manuscript No.
(will be inserted by the editor)

Stochastic Wasserstein Hamiltonian Flows

Jianbo Cui · **Shu Liu** · **Haomin Zhou**

Received: date / Accepted: date

Abstract In this paper, we study the stochastic Hamiltonian flow in Wasserstein manifold, the probability density space equipped with L^2-Wasserstein metric tensor, via the Wong–Zakai approximation. We begin our investigation by showing that the stochastic Euler–Lagrange equation, regardless it is deduced from either the variational principle or particle dynamics, can be interpreted as the stochastic kinetic Hamiltonian flows in Wasserstein manifold. We further propose a novel variational formulation to derive more general stochastic Wasserstein Hamiltonian flows, and demonstrate that this new formulation is applicable to various systems including the stochastic Schrödinger equation, Schrödinger equation with random dispersion, and Schrödinger bridge problem with common noise.

Keywords stochastic Hamiltonian flow · density manifold · Wong–Zakai approximation

Mathematics Subject Classification (2000) Primary 58B20, Secondary 35R60, 35Q41, 35Q83, 65M75

1 Introduction

The density space equipped with L^2-Wasserstein metric forms an infinite dimensional Riemannain manifold, often called Wasserstein manifold or density manifold in literature (see e.g. [40]). It plays an important role in optimal transport theory [54]. Many well-known equations, such as Schrödinger equation, Schrödinger bridge problem and Vlasov equation, can be written as Hamiltonian systems on the density manifold. In this sense, they can be considered as members of the so-called Wasserstein Hamiltonian flows ([54, 4, 29, 17, 14, 15, 20]).

J. Cui
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
E-mail: jianbo.cui@polyu.edu.hk (corresponding author)

S. Liu
School of Mathematics, Georgia Tech, Atlanta, GA 30332, USA
E-mail: sliu459@gatech.edu

H. Zhou
School of Mathematics, Georgia Tech, Atlanta, GA 30332, USA
E-mail: hmzhou@math.gatech.edu

The study of Wasserstein Hamiltonian flow can be traced back to Nelson's mechanics ([47–49]), where a probabilistic interpretation of the linear Schrödinger's equation is given. The rigorous probabilistic contents in Quantum Physics were understood as stochastic variation boundary problems for the probability densities with given marginals (cf. [3,16]). The work of Bismut [7], which is closely related to the principles of stochastic optimal transport theory, showed how the random perturbations affects the classical optimization problem in the expectation sense for both Lagrangian and Hamiltonian formalism. Motivated by the ideas of Schrödinger [51] and Bernstein [6], the connection between the Nelson's approach and hydrodynamics on the Wasserstein space was first discovered by [45]. For more contents on the stochastic optimal transport problem, we refer to [46]. By using Madelung transformation, it is known that a polar representation reveals the Hamiltonian structure of classical Schrödinger equations. We refer to [36] for a more comprehensive review on the geometric hydrodynamics and its relationship with the optimal transport theory. Another framework of second-order differential geometry to derive stochastic Lagarangian and Hamiltonian mechanics and to establish their related Hamilton–Jacobi–Bellman equations are presented in [33]. Recently, it is shown in [15] that the kinetic Hamiltonian flows in density space are probability transition equations of classical Hamiltonian ordinary differential equations (ODEs). In other words, this reveals that the density of a Hamiltonian flow in sample space is a Hamiltonian flow on density manifold.

In the existing works on Wasserstein Hamiltonian flows, random perturbations of common noise type (see e.g. [21,22]) to the Lagrangian functional are not considered in the continuous space. Consequently, the theory is neither directly applicable to the Wasserstein Hamiltonian flows subjected to random perturbations, nor to the systems whose parameters are not given deterministically. The main goal of this article is developing a theory to cover these scenarios in which the stochasticity is presented. More precisely, we mainly focus on the stochastic perturbation of the Wasserstein Hamiltonian flow,

$$
\begin{aligned}
d\rho_t &= \frac{\delta}{\delta S_t}\mathscr{H}_0(\rho_t,S_t)dt,\\
dS_t &= -\frac{\delta}{\delta \rho_t}\mathscr{H}_0(\rho_t,S_t)dt,
\end{aligned}
$$

with a Hamiltonian $\mathscr{H}_0$ on the density manifold and $\frac{\delta}{\delta S}, \frac{\delta}{\delta \rho}$ being the variational derivatives, which is proposed by only imposing randomness on the initial position in the phase space [15]. This is different from the Hamiltonian flows considered in [4], where the authors construct the solutions of the ODEs in the measure space of even dimensional phase variables equipped with the Wasserstein metric. More precisely, the Hamiltonian functional in [4] is defined on the Wasserstein manifold $\mathscr{P}_2(\mathbb{R}^{2d})$, which contains densities of joint distributions of both position and momentum variables, while the system in the current study is mainly defined on the density manifold for the position variable only.

To study the stochastic variational principle on density manifold, we may confront several challenges. First and the foremost, the Wasserstein Hamiltonian flow studied in [15] is induced based on the principle that the density of a Hamiltonian flow in sample space is a Wasserstein Hamiltonian flow in density manifold. This principle may no long hold if the Hamiltonian flow in sample space is perturbed by random noise. Second, the stochastic variational framework must be carefully designed in order to induce stochastic dynamics that possess Hamiltonian structures on Wasserstein manifold. As indicated in [15,43], the Christoffel symbol in Wasserstein space plays an important role in the typical kinetic Hamiltonian dynamics since it induces a certain velocity-momentum transformation that allows us to transfer between the second order Euler–Lagrange equations and the Hamiltonian system in density manifold. However, for the noise perturbed Wasserstein Hamiltonian flows, it is complicated and difficult to introduce such tools for transforming the Euler–Lagrange equations into Hamiltonian dynamics in general.

To overcome the difficulties, we begin our study by investigating the classical Lagrangian functional perturbed by the Wong–Zakai approximation (see e.g. [52,57]) on the phase space, and show that its critical point gives the stochastic Hamiltonian flow driven by the Wong–Zakai approximation. With the help of the equivalence of the particle stochastic ODE system and the macro density formulation, in section 3 we prove that the stochastic Hamiltonian flow driven by the Wong–Zakai approximation coincides with the critical point of a stochastic variational principle (see e.g. [55]). In particular, Proposition 3.3 presents the convergence result of the Wong–Zakai approximation to the stochastic Wasserstein Hamiltonian system in Stratonovich sense. However, in general stochastic case, it is still hard to use the Christoffel symbols to derive the stochastic Hamiltonian dynamics.

Furthermore, based on the cotangent bundles of density manifold, we propose a general variational principle to derive a large class of stochastic Hamiltonian equations on density manifold via Wong–Zakai approximation, such as stochastic nonlinear Schrödinger equation (see, e.g., [5,26,38,53]), nonlinear Schrödinger equation with white noise dispersion (see, e.g., [1,2]), and the mean-field game system with common noise (see, e.g., [30,9,10]). We would like to mention that although the Wong–Zakai approximation of stochastic differential equations has been studied for many years (see, e.g., [57,52,8,56]), few results are known for the convergence on the density manifold. In this work, we also provide some new convergence results of Wong–Zakai approximation for the continuity equation induced by stochastic Hamiltonian system and the stochastic Schrödinger equation on density space under suitable assumptions.

Another main message that we would like to convey in this paper is that the stochastic Hamiltonian flow on phase space, when viewed through the lens of conditional probability, induces the stochastic Wasserstein Hamiltonian flow on density manifold, and it is hard to observe those stochastic Hamiltonian structures in the density manifold without the help of conditional probability (see section 3).

The organization of this article is as follows. In section 2, we review the formulation and derivation of Hamiltonian ordinary differential equations (ODEs), and use the Wong–Zakai approximation of the Lagrangian functional to connect the classic and stochastic variational principles on phase space. In section 3, we study the macro behaviors of stochastic Hamiltonian ODE and its Wong–Zakai approximation, including the stochastic Euler–Lagrange equation on density space, Vlasov equation, as well as the generalized stochastic Wasserstein Hamiltonian flow. Several examples are demonstrated in section 4. Throughout this paper, we denote C and c as positive constants which may differ from line to line.

2 Stochastic Hamiltonian ODEs

In this section, we briefly review the classical and stochastic Hamiltonian flows on a finite dimensional Riemannian manifold.

The classical Hamiltonian flow on a smooth d-dimensional Riemannian manifold $(\mathcal{M},g)$ with g being the metric tensor of $\mathcal{M}$, is derived by the variational problem

$$I(x_0,x_T)=\inf_{(x(t))_{t\in[0,T]}}\{\int_0^T L_0(x,\dot{x})dt : x(0)=x_0,\ x(T)=x_T\}.$$

Here the Lagrangian L_0 is a functional (also called Lagrange action functional) defined on the tangent bundle of $\mathcal{M}$. Its critical point induces the Euler-Lagrange equation

$$\frac{d}{dt}\frac{d}{d\dot{x}}L_0(x,\dot{x})=\frac{d}{dx}L_0(x,\dot{x}).$$

When $L_0(x,\dot{x}) = \frac{1}{2}\dot{x}^\top g(x)\dot{x} - f(x)$ with a smooth potential functional f on $\mathscr{M}$, the Euler-Lagrange equation can be rewritten as a Hamiltonian system,

$$\dot{x} = g(x)^{-1}p,\ \dot{p} = -\frac{1}{2}p^\top d_x g^{-1}(x)p - d_x f(x)$$

Here $\top$ denotes the transpose, $p = g(x)\dot{x}$ and the Hamiltonian is

$$H_0(x,p) = \frac{1}{2}p^\top g^{-1}(x)p + f(x).$$

However, the Lagrange action functional $L_0(x,\dot{x})$ may not be homogeneous or it can by impacted by random perturbations in some problems, which is the reason to introduce stochastic Hamiltonian flows.

