense of Stroock and Varadhan [9], see also Ethier and Kurtz [3]): for a class of test functions f_n (with precise form to be identified later),

$$f_n(\gamma_n(t)) - f_n(\gamma_n(0)) - \int_0^t A_n f_n(\gamma_n(s)) ds = M_n^{f_n}(t)$$

where $M_n^{f_n}$ is a martingale. Second, we show that for a sufficiently large class of smooth test function of the form $f(\gamma) = f(\rho)$ (i.e. f only depends on γ through its X, V-marginals), we find $f_n \in D(A_n)$,

$$f_n(\gamma) = f(\rho) + n^{-1} g(\gamma) + n^{-2} h(\gamma),$$

such that there is a limit operator A with domain containing functions fs, in the sense that

$$A_n f_n(\gamma_n) \to A f(\rho)$$

whenever the X, V-marginal of γ_n converges to ρ in weak convergence of probability measure sense. Third, we show that $\{\gamma_n(\cdot) : n = 1, 2, \dots\}$ is a relatively compact sequence. Then, putting every step together by passing $n \to +\infty$, we have that for any limit point ρ of the relative compact sequence from ρ_n, it satisfies

$$f(\rho(t)) - f(\rho(0)) - \int_0^t Af(\rho(s))ds = 0.$$

We show that this is just another way to express the weak (Schwartz distributional) solution of (1.1). Finally, we demonstrate such solution is unique and consequently conclude that the previous convergence along subsequence claim is true for convergence along the original sequence.

1.4. **Notations.** For any two vectors $\xi = (\xi_1, \dots, \xi_d), \eta = (\eta_1, \dots, \eta_d) \in R^d$, we define matrix

$$\xi \times \eta = (\xi_i \eta_j)_{ij}. \tag{1.20}$$

If E_0 is a metric space, we use $B(E_0)$ to denote bounded measurable functions, $C(E_0)$ continuous functions, $C_b(E_0)$ bounded continuous functions, $\mathcal{P}(E_0)$ the space of probability measures on E_0, and $M_\pm(E_0)$ the space of signed Borel measures on E_0. Let $f \in B(E_0)$, we denote $\|f\| = \sup_{x \in E_0} |f(x)|$. For a function $g : \gamma \in M_\pm(E_0) \mapsto R$, we define its first and second variational derivatives as functions

$$\frac{\delta g}{\delta \gamma} : E_0 \mapsto R, \quad \frac{\delta^2 g}{\delta \gamma^2} : E_0 \times E_0 \mapsto R$$

satisfying the Taylor's expansion

$$g(\gamma + t\gamma') - g(\gamma) = t\langle \frac{\delta g}{\delta \gamma}, \gamma' \rangle + \frac{1}{2} t^2 \langle \frac{\delta^2 g}{\delta \gamma^2}, \gamma' \otimes \gamma' \rangle + o(t^2) \tag{1.21}$$

for each $\gamma' \in M_\pm(E_0)$ with compact support. In the above, $\gamma' \otimes \gamma'$ means the product measure on $E_0 \times E_0$.

As an example, we consider test function

$$f(\gamma) = \psi(\langle \varphi_1, \gamma \rangle, \dots, \langle \varphi_m, \gamma \rangle) : \quad \varphi_k \in B(R^d \times R^d \times S), \psi \in C^2(R^m) \tag{1.22}$$

for γ a signed measure on $R^d \times R^d \times S$. Then

$$\frac{\delta f}{\delta \gamma}(x, v; y) = \sum_{k=1}^m \partial_k \psi(\langle \varphi_1, \gamma \rangle, \dots, \langle \varphi_m, \gamma \rangle) \varphi_k(x, v, y),$$

and

$$\frac{\delta^2 f}{\delta \gamma^2}(x, v, y; \bar{x}, \bar{v}, \bar{y}) = \sum_{k,l=1}^m \partial^2_{kl} \psi(\langle \varphi_1, \gamma \rangle, \dots, \langle \varphi_m, \gamma \rangle) \varphi_k(x, v, y) \varphi_l(\bar{x}, \bar{v}, \bar{y}).$$

We also consider another class of test functions. To simplify notation we consider the general case where E_0 is a metric space. Our main interest is when $E_0 = R^d \times R^d \times S$. Let

$$f(\gamma) = \int_{E_0} \dots \int_{E_0} \varphi(x_1, x_2, x_3, x_4) \gamma(dx_1) \gamma(dx_2) \gamma(dx_3) \gamma(dx_4) : M_\pm(E_0) \mapsto R,$$

where φ is bounded and continuous. Then

$$\begin{aligned}\frac{\delta f}{\delta\gamma}(x) &= \int_{E_0}\int_{E_0}\int_{E_0}\varphi(x,x_2,x_3,x_4)\gamma(dx_2)\gamma(dx_3)\gamma(dx_4)\\ &\quad+\int_{E_0}\int_{E_0}\int_{E_0}\varphi(x_1,x,x_3,x_4)\gamma(dx_1)\gamma(dx_3)\gamma(dx_4)\\ &\quad\quad+\int_{E_0}\int_{E_0}\int_{E_0}\varphi(x_1,x_2,x,x_4)\gamma(dx_1)\gamma(dx_2)\gamma(dx_4)\\ &\quad\quad\quad+\int_{E_0}\int_{E_0}\int_{E_0}\varphi(x_1,x_2,x_3,x)\gamma(dx_1)\gamma(dx_2)\gamma(dx_3),\end{aligned}$$

Taking another variational derivative, the expression of $(\delta^2/\delta\gamma^2)f(x,\bar{x})$ can be similarly expressed. It is involved with more terms.

Throughout this paper, we denote

$$(1.23)\quad E_n' = \{\gamma(dx,dv,dy) = N^{-1}\sum_{i=1}^{N}\delta_{\{x_i,v_i,y_i\}}(dx,dv,dy) : x_i, v_i \in R^d, y_i \in S\},$$

and

$$(1.24)\qquad E_n = \{\rho(dx,dv) = \gamma(dx,dv;S) : \gamma \in E_n'\}.$$

We also denote $E = \mathcal{P}(R^d \times R^d)$ the space of probability measures on $R^d \times R^d$ with a typical element written $\rho = \rho(dx,dv)$. Similarly, we denote $E' = \mathcal{P}(R^d \times R^d \times S)$ with a typical element $\gamma = \gamma(dx,dv,dy)$. Unless specified otherwise, the topology on these spaces we use is always the weak convergence of probability measure topology.

We note that for $\rho \in E_n$, it has moments upto all orders (both in x and in v variables).

For an operator B, $D(B)$ denote the domain of B. If $f = f(x,v,y) : R^d \times R^d \times S \mapsto R$ and $f(x,v,\cdot) \in D(B)$, we write $B_y f(x,v,y) = (Bf(x,v,\cdot))(y)$.

For any sequence of problem measures ρ_n, ρ on a complete separable metric space, $\rho_n \Rightarrow \rho$ denote convergence under the weak convergence of probability measure topology.

2. Martingale problem

First, we identify A_n for a relatively simpler class of test functions (1.22). Later, we will find that another type of test functions is also needed. The procedure explained below can be then analogously applied to derive the extension to the new test functions.

We let $\mathcal{F}_t^n = \sigma(Y_i(r) : n = 1,2,\ldots,N; 0 \le r \le n^2 t\}$. Let $\varphi = \varphi(x,v;y), \nabla_x\varphi, \nabla_v\varphi \in C(R^d \times R^d \times S)$, and $\varphi(x,v,\cdot) \in D(B)$ and $B_y\varphi(x,v,y)$ is bounded continuous. Then for every $t \ge s \ge 0$,

$$\begin{aligned}&\varphi(x_i^{(n)}(t), v_i^{(n)}(t); Y_i(n^2t)) - \varphi(x_i^{(n)}(s), v_i^{(n)}(s); Y_i(n^2t))\\ =&\int_s^t \Big\{ v_i^{(n)}(r)\cdot\nabla_x\varphi(x_i^{(n)}(r), v_i^{(n)}(r); Y_i(n^2t))\\ &\quad -n\Big(\int_{R^d\times R^d\times S}\nabla\Phi_n(x_i^{(n)}(r)-\bar{x}; Y_i(n^2r),\bar{y})\gamma_n(r; d\bar{x}, d\bar{v}, d\bar{y})\\ &\qquad\quad +\nabla\Psi_n(x_i^{(n)}(r); Y_i(n^2r))\Big)\cdot\nabla_v\varphi(x_i^{(n)}(r), v_i^{(n)}(r), Y_i(n^2t))\Big\}dr;\end{aligned}$$

and

$$\begin{aligned} & E[\varphi(x_i^{(n)}(s), v_i^{(n)}(s), Y_i(n^2t)) - \varphi(x_i^{(n)}(s), v_i^{(n)}(s), Y_i(n^2s))|\mathcal{F}_s^n] \\ = \ & E[\int_s^t n^2 B\varphi(x_i^{(n)}(s), v_i^{(n)}(s), Y_i(n^2r))dr|\mathcal{F}_s^n]. \end{aligned}$$

Therefore, by Lemma 4.3.4 of Ethier and Kurtz [3],

$$\begin{aligned} \varphi(x_i^{(n)}(t), v_i^{(n)}(t), Y_i(n^2t)) - \int_0^t \Big(& v_i^{(n)}(r)\nabla_x\varphi(x_i^{(n)}(r), v_i^{(n)}(r), Y_i(n^2r))dr \\ & -n(\int_{R^d\times R^d\times S} \nabla\Phi_n(x_i^{(n)}(r) - x; Y_i(n^2r), y)\gamma_n(r; dx, dv, dy) \\ & +\nabla\Psi_n(x_i^{(n)}(r); Y_i(n^2r))) \cdot \nabla_v\varphi(x_i^{(n)}(r), v_i^{(n)}(r), Y_i(n^2r)) \\ & -n^2 B\varphi(x_i^{(n)}(r), v_i^{(n)}(r), Y_i(n^2r))\Big)ds \equiv M_i^{\varphi}(t); \end{aligned}$$

is a martingale. Let

$$(\varphi_1, \varphi_2)_B(x, v, y) = B_y(\varphi_1\varphi_2)(x, v, y) - \varphi_1(x, v, y)B_y\varphi_2(x, v, y) - \varphi_2(x, v, y)B_y\varphi_1(x, v, y).$$