Let us start with the case that $L(x,\dot{x})$ is composed by the deterministic Lagrange functional $L_0(x,\dot{x})$ and a random perturbation $\eta\sigma(x)\dot{\xi}_\delta(t)$. Here ξ_δ can be chosen as a piecewise continuous differentiable function which obeys certain distribution law in a complete probability space $(\Omega,\mathbb{F},\mathbb{P})$ with a filtration $\{\mathbb{F}_t\}_{t\geq 0}$, $\sigma(\cdot)$ is a potential function and $\eta\in\mathbb{R}$ characterizes the noise intensity. In this paper, ξ_δ is taken as a Wong-Zakai approximation (see e.g. [57]) of the standard Brownian motion $B(t)$ such that $\dot{\xi}_\delta$ is a real function. When $\delta\to 0$, $\xi_\delta(t)$ is convergent to $B(t)$ in pathwise sense or strong sense [57]. For fixed $\omega\in\Omega$, since $\xi_\delta(t)$ is a stochastic process on $(\Omega,\mathbb{F},\mathbb{P})$ with piecewise continuous trajectory, the value of the action functional $\int_0^T L_0(x,\dot{x}) - \eta\sigma(x)\dot{\xi}_\delta(t)dt$ is finite for any given $x(0) = x_0, x(T) = x_T$.

Throughout this paper, we assume that the initial position x_0 of the particle system is a $\mathbb{F}_0$-measurable random variable with the density ρ_0. Let $\mathbb{F}_t, t\geq 0$ be the completion of the filtration generated by the standard Brownian motion. For convenience, we also suppose that x_0 is independent of $B(t), t\geq 0$. To satisfies the above assumptions, we let $(\Omega,\mathbb{F},\mathbb{P}) = (\Omega_B,\{\mathbb{F}_t\}_{t\geq 0},\mathbb{P}_B)\times(\widetilde{\Omega},\widetilde{\mathbb{F}},\widetilde{\mathbb{P}})$, where $B(\cdot)$ is the Brownian motion on Ω_B and x_0 is a random variable on $\widetilde{\Omega}$ independent of Ω_B. Denote $\mathbb{E}$ the expectation with respect to $(\Omega,\mathbb{P})$ and $\mathbb{E}_{\widetilde{\Omega}}$ the *conditional probability* with respect to $(\widetilde{\Omega},\widetilde{\mathbb{P}})$.

Newton's law can be used to derive the Euler–Lagrange equation or the Hamiltonian system in the stochastic case. In order to find out the critical point of $\int_0^T L_0(x,\dot{x}) - \eta\sigma(x)\dot{\xi}_\delta(t)dt$, we calculate its Gâteaux derivative (see, e.g., [31]). Set $x_\varepsilon(t) = x(t) + \varepsilon h(t)$, $h(0) = h(T) = 0$, the Newton's law indicates the critical point satisfies

$$\frac{d}{dt}\frac{\partial}{\partial\dot{x}}L(x,\dot{x}) = \frac{\partial}{\partial x}L(x,\dot{x}) = \frac{\partial}{\partial x}L_0(x,\dot{x}) - \eta\frac{\partial}{\partial x}\sigma(x)\dot{\xi}_{\delta t},$$

which is equivalent to the integral equation

$$\frac{\partial}{\partial\dot{x}}L(x(t),\dot{x}(t)) - \frac{\partial}{\partial\dot{x}}L(x(0),\dot{x}(0)) = \int_0^t \frac{\partial}{\partial x}L_0(x,\dot{x})ds - \eta\int_0^t \frac{\partial}{\partial x}\sigma(x)d\xi_{\delta t}.$$

One can also introduce the Legendre transformation $p = g(x)\dot{x}$,, and get

$$\dot{x} = g(x)^{-1}p,\ \dot{p} = -\frac{1}{2}p^\top d_x g^{-1}(x)p - d_x f(x) - \eta d_x\sigma(x)\dot{\xi}_\delta. \tag{2.1}$$

Since it can be rewritten as

$$\dot{x} = \frac{\partial}{\partial p}H_0(x,p) + \frac{\partial}{\partial p}H_1(x,p)\dot{\xi}_\delta,\ \dot{p} = -\frac{\partial}{\partial x}H_0(x,p) - \frac{\partial}{\partial x}H_1(x,p)\dot{\xi}_\delta,$$

where $H_1(x,p) = \sigma(x)$, the equations form a stochastic Hamiltonian system.

Remark 2.1 *When $\dot{\xi}_\delta$ is a constant, the Hamilton's principle gives a Hamiltonian system with a homogenous perturbation. Otherwise, for a fixed ω, the Hamilton's principle leads to a Hamiltonian system with an inhomogenous perturbation.*

2.1 Wong–Zakai approximation in $\mathcal{M} = \mathbb{R}^d$

In this part, we show that the limit of the Wong-Zakai approximation (2.1) is a stochastic Hamiltonain system.

Lemma 2.1 *Let $\mathcal{M} = \mathbb{R}^d$ and $T > 0$, g be the identity matrix $\mathbb{I}_{d\times d}$. Assume that $f, \sigma \in \mathcal{C}_b^2(\mathcal{M})$, ξ_δ is the linear interpolation of $B(t)$ with width δ and that x_0, p_0 is $\mathbb{F}_0$-apdated. Then* (2.1) *on* $[0,T]$ *is convergent to*

$$dx = p,\ dp = -d_x f(x) - \eta d_x \sigma(x) \circ dB(t),\ \ a.s., \tag{2.2}$$

where $\circ$ denotes the stochastic integral in the Stratonovich sense.

Proof The condition of σ, f ensures the global existence of a unique strong solution for (2.1) and (2.2) by using standard Picard iterations. Then one can follow the classical arguments (see e.g. [52]) to show that the solution of (2.1) is convergent to that of (2.2) and that the right hand side of (2.1) is convergent to that of (2.2).

The following lemma relaxes the classical conditions on the convergence of Wong–Zakai approximation whose proof is presented in Appendix. We call that g is equivalent to $\mathbb{I}_{d\times d}$ if $g \in \mathcal{C}_b^\infty(\mathbb{R}^d;\mathbb{R}^d)$ is symmetric satisfying $\Lambda \mathbb{I}_{d\times d} \succeq g(x) \succeq \lambda \mathbb{I}_{d\times d}$ for some constant $0 < \lambda \le \Lambda$. In the following, we will use the standard notation for the matrix product, that is, $g(x)\cdot(y,z) = y^\top g(x) z$ and $g(x)\cdot y = g(x)y$.

Lemma 2.2 *Let $\mathcal{M} = \mathbb{R}^d$, $T > 0$, g be equivalent to $\mathbb{I}_{d\times d}$. Assume that $f, \sigma \in \mathcal{C}_p^2(\mathcal{M})$, ξ_δ is the linear interpolation of $B(t)$ with the width δ, that x_0, p_0 are $\mathbb{F}_0$-apdated and possess any finite q-moment, $q \in \mathbb{N}^+$, and that*

$$\begin{aligned}
&H_0(x,p) \ge c_0|p| + c_1|x|, \text{ for large enough } |x|,|p| \\
&\eta^2 |\nabla_{pp} H_0(x,p)\cdot(\nabla_x \sigma(x), \nabla_x \sigma(x))| + \eta |\nabla_{pp} H_0(x,p)\cdot(p, \nabla_x \sigma(x))| \\
&+ \eta |\nabla_{pp} H_0(x,p)\cdot(\nabla_x \sigma(x), -\frac{1}{2} p^\top d_x g^{-1}(x) p - \nabla_x f(x))| + \eta |\nabla_{px} H_0(x,p)\cdot(\nabla_x \sigma, g^{-1}(x) p)| \\
&\quad + \eta |\nabla_p H_0(x,p)\cdot \nabla_{xx}\sigma(x) g^{-1}(x) p| \le C_1 + c_1 H_0(x,p).
\end{aligned} \tag{2.3}$$

Then the solution of (2.1) *on* $[0,T]$ *is convergent in probability to the solution of*

$$dx = g^{-1}(x)p,\ dp = -\frac{1}{2} p^\top d_x g^{-1}(x) p - d_x f(x) - \eta d_x \sigma(x) \circ dB(t). \tag{2.4}$$

Denote the solution of (2.1) by $(x^\delta(\cdot, x_0, p_0), p^\delta(\cdot, x_0, p_0))$. According to Lemma 2.2, by studying the equation of $\frac{\partial}{\partial x_0} x^\delta(t, x_0, p_0)$ and $\frac{\partial}{\partial p_0} x^\delta(t, x_0, p_0)$, one could obtain the following convergence result.

Corollary 2.1 *Under the condition of Lemma 2.2, let $f, \sigma \in \mathcal{C}_p^3(\mathcal{M})$. Then for any $\epsilon > 0$, it holds that*

$$\begin{aligned}
&\lim_{\delta \to 0} \mathbb{P}\Big(\sup_{t\in[0,T]} |\frac{\partial}{\partial x_0} x^\delta(t, x_0, p_0) - \frac{\partial}{\partial x_0} x(t, x_0, p_0)| \\
&+ \sup_{t\in[0,T]} |\frac{\partial}{\partial p_0} x^\delta(t, x_0, p_0) - \frac{\partial}{\partial p_0} x(t, x_0, p_0)| \ge \epsilon \Big) = 0.
\end{aligned}$$

Remark 2.2 *One may impose more additional conditions on the coefficients f,σ to obtain the strong convergence order $\frac{1}{2}$ of the Wong–Zakai approximation, that is,*

$$\mathbb{E}\Big[\sup_{t\in[0,T]}|x^{\delta}(t)-x(t)|^{p}\Big]+\mathbb{E}\Big[\sup_{t\in[0,T]}|x^{\delta}(t)-x(t)|^{p}\Big]\leq C\delta^{\frac{p}{2}}.$$

The convergence in probability yields that there exists a pathwise convergent subsequence. In this sense, the limit equation of (2.1) *is* (2.4) *on* $[0,T]$. *When the growth condition* (2.3) *fails, one could also obtain the convergence in probability of* (x^{δ},p^{δ}) *before the stopping time* $\tau_R\wedge\tau_{R_1}$ *(see Appendix for the definition of* τ_R *and* τ_{R_1}*). One could also choose different type of Wong–Zakai approximation of the Wiener process and obtain similar results (see, e.g., [57]).*

2.2 Wong–Zakai approximation on a differential manifold $\mathscr{M}$

Assume that $\mathscr{M}\subset\mathbb{R}^k$ is a d-dimensional differential manifold of class $\mathscr{C}^{\alpha},\alpha\in\mathbb{N}^{+}\cup\infty$ without boundary. Given a $\mathscr{C}^{\alpha}$-diffeomorphism $\phi:W\to V\subset\mathscr{M}$ from an open subset W of $\mathbb{R}^d$ to an open set V of $\mathscr{M}$, the inverse $\phi^{-1}:V\to W$ is called a chart or coordinate system on $\mathscr{M}$. The coordinate components are denoted by $\Phi_1,\Phi_2,\cdots,\Phi_d$, $d\in\mathbb{N}^{+}$. The tangent bundle of $\mathscr{M}$ is denoted by $\mathscr{T}\mathscr{M}:=\{(x,y)\in\mathbb{R}^k\times\mathbb{R}^k|x\in\mathscr{M},y\in\mathscr{T}_x(\mathscr{M})\}$. Moreover, $dim\mathscr{T}_x(\mathscr{M})=d$. The canonical projection is denoted by $\pi:\mathscr{T}\mathscr{M}\to\mathscr{M}$.

In the following, we start from the deterministic Hamiltonian system

$$\dot{x}=p,$$
$$\dot{p}=-d_xf(x),$$

where the vector field $(p,-d_xf(x))\in\mathscr{T}_{(x,p)}\mathscr{T}\mathscr{M}$ for all $(x,p)\in\mathscr{T}\mathscr{M}$. We show how the random force can be added to the system so that $(\dot{x},\dot{p})\in\mathbb{R}^k\times\mathbb{R}^k$ is still tangent to $\mathscr{T}\mathscr{M}$ at (x,p). As a physical interpretation, this tangent condition represents the constrain of the motion equations and is to ensure that the physical motion is living in $\mathscr{T}\mathscr{M}$ by the Kamke property of the maximal solutions (see e.g. [28, Chapter 3]). Consider $\mathscr{M}$ which is regularly defined as the zero level set of a $\mathscr{C}^{\infty}$ map F from $\mathbb{R}^k$ to $\mathbb{R}^{k-d}$. Then we have that the tangent space to $\mathscr{M}$ at x is $\mathscr{T}T_x\mathscr{M}:=\{p\in\mathbb{R}^k|F'(x)p=0\}$, and $T\mathscr{M}=\{(x,p)\in\mathbb{R}^k\times\mathbb{R}^k|F(x)=0,F'(x)p=0\}$. We can also obtain

$$\mathscr{T}\mathscr{T}\mathscr{M}=\{(x,p,\dot{x},\dot{p})|F(x)=0,F'(x)p=0,F'(x)\dot{x}=0,F''(x)(\dot{x},p)+F'(x)\dot{p}=0\}.$$

Therefore, if the added random force satisfies,

$$F'(x)\dot{p}=-F''(x)(\dot{x},p)=\psi(x;p,\dot{x}),\ \dot{x}\in T_x(\mathscr{M}),\tag{2.5}$$

we have $(\dot{x},\dot{p})\in T_{(x,p)}(T\mathscr{M})$. Following [28], we denote a smooth mapping ψ from the vector bundle $\{(x;u,v)\in\mathbb{R}^k\times(\mathbb{R}^k\times\mathbb{R}^k)|x\in\mathscr{M},u,v\in\mathscr{T}_x(\mathscr{M})\}$ to $\mathbb{R}^{k-d}$. Given any vector $z\in\mathbb{R}^{k-d}$, denote by $Az\in(Ker\ F'(x))^{\perp}=(\mathscr{T}_x\mathscr{M})^{\perp}$ the unique solution of $F'(x)\dot{p}=z$. Hence, the solution of (2.5) satisfies

$$\dot{p}=\mu(x;p,\dot{x})+w,$$

where $\mu(x;p,\dot{x})=A\psi(x;p,\dot{x})\in(\mathscr{T}_x(\mathscr{M}))^{\perp}$ and $w\in\mathscr{T}_x(\mathscr{M})$. We observe that to ensure $(\dot{x},\dot{p})\in\mathscr{T}_{(x,p)}(\mathscr{T}\mathscr{M})$, it suffices to take $u,w\in\mathscr{T}_x(\mathscr{M})$ and define $(\dot{x},\dot{p})=(u,\mu(x;p,u)+w)$. In Eq. (2.1) with the driving noise being $-d_x\sigma(x)\dot{\xi}_\delta$, using the above condition, we can verify that it satisfies that $(\dot{x},\dot{p}) \in \mathscr{T}_{(x,p)}(\mathscr{T}\mathscr{M})$. Similarly, a second order differential equation with random force satisfies

$$\ddot{x} = \mu(x;\dot{x},\dot{x}) + \mathscr{R}(t,x,\dot{x}),$$

where $\mathscr{R}_t : \mathscr{T}\mathscr{M} \ni (x,\dot{x}) \mapsto \mathscr{R}(t,x,\dot{x}) \in \mathbb{R}^k$ is a tangent vector field on $\mathscr{M}$. A typical example is that $\mathscr{R} = -\alpha\dot{x} + a(t,x)$ with the frictional force $-\alpha\dot{x}$ and applied random force $a(t,x) = -d_x\sigma(x)\dot{\xi}_\delta(t)$. When $\mathscr{R} = 0$, the above equation is inertial and is so-called geodesic equation on $\mathscr{M}$, which plays an important role in the optimal transport theory (see e.g. [54,29,17,13]).

Lemma 2.3 *Suppose that $\mathscr{M}$ is a d-dimensional compact smooth differential manifold. Let $g = \mathbb{I}$, f,σ be smooth functions on $\mathscr{M}$, ξ_δ be the linear interpolation of $B(t)$ with width δ, and that x_0, p_0 are $\mathbb{F}_0$-adapted and possess any finite q-moment, $q \in \mathbb{N}^+$. Then (x^δ, p^δ) is convergent in probability to the solution (x,p) of* (2.4).

Proof The existence and uniqueness of (x,p) can be found in [32]. The global existence of (x^δ, p^δ) could be also obtained by the fact that $g = \mathbb{I}$, f and σ are globally Lipschitz and that the growth condition (2.3) holds. We only need to show the convergence of (x^δ, p^δ) in probability to (x,p). Since $\mathscr{T}\mathscr{M}$ is $2d$-dimensional manifold which could be embedding to $\mathbb{R}^{2k}$, we can extend the vector field $V(x,p) := (p, -d_x f(x) - \eta d_x\sigma(x))$ to a vector field $\widetilde{V}(\cdot,\cdot)$ on $\mathbb{R}^{2k}$. And thus the equations of (x,p) and (x^δ, p^δ) can be viewed as the equations on $\mathbb{R}^{2k}$. The global existence of (x,p) and (x^δ, p^δ), together with Lemma 2.2, yield the convergence in probability of (x^δ, p^δ).

Remark 2.3 *The above result relies on the particular structure of $g = \mathbb{I}$ and the growth condition* (2.3)*. If this condition* (2.3) *fails, the explosion time $e(x^\delta, p^\delta)$ of (x^δ, p^δ) may depend on δ. And the convergence in probability may only hold before $e(x,p) \wedge \inf_{\delta>0} e(x^\delta, p^\delta)$. When applying different type of Wong–Zakai approximations, the different type of stochastic ODEs could be derived (see e.g. [34]).*

To end this section, we give a special example of stochastic Hamiltonian flows which concentrates on a submanifold with conserved quantities.

Example 2.1 Let $\mathscr{M} = \mathbb{R}^d$, g and $\widetilde{g}$ be metrics equivalent to $\mathbb{I}_{d\times d}$. Define an action functional with random perturbation in dual coordinates,

$$-\int_0^T (\langle p,\dot{x}\rangle - H_0(x,p))dt + \int_0^T H_1(x,p)d\xi_\delta(t),$$

where $H_0(x,p) = \frac{1}{2}p^\top g^{-1}(x)p + f(x)$, $H_1(x,p) = \eta\frac{1}{2}p^\top \widetilde{g}^{-1}(x)p + \eta\sigma(x)$ with smooth potentials f and σ. Then the critical points under the constrain $x(0) = x_0, x(T) = x_T$ satisfies the stochastic Hamiltonian flows

$$\begin{aligned}
\dot{x^\delta} &= \frac{\partial H_0}{\partial p}(x,p) + \frac{\partial H_1}{\partial p}(x^\delta, p^\delta)\dot{\xi}_\delta, \\
\dot{p^\delta} &= -\frac{\partial H_0}{\partial p}(x^\delta, p^\delta) - \frac{\partial H_1}{\partial p}(x^\delta, p^\delta)\dot{\xi}_\delta.
\end{aligned}$$

The solution (x^δ, p^δ) and its limit (x,p) lie on the manifold $\{H_0(x,p) = H_0(x_0,p_0), H_1(x,p) = H_1(x_0,p_0)\}$ when the Hamiltonians satisfies that $\{H_0,H_1\} = 0$ with $\{\cdot,\cdot\}$ being the Possion bracket. Similar to Lemma 2.2, it can be shown that (x^δ, p^δ) converges globally to (x,p) in probability if H_0 or H_1 satisfies the growth condition (2.3).