Then the co-quadratic variation process

$$[M_i^{\varphi_1}, M_j^{\varphi_2}](t) = \int_0^t n^2(\varphi_1, \varphi_2)_B(x_i^{(n)}(r), v_i^{(n)}(r), Y_i(n^2r))dr, \text{ when } i = j; \quad \text{ and } = 0, \text{ when } i \neq j,$$

First, we consider test functions of the form (1.22). Noting the process $\gamma_n(t)$ in (1.15) is essentially identified from a finite dimensional dynamic, by the classical finite dimensional Ito's formula, if we define

(2.1) $A_n f(\gamma)$

$$\begin{aligned} = \ & \langle\gamma, \Big(v\cdot\nabla_x - n(\int_{R^d\times R^d\times S} \nabla\Phi_n(x-\bar{x}; y, \bar{y})\gamma(d\bar{x}, d\bar{v}, d\bar{y}) + \nabla\Psi_n(x, y))\cdot\nabla_v + n^2B_y\Big)\frac{\delta f}{\delta\gamma}\rangle \\ & +\frac{n^2}{2N}\langle\gamma, \sum_{k,l=1}^m \partial^2_{kl}\psi(\langle\varphi_1, \gamma\rangle, \ldots, \langle\varphi_m, \gamma\rangle)\langle\varphi_k, \varphi_l\rangle_B\rangle \\ = \ & \langle\gamma, v\cdot\nabla_x\frac{\delta f}{\delta\gamma} - \Big(\int \nabla\Phi_1(x-\bar{x}; y, \bar{y})\gamma(d\bar{x}, d\bar{v}, d\bar{y}) + \nabla\Psi_1(x; y)\Big)\cdot\nabla_v\frac{\delta f}{\delta\gamma}\rangle \\ & +n\langle\gamma, \Big(\int_{R^d\times R^d\times S} \nabla\Phi_2(x-\bar{x}; y, \bar{y})\gamma(d\bar{x}, d\bar{v}, d\bar{y}) + \nabla\Psi_2(x, y)\Big)\cdot\nabla_v\frac{\delta f}{\delta\gamma}\rangle \\ & +n^2\langle\gamma, B_y\frac{\delta f}{\delta\gamma}\rangle \\ & +\frac{n^2}{2N}\Big(\langle\gamma, B\frac{\delta^2 f}{\delta\gamma^2}(x, v, \cdot; x, v, \cdot)(y)\rangle \\ & -2\langle\gamma(dx, dv, dy)\otimes\delta_{\{x,v,y\}}(d\bar{x}, d\bar{v}, d\bar{y}), B_y\frac{\delta^2 f}{\delta\gamma^2}(x, v, y; \bar{x}, \bar{v}, \bar{y})\rangle\Big), \end{aligned}$$

then

$$f(\gamma_n(t)) - f(\gamma_n(0)) - \int_0^t A_n f(\gamma_n(s))ds \quad (2.2)$$

is a martingale. That is, A_n is the generator for the martingale problem determining the probability-measure-valued process $\gamma_n(\cdot)$.

Having only test functions f of the form (1.22) is not good enough. Later, we need to consider test functions of the form such as g in (3.4) and h in (3.5). These are special cases of

$$\begin{aligned} f(\gamma) &= \phi_1(\langle\rho,\varphi_1\rangle,\dots,\langle\rho,\varphi_m\rangle)\int\int\int\int \phi_2(x,\bar{x},\bar{\bar{x}},\hat{x};y,\bar{y},\bar{\bar{y}},\hat{y})\phi_3(x,v) \\ &\qquad \gamma(dx,dv,dy)\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}})\gamma(d\hat{x},d\hat{v},d\hat{y}), \end{aligned}$$

where $\varphi_k \in C_c^\infty(R^d \times R^d)$, $\phi_1 \in C^2(R^m)$ and ϕ_2,ϕ_3 are bounded continuous functions. For such case, at least when $\gamma \in E'_n$, the first two variational derivatives $\delta f/\delta\gamma$ and $\delta^2 f/\delta\gamma^2$ are well defined smooth functions – see the second example of variational derivatives in last section. For such f, we define $A_n f$ using variational derivatives by (2.1). Note that $\gamma_n(t) \in E'_n$ (defined in (1.15)) is identified with a finite dimensional stochastic dynamic. Therefore, essentially the same arguments as in the early part of this section applies, and finite dimensional Ito's formula allow us to conclude that (2.2) is still a martingale.

3. Generator convergence for a class of perturbed test functions

Again, throughout this paper, ρ denote the X,V-marginal of γ (i.e. $\rho(dx,dv)=\gamma(dx,dv,S)$). We will find a particular sub-class of test functions of (1.22) useful and we define

$$D_0 = \{f(\rho)=\psi(\langle\varphi_1,\rho\rangle,\dots,\langle\varphi_m,\rho\rangle): \quad \varphi_k \in C_c^\infty(R^d\times R^d), \psi\in C^2(R^m)\}. \tag{3.1}$$

In this section, we want to identify an operator $A \subset C_b(E)\times C_b(E)$ which is the limit for $A_n \subset C_b(E_n)\times C_b(E_n)$ in the sense that: for every $f \in D(A) = D_0$ (see (3.1) and note that $f=f(\rho)$ only depends on ρ), there exists $f_n \in D(A_n)$ such that

$$\lim_{n\to+\infty} f_n(\gamma_n) = f(\rho), \lim_{n\to+\infty} A_n f_n(\gamma_n) = Af(\rho)$$

whenever $E_n \ni \rho_n \Rightarrow \rho \in E$ (note that ρ_n is the x,v-marginal of $\gamma_n \in E'_n$). Additionally, we will show

$$\sup_n \sup_{\gamma\in E'_n} |f_n(\gamma)| < +\infty,$$

and there exists $C_0, C_1 > 0$,

$$|A_n f_n(\gamma)| \le C_0 + n^{-1}C_1\int |v|d\gamma, \quad \gamma\in E'_n. \tag{3.2}$$

We claim that such f_n can be taken to be

$$f_n(\gamma) = f(\rho) + n^{-1}g(\gamma) + n^{-2}h(\gamma) \tag{3.3}$$

with

(3.4)

$$g(\gamma) = -\int\int \Big(P_{\nu\otimes\nu}\nabla\Phi_2(x-\bar{x};y,\bar{y}) + P_\nu\nabla\Psi_2(x,y)\Big)\cdot\nabla_v\frac{\delta f}{\delta\rho}(x,v)\gamma(dx,dv,dy)\gamma(d\bar{x},d\bar{v},d\bar{y}),$$

where $P_{\nu\otimes\nu}\nabla\Phi_2(x;y,\bar{y}) = \int_{z\in S}\int_{\bar{z}\in S}\nabla\Phi_2(x;z,\bar{z})\nu(y,dz)\nu(\bar{y},d\bar{z})$; and with

$$
\begin{aligned}
(3.5)\; h(\gamma) \;=\; & \int P_\nu(-\nabla\Psi_1)(x,y)\cdot\nabla_v\frac{\delta f}{\delta\rho}(x,v)\gamma(dx,dv,dy) \\
& +\int\int P_{\nu\otimes\nu}\nabla\Phi_1(x-\bar{x};y,\bar{y})\cdot\nabla_v\frac{\delta f}{\delta\rho}(x,v)\gamma(dx,dv,dy)\gamma(d\bar{x},d\bar{v},d\bar{y}) \\
& +\int\int\int P_{\nu\otimes\nu\otimes\nu}a_1(x,y;\bar{x},\bar{y};\bar{\bar{x}},\bar{\bar{y}})\cdot D^2_{vv}\frac{\delta f}{\delta\rho}(x,v) \\
& \qquad \gamma(dx,dv,dy)\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}}) \\
& +\int\int\int\int P_{\nu\otimes\nu\otimes\nu\otimes\nu}a_2(x,y;\bar{x},\bar{y};\bar{\bar{x}},\bar{\bar{y}};\hat{x},\hat{y})\cdot\nabla_{\bar{\bar{v}}}\nabla_v\frac{\delta^2 f}{\delta\rho^2}(x,v;\bar{\bar{x}},\bar{\bar{v}}) \\
& \qquad \gamma(dx,dv,dy)\gamma(d\hat{x},d\hat{v},d\hat{y})\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}}).
\end{aligned}
$$

See (3.14) and (3.15) for the definition of matrices a_1, a_2. In the above (and below), we denote matrix

$$
\nabla_{\bar{\bar{v}}}\nabla_v\frac{\delta^2 f}{\delta\rho^2}(x,v;\bar{\bar{x}},\bar{\bar{v}}) = \Big(\frac{\partial^2}{\partial v_i\partial\bar{\bar{v}}_j}\frac{\delta^2 f}{\delta\rho^2}(x,v;\bar{\bar{x}},\bar{\bar{v}})\Big)_{d\times d}.
$$

Therefore,

$$
\begin{aligned}
& P_{\nu\otimes\nu\otimes\nu\otimes\nu}a_2(x,y;\bar{x},\bar{y};\bar{\bar{x}},\bar{\bar{y}};\hat{x},\hat{y})\cdot\nabla_{\bar{\bar{v}}}\nabla_v\frac{\delta^2 f}{\delta\rho^2}(x,v;\bar{\bar{x}},\bar{\bar{v}}) \\
= \;& \sum_{k,l=1}^m \partial^2_{kl}\psi(\langle\rho,\varphi_1\rangle,\ldots,\langle\rho,\varphi_m\rangle)\nabla_{\bar{\bar{v}}}\varphi_k(\bar{\bar{x}},\bar{\bar{v}})\cdot P_{\nu\otimes\nu\otimes\nu\otimes\nu}a_2(x,y;\bar{x},\bar{y};\bar{\bar{x}},\bar{\bar{y}};\hat{x},\hat{y})\cdot\nabla_v\varphi_l(x,v).
\end{aligned}
$$

Since $\delta f/\delta\rho \in C_c(R^d\times R^d), \delta^2 f/\delta\rho^2 \in C_c(R^d\times R^d\times R^d\times R^d)$, both $g, h \in C_b(E'_n)$. Below, we explain the choice of g and h, together with the convergence of $H_n f_n$ to Hf, through explicit calculations.