3 Stochastic Wasserstein Hamiltonian flow

In this section, we study the behaviors of the inhomogenous Hamiltonian system (2.1) and stochastic Hamiltonian system (2.4) on the density manifold. To illustrate the strategy, let us focus on the case that $(\mathscr{M},g)$ equals $(\mathbb{T}^d,\mathbb{I})$ or $(\mathbb{R}^d,\mathbb{I})$. Given the filtered complete probability space $(\Omega,\mathbb{F},(\mathbb{F}_t)_{t\geq 0},\mathbb{P})$, recall that $\xi_\delta(t)$ is the piecewisely linear Wong–Zakai approximation of a standard Brownian motion. For a fixed $\widetilde{\omega}\in\widetilde{\Omega}$, we denote $\tau^\delta := \inf\{t\in(0,T]|x_t^\delta$ is not a smooth diffeomorphism on $\mathscr{M}\}$, $p_t^\delta = v(t,x_t^\delta)$ is the vector field depending on the position and time. Here we view the momentum p as the function v depending on both time and space. Eq. (2.1) becomes

$$\frac{d}{dt}x_t^\delta = v(t,x_t^\delta),$$
$$\frac{d}{dt}v(t,x_t^\delta) = -\nabla f(x_t^\delta) - \eta\nabla\sigma(x_t^\delta)\dot{\xi}_\delta(t).$$

Differentiating $v(t,x_t^\delta(x_0))$ before τ^δ leads to

$$\begin{aligned}\partial_t v(t,x_t^\delta(x_0)) + \nabla v(t,x_t^\delta(x_0))\cdot\frac{d}{dt}x_t^\delta &= \partial_t v(t,x_t^\delta(x_0)) + \nabla v(t,x_t^\delta(x_0))\cdot v(t,x_t^\delta(x_0))\\ &= -\nabla f(x_t^\delta(x_0)) - \eta\nabla\sigma(x_t^\delta(x_0))\dot{\xi}_\delta(t).\end{aligned}$$

Taking $x_0 = (x_t^\delta)^{-1}(x)$, we obtain the following conservation law with random perturbation,

$$\partial_t v(t,x) + \nabla v(t,x)\cdot v(t,x) = -\nabla f(x) - \eta\nabla\sigma(x)\dot{\xi}_\delta(t). \tag{3.1}$$

Taking any test function ψ in $C^\infty(\mathscr{M})$, it holds that

$$\begin{aligned}\frac{d}{dt}\mathbb{E}_{\widetilde{\Omega}}[\psi(x_t^\delta(x_0))] &= \frac{d}{dt}\int_{\mathscr{M}}\psi(x)\rho(t,x)dx = \int_{\mathscr{M}}\nabla\psi(x_t^\delta(x))\cdot v(t,x_t^\delta(x))\rho_0(x)dx\\ &= \int_{\mathscr{M}}\nabla\psi(x)\cdot v(t,x)\rho_t(x)dx,\end{aligned}$$

which implies that for $\omega_B\in\Omega_B$, $\rho_t = x_t^\delta\#\rho_0$, i.e., ρ_t equals the distribution generated by the push-forward map $x_t(\cdot)$ on ρ_0, satisfies the continuity equation,

$$\partial_t\rho(t,x) + \nabla\cdot(\rho(t,x)v(t,x)) = 0. \tag{3.2}$$

Introducing the pseudo inverse $(-\Delta_\rho)^\dagger$ (see e.g. [15]) of the operator

$$\Delta_\rho(\cdot) := -\nabla\cdot(\rho\nabla(\cdot)) \tag{3.3}$$

for a positive density ρ, we denote $S_t = (-\Delta_{\rho_t})^\dagger\partial_t\rho_t$. When there exists a potential S such that $v=\nabla S$, the conservation law with random influence (3.1) and the continuity equation (3.2) induce a Hamiltonian system in density manifold before τ^δ,

$$\begin{aligned}\partial_t\rho_t &= \frac{\delta}{\delta S_t}\mathscr{H}_0(\rho_t,S_t) = -\nabla\cdot(\rho_t\nabla S_t),\\ \partial_t S_t &= -\frac{\delta}{\delta\rho_t}\mathscr{H}_0(\rho_t,S_t) - \frac{\delta}{\delta\rho_t}\mathscr{H}_1(\rho_t,S_t)\dot{\xi}_\delta(t) + C(t)\\ &= -\frac{1}{2}|\nabla S_t|^2 - \frac{\delta}{\delta\rho_t}\mathscr{F}(\rho_t) - \frac{\delta}{\delta\rho_t}\eta\Sigma(\rho_t)\dot{\xi}_\delta(t) + C(t),\end{aligned} \tag{3.4}$$

where $C(t)$ is an arbitrary stochastic process on $(\Omega_B, \mathbb{P}_B)$ independent of the spatial position x and initial velocity $v(0,\cdot) = \nabla S(0,\cdot)$. Here the dominated average energy is

$$\mathscr{H}_0(\rho, S) := K(\rho, S) + \mathscr{F}(\rho) = \int_{\mathscr{M}} \frac{1}{2} |\nabla S(x)|^2 \rho(x) dx + \int_{\mathscr{M}} f(x) \rho(x) dx,$$

and the perturbed average energy is

$$\mathscr{H}_1(\rho, S, t) = \eta \Sigma(\rho_t) = \eta \int_{\mathscr{M}} \sigma(x) \rho(x) dx.$$

Taking $\delta \to 0$, the limit system becomes a stochastic Hamiltonian system,

$$\begin{aligned} d\rho_t &= \frac{\delta}{\delta S_t} \mathscr{H}_0(\rho_t, S_t) dt, \\ dS_t &= -\frac{\delta}{\delta \rho_t} \mathscr{H}_0(\rho_t, S_t) - \frac{\delta}{\delta \rho_t} \mathscr{H}_1(\rho_t, S_t) \star d\xi + C(t) dt, \end{aligned} \tag{3.5}$$

where ξ is the limit process of ξ_δ in the pathwise sense. We would like to remark that the solution of (3.5) may be not $\mathbb{F}_t$-measurable in general, for example when x_0 is not independent of $B(t)$. We refer to [50, section 3.3] for more discussions on the anticipating stochastic differential equations. We also would like to remark that the Stratonovich integral is nature in the study of stochastic Hamiltonian system due to the presence of the chain rule [21–23]. In our particular case, since $\xi_\delta(t)$ is a piecewisely linear Wong-Zakai approximation of $B(t)$ and x_0 is independent of $B(t)$, the limit of (3.1), (3.2) is the following system in Stratonovich sense,

$$\begin{aligned} d\rho_t &= -\nabla \cdot (\rho(t,x) v(t,x)) dt, \\ dv(t,x) + \nabla v(t,x) \cdot v(t,x) dt &= -\nabla f(x) dt - \eta \nabla \sigma(x) \circ dB_t. \end{aligned} \tag{3.6}$$

We would like to emphasize that the above analysis indicates a principle for deriving the stochastic Hamiltonian system on Wasserstein manifold: *The conditional probability density of stochastic Hamiltonian flow in phase space is a stochastic Hamiltonian flow in density manifold almost surely.* In the following we always assume that the initial distribution $\rho(0,\cdot)$ of x_0 and the initial velocity $v(0,\cdot)$ are smooth and bounded.

Proposition 3.1 *Suppose that $\mathscr{M}$ is a d-dimensional compact smooth differential submanifold and $T > 0$. Let $g = \mathbb{I}$, $v(0,\cdot)$ be a smooth vector field, f, σ be smooth functions on $\mathscr{M}$, ξ_δ be the linear interpolation of $B(t)$ with width δ, and that x_0, p_0 are $\mathbb{F}_0$-adapted and possess any finite q-moment, $q \in \mathbb{N}^+$. Then there exists a stopping time τ such that there exists a subsequence of (ρ^δ, v^δ) which converges in probability to the solution (ρ, v) of* (3.6) *before τ.*

Proof Applying Lemma 2.3, we have that $(x_t^\delta, v(t, x_t^\delta))$ is convergent to $(x_t, v(t, x_t))$ in $[0,T]$, a.s., up to a subsequence. Define the stopping time $\tau = \inf\{t \in (0,T] | x_t$ is not smooth diffeomorphism on $\mathscr{M}\}$. For convenience, let us take a subsequence such that for almost $\omega \in \Omega$, $(x_t^\delta, v(t, x_t^\delta))$ converges to $(x_t, v(t, x_t))$ and $\frac{\partial}{\partial x_0} x_t^\delta(x_0)$ convergences to $\frac{\partial}{\partial x_0} x_t(x_0)$. Before $\tau(\omega)$, there exists $\alpha > 0$ such that $\det(\frac{\partial}{\partial x_0} x_t^{-1}(x_0)) > \alpha$. The pathwise convergence of x^δ implies that for any $\varepsilon > 0$ there exists $\delta_0 = \delta(\varepsilon, \omega) > 0$ such that when $\delta \le \delta_0$, $\det(\frac{\partial}{\partial x_0} (x_t^\delta)^{-1}(x_0)) > \alpha - \varepsilon > 0$. Notice that the density function $\rho^\delta(t,y)$ of x_t^δ satisfies $\rho^\delta(t,y) = |\det(\nabla x_t^\delta(y))| \rho(0, x_t^\delta(y))$. Since $\rho(0,\cdot)$ is smooth for any fixed ω and the pathwise convergence of x^δ holds, it follows that $\rho^\delta(t,y)$ converges to the density function of x_t, which is $\rho(t,y) = |\det(\nabla x_t(y))| \rho(0, x_t(y))$. Similarly, the pathwise convergence of $v^\delta(t, x_t^\delta(y))$ to $v(t, x_t(y))$, together with invertibility of x_t^δ and x_t, implies the convergence of $v^\delta(t,x)$ to $v(t,x)$. Consequently, the solution of (ρ^δ, v^δ) is convergent to (ρ, v) in pathwise sense up to a subsequence.