By the martingale problem (2.2), for f_n of the form (3.3),

$$
\begin{aligned}
(3.6)\qquad & A_n f_n(\gamma) = A_n f(\gamma) + n^{-1}A_n g(\gamma) + n^{-2}A_n h(\gamma) \\
= \;& \langle\gamma, v\cdot\nabla_x\frac{\delta f}{\delta\rho} - \Big(\int\nabla\Phi_1(x-\bar{x};y,\bar{y})\gamma(d\bar{x},d\bar{v},d\bar{y}) + \nabla\Psi_1(x;y)\Big)\cdot\nabla_v\frac{\delta f}{\delta\rho}\rangle \\
& +\langle\gamma, -\Big(\int_{R^d\times R^d\times S}\nabla\Phi_2(x-\bar{x};y,\bar{y})\gamma(d\bar{x},d\bar{v},d\bar{y}) + \nabla\Psi_2(x,y)\Big)\cdot\nabla_v\frac{\delta g}{\delta\gamma}\rangle \\
& +\langle\gamma, B_y\frac{\delta h}{\delta\gamma}\rangle \\
& +n\langle\gamma, -\Big(\int_{R^d\times R^d\times S}\nabla\Phi_2(x-\bar{x};y,\bar{y})\gamma(d\bar{x},d\bar{v},d\bar{y}) + \nabla\Psi_2(x,y)\Big)\cdot\nabla_v\frac{\delta f}{\delta\rho} + B_y\frac{\delta g}{\delta\gamma}\rangle \\
& +o(1),
\end{aligned}
$$

where

$$
\begin{aligned}
o(1) \;=\; & n^{-1}\Big\{\langle\gamma, v\cdot\nabla_x \frac{\delta g}{\delta\gamma} - \Big(\int \nabla\Phi_1(x-\bar{x};y,\bar{y})\gamma(d\bar{x},d\bar{v},d\bar{y}) + \nabla\Psi_1(x;y)\Big)\cdot\nabla_v\frac{\delta g}{\delta\gamma}\rangle \\
& +\frac{n^2}{2N}\Big(\langle\gamma, B\frac{\delta^2 g}{\delta\gamma^2}(x,v,\cdot;x,v,\cdot)(y)\rangle \\
& -2\langle\gamma(dx,dv,dy)\otimes\delta_{\{x,v,y\}}(d\bar{x},d\bar{v},d\bar{y}), B_y\frac{\delta^2 g}{\delta\gamma^2}(x,v,y;\bar{x},\bar{v},\bar{y})\rangle\Big)\Big\} \\
& +n^{-2}\Big\{\langle\gamma, v\cdot\nabla_x\frac{\delta h}{\delta\gamma} - \Big(\int\nabla\Phi_1(x-\bar{x};y,\bar{y})\gamma(d\bar{x},d\bar{v},d\bar{y}) + \nabla\Psi_1(x;y)\Big)\cdot\nabla_v\frac{\delta h}{\delta\gamma}\rangle \\
& +n\langle\gamma, \Big(\int_{R^d\times R^d\times S}\nabla\Phi_2(x-\bar{x};y,\bar{y})\gamma(d\bar{x},d\bar{v},d\bar{y}) + \nabla\Psi_2(x,y)\Big)\cdot\nabla_v\frac{\delta h}{\delta\gamma}\rangle \\
& +\frac{n^2}{2N}\Big(\langle\gamma, B\frac{\delta^2 h}{\delta\gamma^2}(x,v,\cdot;x,v,\cdot)(y)\rangle \\
& -2\langle\gamma(dx,dv,dy)\otimes\delta_{\{x,v,y\}}(d\bar{x},d\bar{v},d\bar{y}), B_y\frac{\delta^2 h}{\delta\gamma^2}(x,v,y;\bar{x},\bar{v},\bar{y})\rangle\Big)\Big\}.
\end{aligned}
$$

If g is taken to be (3.4), then following Taylor expansion (1.21), we identify

$$
\begin{aligned}
(3.7)\quad \frac{\delta g}{\delta\gamma}(x,v,y) \;=\; & -\int P_{\nu\otimes\nu}\nabla\Phi_2(x-\bar{x};y,\bar{y})\cdot\nabla_v\frac{\delta f}{\delta\rho}(x,v)\gamma(d\bar{x},d\bar{v},d\bar{y}) \\
& -\int(P_{\nu\otimes\nu}\nabla\Phi_2(\bar{\bar{x}}-x;\bar{\bar{y}},y)\nabla_{\bar{\bar{v}}}\frac{\delta f}{\delta\rho}(\bar{\bar{x}},\bar{\bar{v}})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}}) \\
& -P_\nu\nabla\Psi_2(x,y)\cdot\nabla_v\frac{\delta f}{\delta\rho}(x,v) \\
& -\int\int\Big(P_{\nu\otimes\nu}\nabla\Phi_2(\bar{\bar{x}}-\bar{x};\bar{\bar{y}},\bar{y}) + P_\nu\nabla\Psi_2(\bar{\bar{x}},\bar{\bar{y}})\Big) \\
& \cdot\nabla_{\bar{\bar{v}}}\frac{\delta^2 f}{\delta\rho^2}(\bar{\bar{x}},\bar{\bar{v}};x,v)\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}})\gamma(d\bar{x},d\bar{v},d\bar{y}).
\end{aligned}
$$

Similarly, we can also compute $\delta^2 g/\delta\gamma^2$ explicitly. From the conditions on Φ_2, Ψ_2 (Condition 1.2), it follows that

$$
(3.8)\qquad \nabla_x\frac{\delta g}{\delta\gamma}(x,v;y), \nabla_v\frac{\delta g}{\delta\gamma}(x,v;y), B\frac{\delta^2 g}{\delta\gamma^2}(x,v,\cdot;x,v,\cdot)(y), B_y\frac{\delta^2 g}{\delta\gamma^2}(x,v,y;\bar{x},\bar{v},\bar{y})
$$

are all bounded over their respective domain. Analogous estimates holds when g is replaced by the h in (3.5). Note that because $\delta f/\delta\rho, \delta^2 f/\delta\rho^2$ have compact support,

$$
(3.9)\qquad \nabla_v\frac{\delta g}{\delta\gamma}(x,v;y)\in C_c(R^d\times R^d\times S).
$$

To emphasize the dependence of $o(1)$ on n and on γ, we write $o(1;n,\gamma)$. By (3.8) and similar bounded estimates for h and by (3.9), there exists $C_1, C_2, C_3 > 0$,

$$
(3.10)\qquad |o(1;n,\gamma)| \le n^{-1}(C_2 + C_1\int|v|d\gamma) + \frac{n}{N}C_3, \quad \gamma\in E'_n.
$$

Therefore

$$\lim_{n\to+\infty} \sup_n \sup_{\gamma\in E'_n, \int |v| d\gamma \le C} |o(1;n,\gamma)| = 0, \quad \forall C>0.$$

Hence $o(1)$ is a higher order term.

We now explain the choice of g and h assuming they are well chosen so that $o(1)$ is a higher order term in the sense that (3.10) holds. Our point of departure is (3.6). In order to have the above $A_n f_n$ staying bounded as we pass $n\to+\infty$, we have to choose g so that

$$\langle \gamma, -\Big(\int_{R^d\times R^d\times S} \nabla\Phi_2(x-\bar{x};y,\bar{y})\gamma(d\bar{x},d\bar{v},d\bar{y}) + \nabla\Psi_2(x,y)\Big)\cdot\nabla_v \frac{\delta f}{\delta\gamma} + B_y\frac{\delta g}{\delta\gamma}\rangle = 0. \tag{3.11}$$

g in (3.4) is chosen so that this is satisfied. We verify this below. $\delta g/\delta\gamma$ is given by (3.7). By (1.10) and by (1.12),

$$\begin{aligned}
&\int B_y \frac{\delta g}{\delta\gamma}(x,v,y)\gamma(dx,dv;dy) \qquad (3.12)\\
= &\int\int \Big\{ \Big(B_y(-P_{\nu\otimes\nu}\nabla\Phi_2(x-\bar{x};y,\bar{y})) + B_{\bar{y}}(-P_{\nu\otimes\nu}\nabla\Phi_2(x-\bar{x};y,\bar{y}))\Big)\cdot\nabla_v\frac{\delta f}{\delta\rho}(x,v)\\
&\qquad + B_y(-P_\nu\nabla\Psi_2(x,y))\cdot\frac{\delta f}{\delta\rho}(x,v)\Big\}\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(dx,dv,dy)\\
= &\int\int (\nabla\Phi_2(x-\bar{x};y,\bar{y}) + \nabla\Psi_2(x,y))\cdot\nabla_v\frac{\delta f}{\delta\rho}(x,v)\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(dx,dv,dy).
\end{aligned}$$

Hence (3.11) is satisfied.

We can now use the above choice of g to simplify the $A_n f_n$ in (3.6), and work further on identifying a good choice of h next. Note that

$$\begin{aligned}
&\nabla_v\frac{\delta g}{\delta\gamma}(x,v,y) \qquad (3.13)\\
= &\Big(-\int P_{\nu\otimes\nu}\nabla\Phi_2(x-\bar{x};y,\bar{y})\gamma(d\bar{x},d\bar{v},d\bar{y}) - P_\nu\nabla\Psi_2(x,y)\Big)\cdot D_{vv}\frac{\delta f}{\delta\rho}(x,v)\\
&-\nabla_v\int\int\Big(P_{\nu\otimes\nu}\nabla\Phi_2(\bar{\bar{x}}-\bar{x};\bar{\bar{y}},\bar{y}) + P_\nu\nabla\Psi_2(\bar{\bar{x}},\bar{\bar{y}})\Big)\\
&\qquad\cdot\nabla_{\bar{\bar{v}}}\frac{\delta^2 f}{\delta\rho^2}(\bar{\bar{x}},\bar{\bar{v}};x,v)\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}})\gamma(d\bar{x},d\bar{v},d\bar{y}).
\end{aligned}$$

where D_{vv} denotes the $d\times d$ Hessian matrix

$$D^2_{vv}\varphi(x,v) = \Big(\frac{\partial^2}{\partial v^{(i)}\partial v^{(j)}}\varphi(x,v)\Big)_{d\times d}, \quad v=(v^{(1)},\dots,v^{(d)}).$$

Let matrices

$$\begin{aligned}
a_1(x,y;\bar{x},\bar{y};\bar{\bar{x}},\bar{\bar{y}}) &= \Big(\nabla\Phi_2(x-\bar{x};y,\bar{y}) + \nabla\Psi_2(x,y)\Big) \qquad (3.14)\\
&\quad\times\Big(P_{\nu\otimes\nu}\nabla\Phi_2(x-\bar{\bar{x}};y,\bar{\bar{y}}) + P_\nu\nabla\Psi_2(x,y)\Big),
\end{aligned}$$

(see (1.20) for the definition of "$\times$") and

$$\begin{aligned}
a_2(x,y;\bar{x},\bar{y};\bar{\bar{x}},\bar{\bar{y}};\hat{x},\hat{y}) &= (\nabla\Phi_2(x-\bar{x};y,\bar{y}) + \nabla\Psi_2(x,y)) \qquad (3.15)\\
&\quad\times(P_{\nu\otimes\nu}\nabla\Phi_2(\bar{\bar{x}}-\hat{x};\bar{\bar{y}},\hat{y}) + P_\nu\nabla\Psi_2(\bar{\bar{x}},\bar{\bar{y}})).
\end{aligned}$$

Then

$$(3.16)\quad \begin{aligned} & A_n f_n(\gamma) \\ = \ & \langle \gamma, v\cdot\nabla_x \frac{\delta f}{\delta\rho} - \Big(\int \nabla\Phi_1(x-\bar{x};y,\bar{y})\gamma(d\bar{x},d\bar{v},d\bar{y}) + \nabla\Psi_1(x;y)\Big)\cdot\nabla_v\frac{\delta f}{\delta\rho}\rangle \\ & +\langle\gamma, \int\int a_1(x,y;\bar{x},\bar{y};\bar{\bar{x}},\bar{\bar{y}})\cdot D^2_{vv}\frac{\delta f}{\delta\rho}(x,v)\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}})\rangle \\ & +\langle\gamma, \int\int\int a_2(x,y;\bar{x},\bar{y};\bar{\bar{x}},\bar{\bar{y}};\hat{x},\hat{y})\cdot\nabla_{\bar{\bar{v}}}\nabla_v\frac{\delta^2 f}{\delta\rho^2}(x,v;\bar{\bar{x}},\bar{\bar{v}}) \\ & \qquad \gamma(d\hat{x},d\hat{v},d\hat{y})\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}})\rangle \\ & +\langle\gamma, B_y\frac{\delta h}{\delta\gamma}\rangle + o(1). \end{aligned}$$

The idea of choosing h is such that the above limit of $A_n f_n$ only depends on ρ (the X,V-marginal of γ). To achieve this, we need to "average out" all the $y,\bar{y},\bar{\bar{y}},\hat{y}$ dependence in (3.16). h in (3.5) serves such a purpose. We verify in several steps.

First, we make three interesting observations (3.17), (3.18) and (3.19) in order to simplify the expression of $\langle\gamma, B_y\delta h/\delta\gamma\rangle$. Let $\varphi_1 = \varphi_1(y,\hat{y},\bar{y},\bar{\bar{y}}) \in C_b(S^4)\cap D(\mathbf{B})$, and $\varphi_2 = \varphi_2(\rho;x,v;\bar{\bar{x}},\bar{\bar{v}}) \in C_b(E\times(R^d\times R^d)^2)$ where $\delta\varphi_2/\delta\rho$ is still a bounded function. By (1.10),

$$(3.17)\quad \begin{aligned} & \langle\gamma, B\frac{\delta}{\delta\gamma}\int\int\int\int P_{\nu\otimes\nu\otimes\nu\otimes\nu}\varphi_1(y,\hat{y},\bar{y},\bar{\bar{y}})\varphi_2(\rho;x,v;\bar{\bar{x}},\bar{\bar{v}}) \\ & \qquad \gamma(dx,dv,dy)\gamma(d\hat{x},d\hat{v},d\hat{y})\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}})\rangle \\ = \ & \int\int\int\int\Big(\mathbf{B}P_{\nu\otimes\nu\otimes\nu\otimes\nu}\varphi_1(y,\hat{y},\bar{y},\bar{\bar{y}})\Big)\varphi_2(\rho;x,v;\bar{\bar{x}},\bar{\bar{v}}) \\ & \qquad \gamma(dx,dv,dy)\gamma(d\hat{x},d\hat{v},d\hat{y})\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}}) \\ = \ & \int\int\int\int\Big(-\varphi_1(y,\hat{y},\bar{y},\bar{\bar{y}}) + P_{\pi_0\otimes\pi_0\otimes\pi_0\otimes\pi_0}\varphi_1\Big)\varphi_2(\rho;x,v;\bar{\bar{x}},\bar{\bar{v}}) \\ & \qquad \gamma(dx,dv,dy)\gamma(d\hat{x},d\hat{v},d\hat{y})\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}}). \end{aligned}$$

We note that

$$(3.18)\quad \begin{aligned} & P_{\pi_0\otimes\pi_0\otimes\pi_0\otimes\pi_0}a_2(x,y;\bar{x},\bar{y};\bar{\bar{x}},\bar{\bar{y}};\hat{x},\hat{y}) \\ = \ & P_{\pi_0\otimes\pi_0\otimes\pi_0\otimes\pi_0}\Big\{(\nabla\Phi_2(x-\bar{x};y,\bar{y}) + \nabla\Psi_2(x,y))\times(P_{\nu\otimes\nu}\nabla\Phi_2(\bar{\bar{x}}-\hat{x};\bar{\bar{y}},\hat{y}) + P_\nu\nabla\Psi_2(\bar{\bar{x}},\bar{\bar{y}}))\Big\} \\ = \ & \Big\{\Big(P_{\pi_0\otimes\pi_0}\mathbf{B}(-P_{\nu\otimes\nu}\nabla\Phi_2(x-\bar{x};y,\bar{y}) - P_\nu\nabla\Psi_2(x,y))\Big) \\ & \qquad \times\Big(P_{\pi_0\otimes\pi_0}(P_{\nu\otimes\nu}\nabla\Phi_2(\bar{\bar{x}}-\hat{x};\bar{\bar{y}},\hat{y}) + P_\nu\nabla\Psi_2(\bar{\bar{x}},\bar{\bar{y}}))\Big)\Big\} \\ = \ & 0, \end{aligned}$$

and that

$$(3.19)\quad \int\int P_{\pi_0\otimes\pi_0\otimes\pi_0} a_1(x,y;\bar{x},\bar{y};\bar{\bar{x}},\bar{\bar{y}})\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}})$$

$$= \int\int P_{\pi_0\otimes\pi_0\otimes\pi_0}\Big\{\Big(\nabla\Phi_2(x-\bar{x};y,\bar{y})+\nabla\Psi_2(x,y)\Big)$$
$$\times\Big(P_{\nu\otimes\nu}\nabla\Phi_2(x-\bar{\bar{x}};y,\bar{\bar{y}})+P_\nu\nabla\Psi_2(x,y)\Big)\Big\}\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}})$$

$$= \int\int P_{\pi_0\otimes\pi_0\otimes\pi_0}\Big\{\Big(-\mathbf{B}(P_{\nu\otimes\nu}\nabla\Phi_2(x-\bar{x};y,\bar{y})+P_\nu\nabla\Psi_2(x,y))\Big)$$
$$\times\Big(P_{\nu\otimes\nu}\nabla\Phi_2(x-\bar{\bar{x}};y,\bar{\bar{y}})+P_\nu\nabla\Psi_2(x,y)\Big)\Big\}\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}})$$

$$= \int\int P_{\pi_0}\Big\{\Big(-B_yP_\nu(\nabla\bar{\bar{\Phi}}_2(x-\bar{x};y)+\nabla\Psi_2(x,y))\Big)$$
$$\times\Big(P_\nu(\nabla\bar{\bar{\Phi}}_2(x-\bar{\bar{x}};y)+\nabla\Psi_2(x,y))\Big)\Big\}\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}}),$$

$$= \int_{\bar{x},\bar{v},\bar{y}}\int_{\bar{\bar{x}},\bar{\bar{v}},\bar{\bar{y}}}\int_{y\in S}\Big\{\Big(-BP_\nu(\nabla\bar{\bar{\Phi}}_2(x-\bar{x};y)+\nabla\Psi_2(x,y))\Big)$$
$$\times\Big(P_\nu(\nabla\bar{\bar{\Phi}}_2(x-\bar{\bar{x}};y)+\nabla\Psi_2(x,y))\Big)\Big\}\pi_0(dy)\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}})\gamma(d\bar{x},d\bar{v},d\bar{y})$$

$$= \int_{y\in S}\Big\{\Big(-BP_\nu(\rho_X*\nabla\bar{\bar{\Phi}}_2(x;y)+\nabla\Psi_2(x,y))\Big)\times\Big(P_\nu(\rho_X*\nabla\bar{\bar{\Phi}}_2(x;y)+\nabla\Psi_2(x,y))\Big)\Big\}\pi_0(dy),$$

where $\bar{\bar{\Phi}}_2(x;y)$ is defined in (1.18).

In view of the above three identities, the right hand side of (3.16) reduces to

$$(3.20)\qquad A_nf_n(\gamma) = Af(\rho)+o(1),$$

where

(3.21)

$$Af(\rho) = \langle\rho, v\cdot\nabla_x\frac{\delta f}{\delta\rho}(x,v)-(\rho_X*\nabla\bar{\Phi}_1+\nabla\bar{\Psi}_1)(x,y)\cdot\nabla_v\frac{\delta f}{\delta\rho}+a(\rho,x)\cdot D^2_{vv}\frac{\delta f}{\delta\rho}(x,v)\rangle$$

with $\bar{\Phi}_1, \bar{\Psi}_1$ defined by (1.14) and square matrix $a(\rho,x)$ defined by (1.19).

Lemma 3.1. *Let A be defined according to (3.21) and f_n by (3.3) with $g,h\in C_b(E')$ given by (3.4), (3.5). Then*

$$\lim_{n\to+\infty} A_nf_n(\gamma_n) = Af(\rho),$$

whenever $\rho_n(dx,dv)=\gamma_n(dx,dv,S)\Rightarrow\rho(dx,dv)$ in the weak convergence of probability measure topology and $\sup_n\int|v|d\gamma_n<+\infty$. Moreover (3.2) holds.