3.1 Vlasov equation

We would like to present the connections and differences between the classic Vlasov equation and stochastic Wasserstein Hamiltonian flow in this part. For simplicity, let us consider the case that $\mathscr{M}=\mathbb{R}^d$. We fix $\widetilde{\omega}\in\widetilde{\Omega}$, and consider (2.1). Taking differential on $\mathbb{E}_\Omega[\phi(x_t^\delta,p_t^\delta)]$ where ϕ is a sufficient smooth test function, we get

$$\begin{aligned}\frac{d}{dt}\mathbb{E}_\Omega[\phi(x_t^\delta,p_t^\delta)]&=\mathbb{E}_\Omega[\nabla_x\phi(x_t^\delta,p_t^\delta)\frac{d}{dt}x_t^\delta+\nabla_p\phi(x_t^\delta,p_t^\delta)\frac{d}{dt}p_t^\delta]\\&=\mathbb{E}_\Omega[\nabla_x\phi(x_t^\delta,p_t^\delta)p_t+\nabla_p\phi(x_t,p_t)(-\nabla_x f(x_t^\delta)-\eta\nabla_x\sigma(x_t^\delta)\dot{\xi}_\delta)].\end{aligned}$$

Denoting the initial joint probability density function by $F_0(x,p)$, it holds that

$$\begin{aligned}&\frac{d}{dt}\int_{\mathbb{R}^d\times\mathbb{R}^d}\phi(x_t^\delta,p_t^\delta)F_0(x,p)dxdp\\&=\int_{\mathbb{R}^d\times\mathbb{R}^d}\Big(\nabla_x\phi(x_t^\delta,p_t^\delta)p_t^\delta+\nabla_p\phi(x_t^\delta,p_t^\delta)(-\nabla_x f(x_t^\delta)-\eta\nabla_x\sigma(x_t^\delta)\dot{\xi}_\delta)\Big)F_0(x,p)dxdp\end{aligned}$$

Thus the joint distribution on Ω, $F_t^\delta=(x_t^\delta,p_t^\delta)^\# F_0$, satisfies

$$\begin{aligned}&\int_{\mathbb{R}^d\times\mathbb{R}^d}\phi(x,p)\frac{d}{dt}F_t^\delta(x,p)dxdp\\&=\int_{\mathbb{R}^d\times\mathbb{R}^d}\Big(\nabla_x\phi(x,p)p+\nabla_p\phi(x,p)(-\nabla_x f(x)\Big)F_t(x,p)dxdp\\&+\mathbb{E}_\Omega[\nabla_p\phi(x_t^\delta,p_t^\delta)(-\eta\nabla_x\sigma(x_t^\delta))\dot{\xi}_\delta(t)\Big].\end{aligned}$$

Notice that the solution process x_t^δ is $\mathbb{F}_{t_{k+1}}$-measurable when $t\in(t_k,t_{k+1}],t_k=k\delta t$ and $\mathbb{F}_{t_k}$-measurable when $t=t_k$, and x_t is $\mathbb{F}_t$-measurable. By applying the chain rule, we have that for $t\in(t_k,t_{k+1}]$,

$$\begin{aligned}&\int_0^t\mathbb{E}_\Omega[\nabla_p\phi(x_s^\delta,p_s^\delta)(-\eta\nabla_x\sigma(x_s^\delta))\dot{\xi}_\delta(s)\Big]ds\\&=\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}}\mathbb{E}_\Omega[\nabla_p\phi(x_s^\delta,p_s^\delta)(-\eta\nabla_x\sigma(x_s^\delta))\dot{\xi}_\delta(s)\Big]ds\\&+\int_{t_k}^t\mathbb{E}_\Omega[\nabla_p\phi(x_s^\delta,p_s^\delta)(-\eta\nabla_x\sigma(x_s^\delta))\dot{\xi}_\delta(s)\Big]ds\\&=\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}}\mathbb{E}_\Omega[\nabla_p\phi(x_{t_j}^\delta,p_{t_j}^\delta)(-\eta\nabla_x\sigma(x_{t_j}^\delta))\frac{B_{t_{j+1}}-B_{t_j}}{\delta}\Big]ds\\&+\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}}\mathbb{E}_\Omega\Big[\Big(\nabla_p\phi(x_s^\delta,p_s^\delta)(-\eta\nabla_x\sigma(x_s))-\nabla_p\phi(x_{t_j}^\delta,p_{t_j}^\delta)(-\eta\nabla_x\sigma(x_{t_j}^\delta))\Big)\frac{B_{t_{j+1}}-B_{t_j}}{\delta}\Big]ds\\&+\int_{t_k}^t\mathbb{E}_\Omega[\nabla_p\phi(x_{t_k}^\delta,p_{t_k}^\delta)(-\eta\nabla_x\sigma(x_{t_k}^\delta))\frac{B_{t_{k+1}}-B_{t_k}}{\delta}\Big]ds\\&+\int_{t_k}^t\mathbb{E}_\Omega[\Big(\nabla_p\phi(x_s^\delta,p_s^\delta)(-\eta\nabla_x\sigma(x_s^\delta))-\nabla_p\phi(x_{t_k}^\delta,p_{t_k}^\delta)(-\eta\nabla_x\sigma(X_s^\delta))\Big)\frac{B_{t_{k+1}}-B_{t_k}}{\delta}\Big]ds\end{aligned}$$

Then repeating similar arguments in the proof of Lemma 2.2, we have that

$$\begin{aligned}
&\int_0^t \mathbb{E}_\Omega[\nabla_p\phi(x_t^\delta,p_t^\delta)(-\eta\nabla_x\sigma(X_t^\delta))\dot{\xi}_\delta(t)\Big]ds \\
&=\int_0^t \mathbb{E}_\Omega[\nabla_p\phi(x^\delta_{[t]_\delta\delta},p^\delta_{[t]_\delta\delta})(-\eta\nabla_x\sigma(x^\delta_{[t]_\delta\delta}))\dot{\xi}_\delta(t)\Big]ds \\
&+\int_0^t \frac{1}{2}\mathbb{E}_\Omega[(\Delta_{pp}\phi(X^\delta_{[t]_\delta\delta},p_{[t]_\delta\delta})(-\eta\nabla_x\sigma(x^\delta_{[t]_\delta\delta}))(-\eta\nabla_x\sigma(x^\delta_{[t]_\delta\delta}))(\dot{\xi}_\delta(t))^2\Big]ds \\
&+o(\delta^\beta),
\end{aligned}$$

where $\beta\in(0,\frac{1}{2})$. Taking $\delta\to 0$ yield that the second order Vlasov equation

$$\begin{aligned}
\partial_t F(t,x,p) &= -\nabla_x\cdot(F(t,x,p)\frac{\partial H_0}{\partial p})+\nabla_p\cdot(F(t,x,p)\frac{\partial H_0}{\partial x}) \\
&+\frac{1}{2}\Delta_{pp}F(t,x,p)\cdot(\frac{\partial H_1}{\partial x},\frac{\partial H_1}{\partial x}).
\end{aligned}$$

This implies that when we consider the joint distribution on Ω, the density function satisfies the second order Vlasov equation. However, when we consider the conditional probability on $\widetilde{\Omega}$ instead of Ω, the conditional joint probability of Wong–Zakai approximation satisfies the following first order Vlasov equation,

$$\begin{aligned}
\partial_t F^\delta(t,x,p) &= -\nabla_x\cdot(F^\delta(t,x,p)\frac{\partial H_0}{\partial p})+\nabla_p\cdot(F^\delta(t,x,p)\frac{\partial H_0}{\partial x}) \\
&+\nabla_p\cdot(F^\delta(t,x,p)\frac{\partial H_1}{\partial x})\dot{\xi}_\delta.
\end{aligned}$$

Its limit equation becomes

$$\begin{aligned}
dF(t,x,p) &= -\nabla_x\cdot(F(t,x,p)\frac{\partial H_0}{\partial p})dt+\nabla_p\cdot(F(t,x,p)\frac{\partial H_0}{\partial x})dt \\
&+\nabla_p\cdot(F(t,x,p)\frac{\partial H_1}{\partial x})\circ dB_t.
\end{aligned}$$

3.2 Stochastic Euler-Lagrange equation in density space

In this section, we consider the kinetic Wasserstein Hamiltonian flow with random perturbation, i.e., the second order stochastic Euler-Lagrange equation from the Lagrange functional on density manifold. Let $\mathscr{M}=(\mathbb{T}^d,\mathbb{I})$. The density space $\mathscr{P}(\mathscr{M})$ is defined by

$$\mathscr{P}(\mathscr{M})=\{\rho dvol_{\mathscr{M}}|\rho\in\mathscr{C}^\infty(\mathscr{M}),\rho\geq 0,\int_{\mathscr{M}}\rho dvol_{\mathscr{M}}=1\}.$$

Its interior of $\mathscr{P}(\mathscr{M})$ is denoted by $\mathscr{P}_o(\mathscr{M})$. The tangent space at $\rho\in\mathscr{P}_o(\mathscr{M})$ is defined by

$$\mathscr{T}_\rho\mathscr{P}_o(\mathscr{M})=\{\kappa\in\mathscr{C}^\infty(\mathscr{M})|\int_{\mathscr{M}}\kappa dvol_{\mathscr{M}}=0\}.$$

Define the quotient space of smooth functions $\mathscr{F}(\mathscr{M})/\mathbb{R}=\{[\Phi]|\Phi\in\mathscr{C}^\infty(\mathscr{M})\}$, where $[\Phi]=\{\Phi+c|c\in\mathbb{R}\}$. Then one could identify the element in $\mathscr{F}(\mathscr{M})/\mathbb{R}$ as the tangent vector in $T_\rho\mathscr{P}_o(\mathscr{M})$ by using the map

$\Theta:\mathscr{F}(\mathscr{M})/\mathbb{R}\to\mathscr{T}_\rho\mathscr{P}_o(\mathscr{M})$, $\Theta_\Phi=-\nabla\cdot(\rho\nabla\Phi)$. The boundaryless condition of $\mathscr{M}$ and the property of elliptical operator ensures that Θ is one to one and linear [15]. This implies that $\mathscr{F}(\mathscr{M})/\mathbb{R}\cong\mathscr{T}^*_\rho\mathscr{P}_o(\mathscr{M})$, where $\mathscr{T}^*_\rho\mathscr{P}_o(\mathscr{M})$ is the cotangent space of $\mathscr{P}_o(\mathscr{M})$. The L^2-Wasserstein metric on density manifold $g_W:\mathscr{T}_\rho\mathscr{P}(\mathscr{M})\times\mathscr{T}_\rho\mathscr{P}(\mathscr{M})\to\mathbb{R}$ is defined by

$$g_W(\kappa_1,\kappa_2)=\int_{\mathscr{M}}\langle\nabla\Phi_1,\nabla\Phi_2\rangle\rho dvol_{\mathscr{M}}=\int_{\mathscr{M}}\kappa_1(-\Delta_\rho)^\dagger\kappa_2dvol_{\mathscr{M}},$$

where $\kappa_1=\Theta_{\Phi_1}$, $\kappa_2=\Theta_{\Phi_2}$, and $(-\Delta_\rho)^\dagger$ is the pseudo inverse operator of $-\Delta_\rho$. In the deterministic case, it is known that the critical point of

$$\frac{1}{2}W^2(\rho^0,\rho^1):=\inf_{\rho_t\in\mathscr{P}_o(\mathscr{M})}\Big\{\int_0^1\int_{\mathscr{M}}\frac{1}{2}g_W(\partial_t\rho_t,\partial\rho_t)dvol_{\mathscr{M}}dt\Big\}$$

satisfies the geodesic equation in cotangent bundle on density manifold (see e.g. [18]), that is,

$$\begin{aligned}\partial_t\rho_t&=-\nabla\cdot(\rho_t\nabla\Phi_t),\\ \partial_t\Phi_t&=-\frac{1}{2}|\nabla\Phi_t|^2+C_t,\end{aligned}$$

where $\Phi_t=(-\Delta_{\rho_t})^\dagger\partial_t\rho_t$, C_t is independent of $x\in\mathscr{M}$. The above geodesic equation in primal coordinates is the Euler–Lagrange equation,

$$\partial_t\frac{\delta}{\delta\partial_t\rho_t}\mathscr{L}(\rho_t,\partial_t\rho_t)=\frac{\delta}{\delta\rho_t}\mathscr{L}(\rho_t,\partial_t\rho_t)+C_t,$$

where $\mathscr{L}(\rho_t,\partial_t\rho_t)=\frac{1}{2}g_W(\partial_t\rho_t,\partial_t\rho_t)$.