Proof. The conclusion follows from (3.20) and (3.10). □

4. Energy estimates and tightness

Again, ρ_n denotes the x,v-marginal of $\gamma_n\in E'_n\subset\mathcal{P}(R^d\times R^d\times S)$ (i.e. $\rho_n(dx,dv)=\gamma_n(dx,dv,S)\in E_n$). We prove that, under appropriate initial conditions, the sequence of stochastic processes $\{\rho_n(t;dx,dv), 0\le t<+\infty : n=1,2,\dots\}$ is tight (i.e. relatively compact in weak convergence topology for the probability measures in the trajectory space). See Theorem 4.4.

Following the general method given by Theorem 9.1 in Chapter 3 of [3], there are two key steps in proving tightness for processes. First, we prove a compact containment property (Lemma 4.1) by studying the time evolution of an energy function which is effectively averaged at the macroscopic scale.

Let

$$\mathcal{E}(\rho) = \int_{x,v} \int_{\bar{x},\bar{v}} \Big(\frac{1}{2}|v|^2 + \frac{1}{2}\bar{\Phi}_1(x-\bar{x}) + \bar{\Psi}_1(x)\Big)\rho(d\bar{x}, d\bar{v})\rho(dx, dv), \tag{4.1}$$

and

$$V(\rho) = \log(1+\mathcal{E}(\rho)). \tag{4.2}$$

Since $\bar{\Phi}_1, \bar{\Psi}_1 \geq 0$ and $\bar{\Psi}_1$ has compact level set in R^d (Condition 1.2.3), $\mathcal{E}$ and V have compact level sets in $\mathcal{P}(R^d \times R^d)$ under the weak convergence of probability measure topology.

Lemma 4.1. *Suppose that*

$$\sup_n E[V(\rho_n(0))] < +\infty. \tag{4.3}$$

Then

$$\lim_{M\to+\infty} \sup_n P(\tau_{n,M} \leq T) = 0, \qquad \forall T > 0. \tag{4.4}$$

Consequently, we have a compact containment property: for each $T > 0$, $\epsilon > 0$, there exists compact set $K = K(T, \epsilon) \subset \mathcal{P}(R^d \times R^d)$ such that

$$P(\exists t, 0 \leq t \leq T, \rho_n(t) \notin K) < \epsilon.$$

Furthermore, we have energy estimate

$$E[V(\rho_n(t))] \leq E[V(\rho_n(0))] + t(n^{-1}C_0 + \|\mathrm{tr}(a_1)\|_\infty) \tag{4.5}$$

where $C_0 > 0$ is some deterministic constant depending on Ψ_i, Φ_i and B.

We prove this result using a stochastic Lyapunov function technique. Let $\rho \in E_n$. First, we note that for $\varphi(x,v) : R^d \times R^d \mapsto R$ which is bounded from below, $\int \varphi d\rho = N^{-1}\sum_{i=1}^N \varphi(x_i, v_i) < +\infty$. In particular, $\int(\bar{\Psi}_1(x) + |v|^2)d\rho < +\infty$.

Proof. From the definition of the first two variational derivatives in (1.21), for $\rho \in E_n$ and $\gamma \in E'_n$ (see (1.24) and (1.23)),

$$\frac{\delta V}{\delta \rho}(x,v) = (1+\mathcal{E}(\rho))^{-1}(\frac{1}{2}|v|^2 + \int_{\bar{x}} \bar{\Phi}_1(x-\bar{x})\rho_X(d\bar{x}) + \bar{\Psi}_1(x)),$$

and

$$\frac{\delta^2 V}{\delta \rho^2}(x,v;\bar{x},\bar{v}) = (1+\mathcal{E}(\rho))^{-1}\bar{\Phi}_1(x-\bar{x}) - \frac{\delta V}{\delta \rho}(x,v)\frac{\delta V}{\delta \rho}(\bar{x},\bar{v}).$$

We let $g_V = g_V(\gamma)$ and $h_V = h_V(\gamma)$ be respectively defined as in (3.4) and (3.5) with $\delta f/\delta\rho$ and $\delta^2 f/\delta\rho^2$ replaced by $\delta V/\delta\rho$ and $\delta^2 V/\delta\rho^2$. In view of the special form for $\rho \in E_n, \gamma \in E_n'$,

$$\begin{aligned} g_V(\gamma) &= -(1+\mathcal{E}(\rho))^{-1}\int\int \Big(P_{\nu\otimes\nu}\nabla\Phi_2(x-\bar{x};y,\bar{y}) + P_\nu\nabla\Psi_2(x;y)\Big)\cdot v\gamma(dx,dv,dy)\gamma(d\bar{x},d\bar{v},d\bar{y}) \\ &= -(1+\mathcal{E}(\rho))^{-1}N^{-2}\sum_{i=1}^N\sum_{j=1}^N \Big(P_{\nu\otimes\nu}\nabla\Phi_2(x_i-x_j;y_i,y_j) + P_\nu\nabla\Psi_2(x_i;y_i)\Big)\cdot v_i; \end{aligned}$$

and

$$\sup_n \sup_{\gamma\in E_n'} |g_V(\gamma)| < +\infty.$$

Similarly, for $\gamma \in E_n'$,

$$\begin{aligned} h_V(\gamma) &= (1+\mathcal{E}(\rho))^{-1}\{\int\int v\cdot(P_\nu(-\nabla\Psi_1)(x,y) + P_{\nu\otimes\nu}\nabla\Phi_1(x-\bar{x};y,\bar{y})) \\ &\qquad \gamma(dx,dv,dy)\gamma(d\bar{x},d\bar{v},d\bar{y}) \\ &\quad + \int\int\int \mathrm{Tr}(P_{\nu\otimes\nu\otimes\nu}a_1(x,y;\bar{x},\bar{y};\bar{\bar{x}},\bar{\bar{y}}))\gamma(dx,dv,dy)\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}})\} \\ &\quad +(1+\mathcal{E}(\rho))^{-2}\int\int\int\int P_{\nu\otimes\nu\otimes\nu\otimes\nu}a_2(x,y;\bar{x},\bar{y};\bar{\bar{x}},\bar{\bar{y}};\hat{x},\hat{y})\cdot(v\times\bar{\bar{v}}) \\ &\qquad \gamma(dx,dv,dy)\gamma(d\hat{x},d\hat{v},d\hat{y})\gamma(d\bar{x},d\bar{v},d\bar{y})\gamma(d\bar{\bar{x}},d\bar{\bar{v}},d\bar{\bar{y}}); \end{aligned}$$

and it follows

$$\sup_n \sup_{\gamma\in E_n'} |h_V(\gamma)| < +\infty.$$

Furthermore, following similar computations as in (3.7) and (3.13), we can verify

$$\begin{aligned} &\sup_n \sup_{\gamma\in E_n'} (|\langle\gamma, v\cdot\nabla_x\frac{\delta g_V}{\delta\gamma}\rangle| + |\langle\gamma, \nabla\Psi_1(x,y)\cdot\nabla_v\frac{\delta g_V}{\delta\gamma}\rangle| + |\langle\gamma, B_y\frac{\delta^2 g_V}{\delta\gamma^2}(x,v,y;x,v,y)\rangle| \\ &\quad +|\langle\gamma(dx,dv,dy)\otimes\delta_{\{x,v,y\}}(d\bar{x},d\bar{y},d\bar{y}), B_y\frac{\delta^2 g_V}{\delta\gamma^2}(x,v,y;\bar{x},\bar{v},\bar{y})\rangle|) < +\infty. \end{aligned}$$

Analogous estimates hold when g_V is replaced by h_V as well.

Let

$$V_n(\gamma) = V(\rho) + n^{-1}g_V(\gamma) + n^{-2}h_V(\gamma).$$

We extend the definition of A_n to V_n according to the second expression in (2.1) where everything is expressed in terms of $\delta V_n/\delta\gamma$ and $\delta^2 V_n/\delta\gamma^2$. We now define stopping time

$$\tau_{n,M} = \inf\{t \ge 0 : V(\rho_n(t)) > M\}. \tag{4.6}$$

Recall that for each fixed n, $\gamma_n(t)$ is a probability measure concentrated only on a finite number of points. By the usual finite dimensional Ito's formula,

$$V_n(\gamma_n(t\wedge\tau_{n,M})) - V_n(\gamma_n(0)) - \int_0^{t\wedge\tau_{n,M}} A_nV_n(\gamma_n(s))ds$$

is a martingale. We estimate A_nV_n next.

Following similar computations as in the previous section, we have for $\gamma \in E'_n$,

$$
\begin{aligned}
(4.7)\quad & A_n V_n(\gamma) \\
= \ & \langle \rho, v \cdot \nabla_x \frac{\delta V}{\delta \rho}(x,v) - \Big(\rho * \nabla \bar{\Phi}_1(x) + \nabla \bar{\Psi}_1(x)\Big) \cdot \nabla_v \frac{\delta V}{\delta \rho}(x,v) \\
& \quad + a(\rho, x) \cdot D^2_{vv} \frac{\delta V}{\delta \rho}(x,v) \rangle \\
& + n^{-1} \Big\{ \langle \gamma, v \cdot \nabla_x \frac{\delta g_V}{\delta \gamma} - \Big(\int \nabla \Phi_1(x - \bar{x}; y, \bar{y}) \gamma(d\bar{x}, d\bar{v}, d\bar{y}) \\
& \quad + \nabla \Psi_1(x; y) \Big) \cdot \nabla_v \frac{\delta g_V}{\delta \gamma} \rangle + \frac{n^2}{2N} \Big(\langle \gamma, B \frac{\delta^2 g_V}{\delta \gamma^2}(x, v, \cdot; x, v, \cdot)(y) \rangle \\
& \quad - 2 \langle \gamma(dx, dv, dy) \otimes \delta_{\{x,v,y\}}(d\bar{x}, d\bar{v}, d\bar{y}), B_y \frac{\delta^2 g_V}{\delta \gamma^2}(x, v, y; \bar{x}, \bar{v}, \bar{y}) \rangle \Big) \Big\}, \\
& + n^{-2} \Big\{ \Big(\langle \gamma, v \cdot \nabla_x \frac{\delta h_V}{\delta \gamma} - \Big(\int \nabla \Phi_1(x - \bar{x}; y, \bar{y}) \gamma(d\bar{x}, d\bar{v}, d\bar{y}) + \nabla \Psi_1(x; y) \Big) \cdot \nabla_v \frac{\delta h_V}{\delta \gamma} \rangle \Big) \\
& \quad + \frac{n^2}{2N} \Big(\langle \gamma, B \frac{\delta^2 h_V}{\delta \gamma^2}(x, v, \cdot; x, v, \cdot)(y) \rangle \\
& \quad - 2 \langle \gamma(dx, dv, dy) \otimes \delta_{\{x,v,y\}}(d\bar{x}, d\bar{v}, d\bar{y}), B_y \frac{\delta^2 h_V}{\delta \gamma^2}(x, v, y; \bar{x}, \bar{v}, \bar{y}) \rangle \Big) \Big\} \\
& + n^{-1} \Big\{ \langle \gamma, \Big(\int_{R^d \times R^d \times S} \nabla \Phi_2(x - \bar{x}; y, \bar{y}) \gamma(d\bar{x}, d\bar{v}, d\bar{y}) + \nabla \Psi_2(x, y) \Big) \cdot \nabla_v \frac{\delta h_V}{\delta \gamma} \rangle \Big\} \\
\leq \ & \|\mathrm{Tr}(a_1)\|_\infty + n^{-1} C_0
\end{aligned}
$$

where $C_0 > 0$ is some deterministic constant depending on Ψ_i, Φ_i and B.

Consequently,

$$
E[V_n(\gamma_n(t \wedge \tau_{n,M})] \leq E[V_n(\gamma_n(0))] + t(n^{-1} C_0 + \|\mathrm{tr}(a_1)\|_\infty). \tag{4.8}
$$

For each n fixed, $V(\rho_n(t))$ is a process which is continuous in time. Let $T, M > 0$,

$$
\begin{aligned}
& MP(\tau_{n,M} \leq T) \leq E[V(\rho_n(\tau_{n,M})) I_{\tau_{n,M} \leq T}] \leq E[V(\rho_n(T \wedge \tau_{n,M})) I_{\tau_{n,M} \leq T}] \\
\leq \ & E[V_n(\gamma_n(T \wedge \tau_{n,M}))] + \sup_m \sup_{\gamma \in E'_m} (n^{-1} |g_V(\gamma)| + n^{-2} |h_V(\gamma)|) \\
\leq \ & E[V_n(\gamma_n(0))] + T(n^{-1} C_0 + \|\mathrm{tr}(a_1)\|_\infty) + \sup_m \sup_{\gamma \in E'_m} (n^{-1} |g_V(\gamma)| + n^{-2} |h_V(\gamma)|).
\end{aligned}
$$

Therefore (4.4) follows. Let $\epsilon > 0$, by selecting M large enough, then the compact containment properly also follows. Finally, (4.5) follows from (4.8) by taking $M \to +\infty$ and by noting lower semicontinuity of V. $\square$

We choose a special $M = M_n = 2 \log(n)$, and denote

$$
\tau_n = \tau_{n, M_n}.
$$

Then from this definition and the definition of V,

$$
I(r \leq \tau_n) n^{-1} e^{V/2(\rho_n(r))} \leq n^{-1} e^{M_n/2} = 1, \quad r > 0. \tag{4.9}
$$

Lemma 4.2. *For each $f \in D_0$ (see (3.1)), $\{f(\rho_n(\cdot \wedge \tau_n)) : n = 1, 2, \dots\}$ is a sequence of relatively compact real valued processes.*

Proof. We apply Theorem 9.4 in Chapter 3 of Ethier and Kurtz [3].

Let $T > 0$. Let f_n be given as in (3.3) with g, h defined by (3.4), (3.5). First, we note

$$\sup_n E[\sup_{0\le t\le T} |f_n(\gamma_n(t\wedge\tau_n)) - f(\rho_n(t\wedge\tau_n))|] \le n^{-1}\|g\| + n^{-2}\|h\| \to 0, \text{ as } n\to+\infty.$$

This verifies (9.17) in the above mentioned Theorem 9.4 of [3].

With each n fixed, γ_n really only identify with a finite dimensional stochastic dynamic and by the finite dimensional Ito's formula,

$$f_n(\gamma_n(t\wedge\tau_n)) - f_n(\gamma_n(0)) - \int_0^{t\wedge\tau_n} A_n f_n(\gamma_n(r))dr$$

is a martingale. By estimate (3.2) (Lemma 3.1) and noting (4.9), for $T > 0$,

$$\begin{aligned}
&\sup_{n=1,2,\ldots} E[\sup_{0\le t\le T} |\int_0^t I(r\le\tau_n)A_n f_n(\gamma_n(r))dr|] \\
\le\ & \sup_n E[\sup_{0\le t\le T} |\int_0^t I(r\le\tau_n)\Big(C_0 + n^{-1}C_1\sqrt{\int |v|^2 d\gamma_n(r)}\Big)dr|) \\
\le\ & \sup_n E[\sup_{0\le t\le T} |\int_0^t I(r\le\tau_n)(C_0 + \sqrt{2}n^{-1}C_1\exp\{V/2(\rho_n(r))\})dr|) \\
\le\ & (C_0 + 2C_1)T < +\infty.
\end{aligned}$$

This verifies (9.18) in Theorem 9.4 in Chapter 3 of [3].

The conclusion of this lemma follows from the above two estimates. □

We conclude the tightness property.

Lemma 4.3. *The sequence of measure valued processes* $\{\rho_n(\cdot\wedge\tau_n) : n = 1,2,\ldots\}$ *is tight with trajectory in space* $C_{\mathcal{P}(R^d\times R^d)}[0,+\infty)$.

Proof. The class of functions in D_0 is closed under multiplication and linear combinations, and separates points in $\mathcal{P}(R^d\times R^d)$. In addition, for each $\rho\in\mathcal{P}(R^d\times R^d)$, we can find $f\in D_0$, $f(\rho)\ne 0$. By a generalization of Stone-Weierstrass theorem (e.g. Theorem A.8 in Appendix A of Feng and Kurtz [6]), D_0 is a dense subset of $C_b(\mathcal{P}(R^d\times R^d))$ in the topology of uniform convergence on compact set.

Combining Lemmas 4.1 and 4.2, the result follows from Theorem 9.1 in Chapter 3 of Ethier and Kurtz [3]. □

Theorem 4.4. *Suppose that (4.3) hold. Then (4.5) holds and the processes* $\{\rho_n(\cdot) : n = 1,2,\ldots\}$ *is tight with trajectories in space* $C_{\mathcal{P}(R^d\times R^d)}[0,+\infty)$.

Proof. Let $\{\rho_{n_k}(\cdot\wedge\tau_{n_k}) : k = 1,2,\ldots\}$ be a convergent subsequence with limiting point $\rho_0(\cdot)$. The existence of such sequence follows from the tightness result of Lemma 4.3. Let r be any metric which gives the weak convergence of probability measure topology on $\mathcal{P}(R^d\times R^d)$. Then for each $\epsilon > 0$,

$$P(\sup_{0\le t\le T} r(\rho_n(t\wedge\tau_n), \rho_n(t)) > \epsilon) \le P(\tau_n\le T)\to 0, \quad \text{as } n\to+\infty.$$

Therefore by Slusky's theorem (e.g. Corollary 3.3 in Chapter 3 of Ethier and Kurtz [3]), $\{\rho_{n_k}(\cdot) : k = 1,2,\ldots\}$ converges in distribution to process $\rho_0(\cdot)$ as well.

Therefore $\{\rho_n(\cdot) : n = 1,2,\ldots\}$ is relatively compact in the trajectory space. □

5. Identifying the limit equation

Let f_n be given by (3.3) and $\tau_{n,M}$ be given by (4.6). By Ito's formula,

$$E[\Big(f_n(\gamma_n(t\wedge\tau_{n,M}))-f_n(\gamma_n(s\wedge\tau_{n,M}))-\int_{s\wedge\tau_{n,M}}^{t\wedge\tau_{n,M}}A_nf_n(\gamma_n(r))dr\Big)\Pi_{i=1}^k h_i(\rho_n(t_i))]=0,$$

and for every $h_1,\dots,h_k\in C_b(E)$, $0\le t_1\le\dots,t_k\le s\le t$, and $k=1,2,\dots$. Let $\rho_0(\cdot)$ be a limit processes of $\{\rho_n(\cdot):n=1,2,\dots\}$ (Theorem 4.4). Then by the convergence of A_nf_n to Af in Lemma 3.1,

$$E[\Big(f(\rho_0(t))-f(\rho_0(s))-\int_s^t Af(\rho_0(r))dr\Big)\Pi_{i=1}^k h_i(\rho_0(t_i))]=0,$$

implying ρ_0 is a solution to the martingale problem

$$f(\rho_0(t))-f(\rho_0(0))-\int_0^t Af(\rho_0(s))ds=M^f(t) \tag{5.1}$$

where M^f is a martingale with respect to the natural filtration induced by ρ_0.

We notice that A is really a first-order differential operator in infinite dimensions in the sense that $A(fg)=fAg+gAf$, if $f,g\in D(A)$. In particular,

$$A(f^2)-2fAf\equiv 0,\quad f\in D(A).$$

Therefore, by Exercise 29 on page 93 of Ethier and Kurtz [3], the quadratic variation

$$[M^f,M^f](t)=\int_0^t(Af^2(\rho_0(s))-2f(\rho_0(s))Af(\rho_0(s)))ds\equiv 0.$$

This implies that $M^f\equiv 0$. In particular, taking $f(\rho)=\langle\varphi,\rho\rangle$ for $\varphi\in C_c^\infty(R^d\times R^d)$, this means that ρ is just a weak solution (Schwartz distributional sense) to (1.1). Therefore we arrive at

Theorem 5.1. *Suppose (4.3) holds. Then under Condition 1.2, stochastic process $\{\rho_n:n=1,2,\dots\}$ is tight, and every convergent subsequence converges in probability to a weak (i.e. Schwartz distributional) solution of (1.1) in the path space.*

If we can show that weak solution to (1.1) is unique, then the above conclusion can be strengthened from convergence of subsequence to convergence. We pursue this without additional mild conditions on $\Phi_i,\Psi_i,i=1,2$ next.