Next, we consider the Lagrangian in density manifold with random perturbation,

$$\mathscr{L}(\rho_t,\partial_t\rho_t)=\frac{1}{2}g_W(\partial_t\rho_t,\partial_t\rho_t)-\mathscr{F}(\rho_t)-\Sigma(\rho_t)\dot{\xi}_\delta(t),$$

and its variational problem $I_\delta(\rho^0,\rho^T)=\inf_{\rho_t}\{\int_0^T\mathscr{L}(\rho_t,\partial_t\rho_t)dt|\rho_0=\rho^0,\rho_T=\rho^T\}$. Recall that by (3.3), we have that

$$\Delta_{\rho_t}(\cdot)=-\nabla\cdot(\rho_t\nabla(\cdot)),\ \Delta_{\partial_t\rho_t}(\cdot)=-\nabla\cdot(\partial_t\rho_t\nabla(\cdot)).$$

Theorem 3.1 *The Euler Lagrangian equation of the variational problem $I_\delta(\rho^0,\rho^T)$ satisfies*

$$\partial_{tt}\rho_t+\Gamma_W(\partial_t\rho_t,\partial_t\rho_t)=-grad_W\mathscr{F}(\rho_t)-grad_W\Sigma(\rho_t)\dot{\xi}_\delta,\tag{3.7}$$

where $grad_W\mathscr{F}(\rho_t)=-\nabla\cdot(\rho_t\nabla\frac{\delta}{\delta\rho_t}\mathscr{F}(\rho_t))$, $\Gamma_W(\partial_t\rho_t,\partial_t\rho_t)=\Delta_{\partial_t\rho_t}(-\Delta_{\rho_t})^\dagger\partial_t\rho_t+\frac{1}{2}\Delta_{\rho_t}|\nabla(-\Delta_{\rho_t})^\dagger\rho_t|^2$. Furthermore, Eq. (3.7) *can be formulated as the following Hamiltonian system*

$$\begin{aligned}&\partial_t\rho_t+\nabla\cdot(\rho_t\nabla\Phi_t)=0,\\ &\partial_t\Phi_t+\frac{1}{2}|\nabla\Phi_t|^2=-\frac{\delta}{\delta\rho_t}\mathscr{F}(\rho_t)-\frac{\delta}{\delta\rho_t}\Sigma(\rho_t)\dot{\xi}_\delta,\end{aligned}\tag{3.8}$$

where $\Phi_t=(-\Delta_{\rho_t})^\dagger\partial_t\rho_t$ up to a spatially constant stochastic process shift.

Proof Consider a smooth perturbation εh_t satisfying $\int_{\mathcal{M}} h_t dvol_{\mathcal{M}} = 0$, $t\in[0,T]$ and $h_0 = h_T = 0$. Applying Taylor expansion with respect ε and integration by parts, using $h_0 = h_T = 0$ and the fact that $\mathcal{M}$ is compact, we get

$$\begin{aligned}&\int_0^T \mathscr{L}(\rho_t+\varepsilon h_t, \partial_t\rho_t+\varepsilon\partial_t h_t)dt\\&=\int_0^T \mathscr{L}(\rho_t,\partial_t\rho_t)dt+\varepsilon\int_0^T\int_{\mathcal{M}}\Big(\frac{\delta}{\delta\rho_t}\mathscr{L}(\rho_t,\partial_t\rho_t)-\partial_t\frac{\delta}{\delta\partial_t\rho_t}\mathscr{L}(\rho_t,\partial_t\rho_t)\Big)\cdot h_t dvol_{\mathcal{M}}dt+o(\varepsilon).\end{aligned}$$

Similar to the proof of [15, Theorem 1], direct calculations lead to

$$\begin{aligned}\partial_t\frac{\delta}{\delta\partial_t\rho_t}\mathscr{L}(\rho_t,\dot\rho_t)&=\partial_t((-\Delta_{\rho_t})^{\dagger}\partial_t\rho_t)\\&=(-\Delta_{\rho_t})^{\dagger}\partial_{tt}\rho_t-(-\Delta_{\rho_t})^{\dagger}(-\Delta_{\partial_t\rho_t})(-\Delta_{\rho_t})^{\dagger}\partial_t\rho_t,\\\frac{\delta}{\delta\rho_t}\mathscr{L}(\rho_t,\dot\rho_t)&=-\frac{1}{2}\nabla|(-\Delta_{\rho_t}^{\dagger})\partial_t\rho_t|^2-\frac{\delta}{\delta\rho_t}\mathscr{F}(\rho_t)-\frac{\delta}{\delta\rho_t}\Sigma(\rho_t)\dot\xi_\delta(t),\end{aligned}$$

which, together with the property $\int_{\mathcal{M}} h_t dvol_{\mathcal{M}} = 0$, yields (3.7) up to a spatially-constant stochastic process shift by multiplying Δ_{ρ_t} on both sides. By introducing the Legendre transformation $\Phi_t = (-\Delta_{\rho_t})^{\dagger}\partial\rho_t$, we obtain Eq. (3.8) from Eq. (3.7).

Note that the formulation Γ_W for $\partial_t\rho$ is called as the Christoffel symbol in density manifold [15]. The dual coordinate $\Phi_t = (-\Delta_{\rho_t})^{\dagger}\partial_t\rho_t$ is obtained via the Legendre transformation, which is the key to derive the kinetic Hamiltonian formulation. However, it is still hard to use the Christoffel symbol and Lagrangian functional to derive general stochastic Wasserstein Hamiltonian systems.

Proposition 3.2 *The Euler–Lagrange equation of the variational problem $I(\rho^0,\rho^T)$,*

$$I(\rho_0,\rho_T)=\int_0^T\Big(\frac{1}{2}g_W(\partial_t\rho_t,\partial_t\rho_t)-\mathscr{F}(\rho_t)\Big)dt-\int_0^T\Sigma(\rho_t)\circ dB(t)$$

satisfies

$$\partial_{tt}\rho_t+\Gamma_W(\partial_t\rho_t,\partial_t\rho_t)=-grad_W\mathscr{F}(\rho_t)-grad_W\Sigma(\rho_t)\circ dB_t, \tag{3.9}$$

where ρ_t is $\mathbb{F}_t$-measurable. Furthermore, Eq. (3.9) *can be formulated as the following Hamiltonian system*

$$\begin{aligned}&\partial_t\rho_t+\nabla\cdot(\rho_t\nabla\Phi_t)=0,\\&\partial_t\Phi_t+\frac{1}{2}|\nabla\Phi_t|^2=-\frac{\delta}{\delta\rho_t}\mathscr{F}(\rho_t)-\frac{\delta}{\delta\rho_t}\Sigma(\rho_t)\circ dB_t,\end{aligned}\tag{3.10}$$

where $\Phi_t = (-\Delta_{\rho_t})^{\dagger}\partial_t\rho_t$ up to a spatially constant stochastic process shift.

Proof Consider a smooth perturbation εh_t satisfying $\int_{\mathcal{M}} h_t dvol_{\mathcal{M}} = 0$, $t\in[0,T]$ and $h_0 = h_T = 0$. Denote $\mathscr{L}_0(\rho_t,\partial_t\rho_t)=\frac{1}{2}g_W(\partial_t\rho_t,\partial_t\rho_t)-\mathscr{F}(\rho_t)$. Recall the equivalence of stochastic integrals between Itô sense and Stratonovich sense (see e.g. [37]), i.e., for $M(t)=\int_0^t X(s)\circ dW(s)$, it holds that $M(t)=\int_0^t X(s)dW(s)+\frac{1}{2}\langle M(\cdot)\rangle_t$. Here $X(s)$ is $\mathbb{F}_t$-measurable such that the quadratic variation process $\langle M(\cdot)\rangle_s$ is well-defined for $s\geq 0$. By our assumption that $\partial_t\rho_t\in\mathscr{T}_{\rho_t}\mathscr{P}(\mathcal{M})$, there exists some Φ_t such that $\Phi_t=(-\Delta_{\rho_t})^{\dagger}\partial_t\rho_t$. This yields that $\int_0^t \Sigma(\rho_t)\circ dB(t) = \int_0^t \Sigma(\rho_t)dB(t)$ and that $\int_0^t \int_{\mathscr{M}} \frac{\delta}{\delta\rho_t}\Sigma(\rho_t)\cdot h_t dvol_{\mathscr{M}} \circ dB(t) = \int_0^t \int_{\mathscr{M}} \frac{\delta}{\delta\rho_t}\Sigma(\rho_t)\cdot h_t dvol_{\mathscr{M}} dB(t)$ since their quadratic variation processes are 0. As a consequence, we have that

$$\begin{aligned}
&\int_0^T \frac{1}{2} g_W(\partial_t\rho_t+\varepsilon h_t,\partial_t\rho_t+\varepsilon h_t)-\mathscr{F}(\rho_t+\varepsilon h_t)dt-\int_0^T \Sigma(\rho_t+\varepsilon h_t)dB_t\\
&=\int_0^T \mathscr{L}_0(\rho_t,\partial_t\rho_t)dt+\int_0^T \Sigma(\rho_t)dB_t\\
&+\varepsilon\int_0^T\int_{\mathscr{M}}(\frac{\delta}{\delta\rho_t}\mathscr{L}_0(\rho_t,\partial_t\rho_t)-\partial_t\frac{\delta}{\delta\partial_t\rho_t}\mathscr{L}_0(\rho_t,\partial_t\rho_t))\cdot h_t dvol_{\mathscr{M}}dt\\
&+\varepsilon\int_0^T\int_{\mathscr{M}}\frac{\delta}{\delta\rho_t}\Sigma(\rho_t)\cdot h_t dvol_{\mathscr{M}}dB_t+o(\varepsilon).
\end{aligned}$$

Similar to the proof of Theorem 3.1, we obtain (3.9) and its equivalent Hamiltonian system (3.10).