6. Uniqueness of the Vlasov-Fokker-Planck equation - a particle representation approach

Uniqueness of probability-measure-valued solution for (1.1) can be proved using a probabilistic approach.

First, we consider a countably infinite system of stochastic differential equations

$$\begin{aligned}\dot{x}_i(t)&=v_i(t)\\ dv_i(t)&=\rho_X(t)*\nabla\bar\Phi_1(x_i(t))dt+\nabla\bar\Psi_1(x_i(t))dt+\sqrt{2}\sigma(\rho(t);x_i(t))dW_i(t);\end{aligned}\tag{6.1}$$

for some σ such that $\sigma\sigma^T=a$. In the above equation,

$$\rho(t,dx,dv)=\lim_{n\to+\infty}\frac{1}{n}\sum_{j=1}^n\delta_{x_i(t),v_i(t)}(dx,dv),\tag{6.2}$$

and $(W_1, W_2, \dots, W_k, \dots)$ is a countably infinite sequence of i.i.d. standard Brownian motions. In the above, we define $\sigma(\rho; x) = a^{1/2}(\rho, x)$ where a is given by (1.19). Note that a is non-negative definite square matrix, therefore its square root is well defined. The following additional smoothness condition on Φ_2, Ψ_2 will be useful in this section.

Condition 6.1. $\Psi_2(x; y), \Phi_2(x; y_1, y_2)$ *have continuous derivatives in the* x*-variable upto the third order. Moreover,*

$$\sup_{x \in R^d; y_1, y_2 \in S} |D^1 \Phi_2(x; y_1, y_2)| + |D^2 \Phi_2(x; y_1, y_2)| + |D^3 \Phi_2(x; y_1, y_2)| < +\infty$$

and

$$\sup_{x \in R^d, y \in S} |D^1 \Psi_2(x; y)| + |D^2 \Psi_2(x; y)| + |D^3 \Psi_2(x; y)| < +\infty$$

where all derivatives are taken with respect to x*.*

With the above condition, by Theorem 5.2.3 of Stroock and Varadhan [9], and by the defining structure of a in (1.19) (also noting (1.8)), $a^{1/2}$ is Lipschitz in the sense

$$|a^{1/2}(\rho; x) - a^{1/2}(\gamma, y)| \le C(|x - y| + d_W(\rho, \gamma)),$$

where d_W is the order-0 Wasserstein metric on $\mathcal{P}(R^d \times R^d)$:

$$d_W(\rho, \gamma) = \sup_{f \in C_b(R^d \times R^d), |f| \le 1, |f(x,v) - f(y,u)| \le |x-y| + |v-u|} \left| \int f d\rho - \int f d\gamma \right|. \tag{6.3}$$

Condition 6.2. $\nabla \bar{\Phi}_1, \nabla \bar{\Psi}_1 : R^d \mapsto R$ *are Lipschitz continuous.*

By Section 10 of Kurtz and Protter [8], we have the following

Lemma 6.3. *Let Conditions 6.1 and 6.2 hold. Assume that the initial values are random variables and* $\{(x_i(0), v_i(0)) : i = 1, 2, \dots\}$ *is a stationary sequence. Then existence and strong uniqueness holds for the system (6.1).*

We denote $E_0 = R^d \times R^d$. The state space of (6.1) is

$$(E_0)^\infty = \{(\vec{x}, \vec{v}) = (x_1, \dots, x_n, \dots; v_1, \dots, v_n, \dots) : (x_k, v_k) \in E_0\},$$

with the usual product topology. From the symmetry of labels in (6.1), if the initial value $(\vec{x}(0), \vec{v}(0))$ is exchangeable, then the solution (uniqueness follows from Lemma 6.3) $(\vec{x}(t), \vec{v}(t))$ for (6.1) is exchangeable for all $t > 0$. Let ρ be any weak (Schwartz distributional) solution of partial differential equation (1.1). The goal of this section is to apply the Markov mapping theorem of Kurtz [7] to show that (Theorem 6.8)

$$\rho(t) = \lim_{n \to +\infty} \sum_{k=1}^{n} \delta_{(x_i(t), v_i(t))}.$$

The above limit exists because that $\{(x_i(t), v_i(t)), i = 1, 2, \dots\}$ is an exchangeable sequence for all $t > 0$. Therefore, the uniqueness of ρ follows from the uniqueness of (6.1) which is verified in Lemma 6.3. The method is known as particle representation method. We provide the details next.

First, we express solution to (6.1) in terms of a martingale problem. We consider

$$D(A_0) = \{g(\vec{x}, \vec{v}) = \Pi_{k=1}^m \varphi_k(x_k, v_k) : \varphi_k \in C_c^\infty(R^d \times R^d), m = 1, 2, \dots\}. \tag{6.4}$$

For each such $g \in D(A_0)$, we define

$$(6.5) \qquad A_0 g(\vec{x},\vec{v}) = \sum_{k=1}^{m} (\Pi_{j\neq k}\varphi_j(x_j,v_j))\Big(v_k \cdot \nabla_x \varphi_k(x_k,v_k) + (\eta(\vec{x},\vec{v}) * \nabla\bar{\Phi}_1(x_k) + \nabla\bar{\Psi}_1(x_k)) \cdot \nabla_v \varphi_k(x_k,v_k) + a(\eta(\vec{x},\vec{v}),x_k) \cdot D^2_{vv}\varphi_k(x_k,v_k)\Big)$$

Then by Ito's formula, any solution to the infinite system (6.1) is also a solution to the martingale problem

$$(6.6) \qquad g(\vec{x}(t),\vec{v}(t)) - g(\vec{x}(0),\vec{v}(0)) - \int_0^t A_0 g(\vec{x}(s),\vec{v}(s))ds = \text{martinagle}.$$

Let

$$\mathcal{H} = \{h(\vec{x},\vec{v}) = f(x_1,v_1,\dots,x_m,v_m) - f(x_{\sigma_1},v_{\sigma_1},\dots,x_{\sigma_m},v_{\sigma_m}) : \text{all permutation } \sigma, f \in B((E_0)^m), m = 1,\dots\}.$$

Assume additionally that the distribution $\nu_0 \in \mathcal{P}((E_0)^\infty)$ for $((x_k(0),v_k(0)) : k = 1,2,\dots)$ is exchangeable. Because of the symmetry in (6.1), exchangeability of $((x_k(t),v_k(t)) : k = 1,2,\dots)$ follows for all $t > 0$, therefore

$$(6.7) \qquad E[h(\vec{x}(t),\vec{v}(t))] = 0, \quad t \geq 0, h \in \mathcal{H}.$$

We call $(\vec{x},\vec{v})$ a solution to the *restricted* martingale problem for $(A_0,\mathcal{H},\nu_0)$ in the sense that both (6.6) and (6.7) are satisfied.

As in the setting of finite multi-dimensional diffusion processes, we also have the following.

Lemma 6.4. *Let $(\vec{x},\vec{v})$ be a solution to the restricted martingale problem for $(A_0,\mathcal{H},\nu_0)$ with trajectory in $C_{(E_0)^\infty}[0,+\infty)$ and filtration $\{\mathcal{F}_t : t \geq 0\}$, on a probability space $(\Omega,\mathcal{F},P)$. Then there exists countably infinite i.i.d Brownian motions W_is and probability space $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{P})$ with filtration $\tilde{\mathcal{F}}_t \supset \mathcal{F}_t$ and $(\tilde{\vec{x}},\tilde{\vec{v}})$ on probability space $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{P})$ with filtration $\tilde{\mathcal{F}}_t$. $(\tilde{\vec{x}},\tilde{\vec{v}})$ has the same distribution as $(\vec{x},\vec{v})$, and $(\tilde{\vec{x}},\tilde{\vec{v}})$ satisfies Ito's equation (6.1).*

Proof. The same proof of Proposition 3.1 and Theorem 3.3 in Chapter 5 of Ethier and Kurtz [3] apply to this countably infinite system setting. □

Therefore, by Lemma 6.3, we have

Lemma 6.5. *Let the initial distribution ν_0 be exchangeable. Any solution of the restricted martingale problem for $(A_0,\mathcal{H},\nu_0)$, with trajectory in $C_{(E_0)^\infty}[0,+\infty)$, is unique.*

Next, we define a map $\eta : (E_0)^\infty \mapsto \mathcal{P}(E_0)$ (η is the map γ in Kurtz [7]'s notation). Let $\gamma_0 \in \mathcal{P}(E_0)$ be arbitrary but fixed. We define

$$\eta(\vec{x},\vec{v}) = \lim_{n\to+\infty} n^{-1}\sum_{k=1}^{n} \delta_{(x_k,v_k)} \in \mathcal{P}(E_0), \quad (\vec{x},\vec{v}) \in (E_0)^\infty$$

if the limit on the right hand side exists in the weak convergence of probability measure sense; and $\eta(\vec{x},\vec{v}) = \gamma_0$ otherwise. We now define a transition probability measure $\alpha : \mathcal{P}(E_0) \mapsto (E_0)^\infty$:

$$\alpha(\rho; d\vec{x}, d\vec{v}) = \Pi_{k=1}^\infty \rho(dx_k,dv_k).$$

We verify another condition required by Theorem 3.2, Corollaries 3.5, 3.7 of Kurtz [7].

Lemma 6.6. *The set* $\eta^{-1}(\rho) = \{(\vec{x}, \vec{v}) \in (E_0)^\infty : \eta(\vec{x}, \vec{v}) = \rho\}$ *satisfies*

$$\alpha(\rho; \eta^{-1}(\rho)) = 1. \tag{6.8}$$

Proof. We note that if we take $(E_0)^\infty$ to be a probability space endowed with the product measure $\Pi_{i=1}^\infty \rho(dx_i, dv_i)$, then (6.8) holds because of the strong law of large number. □

We verify some regularity properties for A_0, which is needed to apply a version of the Markov mapping theorem (Corollary 3.7) of Kurtz [7].

From (6.4), $D(A_0)$ is closed under multiplication and separates points. It is also a pre-generator in the sense of that paper because system (6.1) can be approximated by similar systems when the Brownian motion terms are replaced by centered i.i.d. Poisson processes. Furthermore, we have

Lemma 6.7. *There exists a countable subset* $\{g_k\} \subset D(A_0)$ *such that the graph of* A_0 *is contained in the bounded pointwise closure of the linear span of* $\{(g_k, A_0 g_k)\}$.

Proof. Noting the form (6.5), we can always approximate the φ_js by countable sequence of polynomials with rational coefficients. The collection of such polynomials (and hence any finite products of them) is countable. □

Theorem 6.8. *[Uniqueness and particle representation of solution] Let Condition 1.2 hold. Define*

$$C = \{(\int g(\vec{x}, \vec{v})\alpha(\cdot; d\vec{x}, d\vec{v}), \int A_0 g(\vec{x}, \vec{v})\alpha(\cdot, d\vec{x}, d\vec{v})) : g \in D(A_0)\}.$$

Then

1. $D(A_0)$ *consists of functions of the form*

$$f(\rho) = \int g(\vec{x}, \vec{v})\alpha(\rho; d\vec{x}, d\vec{v}) = \Pi_{k=1}^m \langle \varphi_k, \rho \rangle,$$

and

$$\begin{aligned} Cf(\rho) &\equiv \int A_0 g(\vec{x}, \vec{v})\alpha(\rho; d\vec{x}, d\vec{v}) \\ &= \sum_{k=1}^m (\Pi_{j \neq k} \langle \varphi_j, \rho \rangle)\Big(\langle v \cdot \nabla_x \varphi_k + (\rho * \nabla\bar{\Phi}_1 + \nabla\bar{\Psi}_1) \cdot \nabla_v \varphi_k \\ &\qquad + a(\rho, x) \cdot D^2_{vv}\varphi_k), \rho \rangle\Big). \end{aligned}$$

2. $D(C)$ *is closed under multiplication,* $Cf^2 = 2fCf$ *and for each* $f \in D(C)$*. Any solution to the martingale problem for* C *satisfies*

$$f(\rho(t)) - f(\rho(0)) - \int_0^t Cf(\rho(s))ds = 0, \tag{6.9}$$

which is equivalent to (Schwartz distributional) solution of (1.1).

3. *Let* $\rho \in C_{\mathcal{P}(R^d \times R^d)}[0, +\infty)$ *be a solution to (6.9) with deterministic initial condition* $\rho(0) = \rho_0 \in \mathcal{P}(E_0)$*. Let* $(\vec{x}, \vec{v})$ *be the unique solution (see Lemma 6.3) to (6.1) with initial distribution* $(\vec{x}, \vec{v}) \sim \nu_0(d\vec{x}, d\vec{v}) = \Pi_{i=1}^\infty \rho_0(dx_i, dv_i)$*. Assume that Conditions 6.1, 6.2 hold. Then the particle representation*

$$\rho(t) = \eta(\vec{x}(t), \vec{v}(t)).$$

holds. Consequently, solution to (6.9), in the path space $C_{\mathcal{P}(R^d\times R^d)}[0,+\infty)$, is unique. That is, weak (Schwartz distributional) solution to (1.1) in $C_{\mathcal{P}(R^d\times R^d)}[0,+\infty)$ is unique.

Proof. The first two parts of the theorem follows by direct verification. Since $Cf^2 = 2fCf$, as in Section 5, the martingale appearing in the martingale problem for C has to be zero and (6.9) follows.

Part three follows by applying Corollary 3.7 of Kurtz [7] and by noting existence and uniqueness for solution of restricted martingale problem for $(A_0, \mathcal{H}, \nu_0)$ is equivalent to existence and uniqueness for infinite system (6.1) with initial condition $(\vec{x}(0), \vec{v}(0)) \sim \nu_0$. □

Using the above conclusion, we can strengthen the existence result in Theorem 5.1 to the following.

Theorem 6.9. *Suppose (4.3) hold. Then under Conditions 1.2, and 6.1 and 6.2, the stochastic processes $\{\rho_n : n = 1, 2, \dots\}$ in (1.16) converges in probability to the unique solution of partial differential equation (1.1) in $C_{\mathcal{P}(R^d\times R^d)}[0,+\infty)$.*

7. Comments on large deviation

Large deviation behavior is also expected to hold for (1.7). At least informally, similar operator convergence (using perturbed test functions) method as employed in this article can be used to prove large deviation as well. For a systematic exposition on generator convergence approach to large deviation, see Feng and Kurtz [6].

Let

$$H_n f(\gamma) = \frac{1}{N} e^{-Nf} A_n e^{Nf}(\gamma) \tag{7.1}$$

Then by Lemma 3.2 in Chapter 4 of Ethier and Kurtz [3],

$$\exp\{Nf(\gamma_n(t)) - Nf(\gamma_n(s)) - \int_s^t NH_n f(\gamma_n(r))dr\} = M_n^f(t) - M_n^f(s) \tag{7.2}$$

where $M_n^f(t)$ is a mean one nonnegative martingale. By the main result in [6], convergence of H_n to a limit H, together with some regularities on $\{\gamma_n : n = 1, 2, \dots\}$ and on H, implies a large deviation principle satisfied by the associate processes:

$$\lim_{n\to+\infty} \frac{1}{N} \log P(\rho_n(\cdot) \in A) = -\inf_{\rho\in A} I(\rho(\cdot)), \quad \forall \text{ sufficiently regular set } A \subset C_{\mathcal{P}(R^d\times R^d)}[0,\infty),$$

where I is a functional characterized by H.

For the specific application here, H is identified as follows: for each $f \in D(H)$, we select g and h, let

$$f_n(\gamma) = f(\rho) + \frac{1}{n} g(\gamma) + \frac{1}{n^2} h(\gamma) \tag{7.3}$$

then

$$H_n f_n(\gamma_n) \to Hf(\rho), \qquad \text{whenever } \rho_n(dx,dv) \equiv \gamma_n(dx,dv;S) \Rightarrow \rho(dx,dv).$$

Compared to the A_n, whose convergence is justified in this article, H_ns are fully nonlinear. However, the same idea on how to construct "perturbed test function" f_ns apply without essential change. In finite dimensions, similar problems have been considered by Evans [4], who also devised the name of perturbed test function method.

Routine but long calculations reveal that H has to be of the following form.

$$\begin{aligned} Hf(\rho) &= \langle \rho, v\cdot\nabla_x \frac{\delta f}{\delta\rho} - (\rho_X * \nabla\bar{\Phi}_1 + \nabla\bar{\Psi}_1)(x,y)\cdot\nabla_v\frac{\delta f}{\delta\rho} + a(\rho_X, x)\cdot D^2_{vv}\frac{\delta f}{\delta\rho}\rangle \\ &\quad + \int \nabla_v\frac{\delta f}{\delta\rho}(x,v)\cdot a(\rho_X; x)\cdot\nabla_v\frac{\delta f}{\delta\rho}(x,v)\rho(dx,dv). \end{aligned}$$

The regularity on H needed here is a type of uniqueness (the comparison principle) result on viscosity solution of

$$(I - \alpha H)f = h, \tag{7.4}$$

for sufficiently many bounded continuous h and all $\alpha > 0$. This turns out to be a hard analysis problem and is interesting to investigate on its own right. Earlier analysis literature had focused on simpler situations where the state space is a subset of a Banach space, instead of space of probability measures. See Crandall and Lions [1]. Relying on mass transport technique and exploring the probabilistic connection here, Feng and Katsoulakis [5], and Sections 9.4 and 13.3 of Feng and Kurtz [6] made progress in extending earlier viscosity techniques to certain problems with probability measure valued state space.

The large deviation rate function can be expressed as a type of cost functional using a control variable $p = p(t,x,v)$ appearing in the following controlled partial differential equation of the Fokker-Planck type

$$\partial_t\rho + v\cdot\nabla_x\rho - (\rho_X * \nabla\bar{\Phi}_1 + \nabla\bar{\Psi}_1)\cdot\nabla_v\rho + \nabla_v(\rho\nabla_v p) = a(\rho_X; x)\cdot D^2_{vv}\rho. \tag{7.5}$$

Details and rigorous justifications of the above mentioned approach will be pursued elsewhere.

References

1. Crandall, Michael G. and Lions, Pierre-Louis. Hamilton-Jacobi Equations in Infinite Dimensions, PART I *J. Funct. Anal.* **62** (1985) 3. 379-396. PART II *J. Funct. Anal.* **65** (1986) 3. 368-405. PART III *J. Funct. Anal.* **68** (1986) 2. 214-247. PART IV *J. Funct. Anal.* **90** (1990) 2. 237-283. PART V *J. Funct. Anal.* **97** (1991) 2. 417-465. PART VII *J. Funct. Anal.* **125** (1994) 1. 111-148.
2. Dawson, Donald. A. and Gärtner, Jürgen. Large Deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics* **20**. 247-308, 1987.
   /*Large Deviations Techniques and Applications*, Jones and Bartlett Publishers, Boston, 1993.
3. Ethier, Stewart N. and Kurtz, Thomas G. *Markov Processes*, John Wiley and Sons, New York, 1986.
4. Evans, Lawrence C. The perturbed test function method for viscosity solutions of nonlinear PDE. *Proc. Roy. Soc. Edinburgh Sect A.* **111** (1989), 359-375.
5. Feng, Jin and Katsoulakis, Markos A Hamilton-Jacobi theory for controlled gradient flows in infinite dimensions. *Submitted* 2003.
6. Feng, Jin and Kurtz, Thomas *Large deviation for stochastic processes* Submitted, 2005.
   /Panagiotis E. PDE-viscosity solution approach to some problems of large deviations. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **13** (1986), 171-192.
7. Kurtz, Thomas. G. Martingale Problems for conditional distributions of Markov Processes. *Electronic Journal of Probability*, **3**, (1998), 1-29.
8. Kurtz, Thomas. G. and Protter, Philip. Limit theorems for solutions of stochastic equations II, *CIME School in Probability, Springer Lecture Notes in Mathematics*, **1627**, (1996), 197-285.
9. Stroock, Daniel. W. and Varadhan, S.R.S. *Multidimensional Diffusion Processes* Springer, New York, 1979 (Corrected Second Printing, 1997).

Department of Mathematics and Statistics,, University of Massachusetts,, Amherst, MA 01002.