3.3 Generalized stochastic Wasserstein–Hamiltonian flow

In the last section, we show that the density of a Hamiltonian ODE with random perturbation satisfies the stochastic Wasserstein Hamiltonian flow. In this section, We derived the general stochastic Wasserstein Hamiltonian flow via the random perturbation in the dual coordinates in density space. It provides a more general framework that can derive a large class of stochastic Wasserstein Hamiltonian flows which can not be obtained from the classic dynamics with perturbations.

Let $\mathscr{M}=(\mathbb{T}^d,\mathbb{I})$. We introduce the following variational problem

$$I_\delta(\rho^0,\rho^T)=\inf\{\mathscr{S}(\rho_t,\Phi_t)|\Delta_{\rho_t}\Phi_t\in\mathscr{T}_{\rho_t}\mathscr{P}_o(\mathscr{M}),\rho(0)=\rho^0,\rho(T)=\rho^T\} \tag{3.11}$$

whose action functional is given by the dual coordinates,

$$\mathscr{S}(\rho_t,\Phi_t)=-\int_0^T\langle\Phi(t),\partial_t\rho_t\rangle+\mathscr{H}_0(\rho_t,\Phi_t)dt+\int_0^T\mathscr{H}_1(\rho_t,\Phi_t)d\xi_\delta(t).$$

Here $\mathscr{H}_0(\rho_t,\Phi_t)=\int_{\mathscr{M}}\frac{1}{2}|\nabla\Phi_t|^2\rho_t dvol_{\mathscr{M}}+\mathscr{F}(\rho_t)$, $\mathscr{H}_1(\rho_t,\Phi_t)=\eta\int_{\mathscr{M}}\frac{1}{2}|\nabla\Phi_t|^2\rho_t dvol_{\mathscr{M}}+\eta\Sigma(\rho_t)$, $\mathscr{F}$ and Σ are smooth potential functions.

Theorem 3.2 *The critical point of the variational problem $I_\delta(\rho^0,\rho^T)$ satisfies the following Hamiltonian system*

$$\begin{aligned}
&\partial_t\rho_t+\nabla\cdot(\rho_t\nabla\Phi_t)+\eta\nabla\cdot(\rho_t\nabla\Phi_t)\dot{\xi}_\delta=0,\\
&\partial_t\Phi_t+\frac{1}{2}|\nabla\Phi_t|^2+\eta\frac{1}{2}|\nabla\Phi_t|^2\dot{\xi}_\delta=-\frac{\delta}{\delta\rho_t}\mathscr{F}(\rho_t)-\eta\frac{\delta}{\delta\rho_t}\Sigma(\rho_t)\dot{\xi}_\delta,
\end{aligned} \tag{3.12}$$

where $(1+\dot{\xi}_\delta(t))\Phi_t=(-\Delta_{\rho_t})^\dagger\partial_t\rho_t$ up to a spatially constant stochastic process shift.

Proof Consider the perturbations on ρ and Φ. Following the arguments in the proof of Proposition 3.2, the critical point satisfies that

$$\begin{aligned}
&\mathscr{S}(\rho_t+\varepsilon\delta\rho_t,\Phi_t+\varepsilon\delta\Phi_t)\\
&=\mathscr{S}(\rho_t,\Phi_t)-\varepsilon\int_0^T\langle\Phi(t),\partial_t\delta\rho_t\rangle dt-\varepsilon\int_0^T\langle\delta\Phi(t),\partial_t\rho_t\rangle dt\\
&+\varepsilon\int_0^T\frac{\delta}{\delta\rho_t}\mathscr{H}_0(\rho_t,\Phi_t)\delta\rho_t+\frac{\delta}{\delta\Phi_t}\mathscr{H}_0(\rho_t,\Phi_t)\delta\Phi_t dt\\
&+\varepsilon\int_0^T\frac{\delta}{\delta\rho_t}\mathscr{H}_1(\rho_t,\Phi_t)\delta\rho_t+\frac{\delta}{\delta\Phi_t}\mathscr{H}_1(\rho_t,\Phi_t)\delta\Phi_t d\xi_\delta(t)+o(\varepsilon)\\
&=\mathscr{S}(\rho_t,\Phi_t)+\varepsilon\int_0^T\langle\partial_t\Phi(t),\delta\rho_t\rangle dt-\varepsilon\int_0^T\langle\delta\Phi(t),\partial_t\rho_t\rangle dt\\
&+\varepsilon\int_0^T\langle\frac{\delta}{\delta\rho_t}\mathscr{H}_0(\rho_t,\Phi_t),\delta\rho_t\rangle+\langle\frac{\delta}{\delta\Phi_t}\mathscr{H}_0(\rho_t,\Phi_t),\delta\Phi_t\rangle dt\\
&+\varepsilon\int_0^T\langle\frac{\delta}{\delta\rho_t}\mathscr{H}_1(\rho_t,\Phi_t),\delta\rho_t\rangle+\langle\frac{\delta}{\delta\Phi_t}\mathscr{H}_1(\rho_t,\Phi_t),\delta\Phi_t\rangle d\xi_\delta(t)+o(\varepsilon).
\end{aligned}$$

Taking $\varepsilon\to 0$, we obtain that

$$\begin{aligned}
\partial_t\rho_t&=\frac{\delta}{\delta\Phi_t}\mathscr{H}_0(\rho_t,\Phi_t)+\frac{\delta}{\delta\Phi_t}\mathscr{H}_0(\rho_t,\Phi_t)\dot{\xi}_\delta(t)\\
\partial_t\Phi_t&=-\frac{\delta}{\delta\rho_t}\mathscr{H}_0(\rho_t,\Phi_t)-\frac{\delta}{\delta\rho_t}\mathscr{H}_0(\rho_t,\Phi_t)\dot{\xi}_\delta(t),
\end{aligned}$$

which leads to (3.12).

Similarly, consider the action functional

$$\widetilde{\mathscr{S}}(\rho_t,\Phi_t)=-\int_0^T\langle\Phi(t),\circ d\rho_t\rangle+\mathscr{H}_0(\rho_t,\Phi_t)dt+\int_0^T\mathscr{H}_1(\rho_t,\Phi_t)\circ dB_t$$

over the $\mathbb{F}_t$-adapted feasible set, we obtain the following result.

Theorem 3.3 *The critical point of the variational problem $I(\rho^0,\rho^T)$ defined by*

$$I(\rho^0,\rho^T)=\inf\{\widetilde{\mathscr{S}}(\rho_t,\Phi_t)|\rho(0)=\rho^0,\rho(T)=\rho^T\}$$

satisfies the following Hamiltonian system

$$\begin{aligned}
&\partial_t\rho_t+\nabla\cdot(\rho_t\nabla\Phi_t)+\eta\nabla\cdot(\rho_t\nabla\Phi_t)\circ dB_t=0,\\
&\partial_t\Phi_t+\frac{1}{2}|\nabla\Phi_t|^2+\eta\frac{1}{2}|\nabla\Phi_t|^2\circ dB_t=-\frac{\delta}{\delta\rho_t}\mathscr{F}(\rho_t)-\eta\frac{\delta}{\delta\rho_t}\Sigma(\rho_t)\circ dB_t
\end{aligned}\tag{3.13}$$

up to a spatially constant stochastic process shift on Φ_t.

Next, we show that the continuity equation and the velocity equation generated by Φ,

$$\begin{aligned}
&\partial_t\rho_t+\nabla\cdot(\rho_t v_t)+\eta\nabla\cdot(\rho_t v_t)\dot{\xi}_\delta=0,\\
&\partial_t v_t+\nabla v_t\cdot v_t+\eta\nabla v_t\cdot v_t\dot{\xi}_\delta=-\nabla\frac{\delta}{\delta\rho_t}\mathscr{F}(\rho_t)-\eta\frac{\delta}{\delta\rho_t}\nabla\Sigma(\rho_t)\dot{\xi}_\delta
\end{aligned}\tag{3.14}$$

is convergent to the corresponding system driven by the Brownian motion.

Proposition 3.3 *Assume that* $v(0,\cdot),\rho(0,\cdot)$ *is* $\mathbb{F}_0$*-measurable and smooth,* $\mathscr{F}(\rho_t)=\int_{\mathscr{M}} f\rho_t dvol_{\mathscr{M}}$ *and* $\Sigma(\rho_t)=\int_{\mathscr{M}}\sigma\rho_t dvol_{\mathscr{M}}$ *with* $f,\sigma\in C_p^3(\mathscr{M})$*. Let* ρ^δ,v^δ *be the solution of* (3.14)*, and* ρ,v *be the solution of*

$$\begin{aligned}&\partial_t\rho_t+\nabla\cdot(\rho_t v_t)+\eta\nabla\cdot(\rho_t v_t)\circ dB_t=0,\\&\partial_t v_t+\nabla v_t\cdot v_t+\eta\nabla v_t\cdot v_t\circ dB_t=-\nabla\frac{\delta}{\delta\rho_t}\mathscr{F}(\rho_t)-\eta\nabla\frac{\delta}{\delta\rho_t}\Sigma(\rho_t)\circ dB_t.\end{aligned}\tag{3.15}$$

Then there exists a stopping time $\tau>0$ *such that for any* $\delta>0$,

$$\lim_{\delta\to0}\mathbb{P}(\sup_{t\in[0,\tau)}[|\rho_t^\delta-\rho_t|_{L^\infty(\mathscr{M})}+|v_t^\delta-v_t|_{L^\infty(\mathscr{M})}]>\varepsilon)=0.$$

Proof Since $\mathscr{M}$ is compact, $f,\sigma\in C_p^3(\mathscr{M})$, similar to the proofs of Lemma 2.2 and Lemma 2.3, we can obtain the global well-posedness of the particle ODE systems

$$\begin{aligned}dX_t&=v(t,X_t)dt+\eta v(t,X_t)\circ dB_t,\\dv(t,X_t)&=-\nabla f(X_t)dt-\eta\nabla\sigma(X_t)\circ dB_t,\end{aligned}$$

and

$$\begin{aligned}dX_t^\delta&=v^\delta(t,X_t^\delta)dt+\eta v(t,X_t^\delta)d\xi_\delta,\\dv^\delta(t,X_t^\delta)&=-\nabla f(X_t^\delta)dt-\eta\nabla\sigma(X_t^\delta)d\xi_\delta.\end{aligned}$$

Following the arguments in the proof Proposition 3.1, we can obtain that there exists a stopping time $\tau>0$ such that X_t is a smooth diffeomorphism before τ. Notice that the density function $\rho^\delta(t,y)$ of X_t^δ satisfies $\rho^\delta(t,y)=|\det(\nabla X_t^\delta(y))|\rho(0,X_t^\delta(y))$. Since $\rho(0,\cdot)$ is smooth for any fixed ω and the pathwise convergence of X^δ holds, it follows that $\rho^\delta(t,y)$ converges to the density function of X_t before τ, which is $\rho(t,y)=|\det(\nabla X_t(y))|\rho(0,X_t(y))$. Similarly, the pathwise convergence of $v^\delta(t,X_t^\delta(y))$ to $v(t,X_t(y))$, together with invertibility of X_t^δ and X_t, implies the convergence of $v^\delta(t,x)$ to $v(t,x)$ before τ.

Remark 3.1 *If one obtains the convergence of the Wong–Zakai approximations of the mean-field SODEs,*

$$\begin{aligned}dX_t&=v(t,X_t)dt+\eta v(t,X_t)\circ dB_t,\\dv(t,X_t)&=-\nabla\frac{\delta}{\delta\rho(t,X_t)}\mathscr{F}(\rho(t,X_t))dt-\eta\nabla\frac{\delta}{\delta\rho(t,X_t)}\Sigma(t,X_t)\circ dB_t,\end{aligned}$$

then the convergence of (3.14) *to* (3.15) *can be shown similarly before the stopping time* τ*, that is, the first time* X_t *is not a smooth diffeomorphism on* $\mathscr{M}$ *or* X_t *escapes* $\mathscr{M}$*.*

4 Examples

In this section, we show that both the stochastic nonlinear Schrödinger (NLS) equation, which models the propagation of nonlinear dispersive waves in random or inhomogenous media in quantum physics (see e.g. [5,23,26,38,53]), and nonlinear Schrödinger equation with random dispersion, which describes the propagation of a signal in an optical fibre with dispersion management (see e.g. [1,2]), are stochastic Wasserstein-Hamiltonian flows. We also discuss that the mean-field game system with common noise (see e.g. [57,52,56]) is a stochastic Wasserstein-Hamiltonian flow under suitable transformations.

4.1 Stochastic NLS equation

The dimensionless stochastic NLS equation is given by

$$du = \mathbf{i}\Delta u dt + \mathbf{i}\lambda f(|u|^2)udt + \mathbf{i}u \circ dW_t, \tag{4.1}$$

where W_t is a Q-Wiener process on the Hilbert space $L^2(\mathscr{M};\mathbb{R})$ and f is a real-valued continuous function. Since the Q-Wiener process W has the Karhunen–Loève expansion $W(t,x) = \sum_{i\in\mathbb{N}^+} Q^{\frac{1}{2}} e_i(x)\beta_i(t)$ (see e.g. [24]), where $\{e_i\}_{i\in\mathbb{N}}$ is an orthonormal basis of $L^2(\mathscr{M};\mathbb{R})$, and $\{\beta_i\}_{i\in\mathbb{N}}$ is a sequence of linearly independent Brownian motions on $(\Omega, \mathbb{F}, \{\mathbb{F}_t\}_{t\geq 0}, \mathbb{P})$. We denote $W_\delta(t,x) = \sum_{i\in\mathbb{N}^+} Q^{\frac{1}{2}} e_i(x)\beta_i^\delta(t)$ as the piecewise linear Wong–Zakai approximation (or other type Wong–Zakai approximation) of W and consider the approximated NLS equation of (4.1)

$$\partial_t u(t,x) = \mathbf{i}\Delta_{xx}u(t,x) + \mathbf{i}\lambda f(|u(t,x)|^2)u(t,x) + \mathbf{i}u(t,x)\dot{W}_\delta(t,x). \tag{4.2}$$

We aim to prove that (4.2) is a stochastic Wasserstein Hamiltonian flow for any $\delta > 0$, and thus its limit (4.1) is also a stochastic Wasserstein Hamiltonian flow. In the following, we assume that f is a real-value function, W is smooth with respect to the space variable, and (4.2) possesses a mild solution or a strong solution on $[0,T]$.

Denote the L^2-inner product by $\langle u, v\rangle = \Re\int_{\mathscr{M}} \bar{u}vdvol_M$, where $\Re$ is the real part of a complex number. The variational problem on density manifold of (4.2) is

$$I_\delta(\rho^0,\rho^T) = \inf\{\mathscr{S}(\rho_t,\Phi_t)|\Delta_{\rho_t}\Phi_t \in \mathscr{T}_{\rho_t}\mathscr{P}_o(\mathscr{M}), \rho(0)=\rho^0, \rho(T)=\rho^T\} \tag{4.3}$$

whose action functional is given by the dual coordinates,

$$\mathscr{S}(\rho_t,\Phi_t) = -\int_0^T \langle \Phi(t), \partial_t\rho_t\rangle dt + \int_0^T \mathscr{H}_0(\rho_t,\Phi_t)dt + \sum_{i\in\mathbb{N}^+}\int_0^T \mathscr{H}_i(\rho_t,\Phi_t)d\beta_i^\delta(t).$$

Here $\mathscr{H}_0(\rho_t,\Phi_t) = \int_{\mathscr{M}} |\nabla\Phi_t|^2\rho_t dvol_{\mathscr{M}} + \frac{1}{4}I(\rho) + \mathscr{F}(\rho_t)$, $\mathscr{H}_i(\rho_t,\Phi_t) = -\Sigma_i(\rho_t) = -\int_{\mathscr{M}} Q^{\frac{1}{2}} e_i\rho_t dvol_{\mathscr{M}}$, $\mathscr{F}(\rho) = -\frac{\lambda}{2}\int_{\mathscr{M}}\int_0^\rho f(s)dsdvol_{\mathscr{M}}$ with a smooth function f, and $I(\rho) = \int_{\mathscr{M}} |\nabla\log(\rho)|^2\rho d\mathrm{vol}_{\mathscr{M}}$.

In the following, we show the relationship between the the variational problem (4.3) and nonlinear Schrödinger equation with Wong–Zakai approximation (4.2) by using the Madelung transform [44].

Proposition 4.1 *The critical point of the variational problem* (4.3) *satisfies the Madelung system of* (4.2) *on the support of* ρ_t. *Conversely, the Madelung transform of* (4.2) *satisfies the critical point of* (4.3) *on the support of* $|u_t|$.

Proof By studying the perturbation on the dual coordinates, the arguments in the proof of Theorem 3.2 yield that the critical point of (4.3) satisfies

$$\begin{aligned}
&\partial_t\rho_t + 2\nabla\cdot(\rho_t\nabla\Phi_t) = 0,\\
&\partial_t\Phi_t + |\nabla\Phi_t|^2 = -1/4\frac{\delta}{\delta\rho_t}I(\rho_t) - \frac{\delta}{\delta\rho_t}\mathscr{F}(\rho_t) - \dot{W}_\delta.
\end{aligned}$$

Define a complex valued function by $\widehat{u}(t,x) = \sqrt{\rho(t,x)}e^{\mathbf{i}\Phi(t,x)}$. One obtains the equation of $\widehat{u}(t,x)$ satisfying (4.2) on the support of ρ_t by direct calculations.

Conversely, using the Madelung transform of the solution $\sqrt{\rho(t,x)}e^{\mathbf{i}S(t,x)} = u(t,x)$ where $\rho = |u|^2$ for (4.2). Then direct calculation leads to

$$
\begin{aligned}
& e^{\frac{1}{2}\log(\rho)+\mathbf{i}S}(\frac{1}{2}\frac{\partial_t \rho}{\rho} + \mathbf{i}\partial_t S) \\
&= \mathbf{i}e^{\frac{1}{2}\log(\rho)+\mathbf{i}S}(\frac{1}{2}\frac{\nabla\rho}{\rho} + \mathbf{i}\nabla S)^2 + \mathbf{i}e^{\frac{1}{2}\log(\rho)+\mathbf{i}S}(\frac{1}{2}\frac{\Delta\rho}{\rho} + \mathbf{i}\Delta S - \frac{1}{2}|\frac{\nabla\rho}{\rho}|^2) \\
&+ \mathbf{i}e^{\frac{1}{2}\log(\rho)+\mathbf{i}S}(\lambda f(\rho) + \dot{W}_\delta) \\
&= \mathbf{i}e^{\frac{1}{2}\log(\rho)+\mathbf{i}S}(\frac{1}{4}(\frac{\nabla\rho}{\rho})^2 - (\nabla S)^2 + \mathbf{i}\frac{\nabla\rho}{\rho}\cdot\nabla S) + \mathbf{i}e^{\frac{1}{2}\log(\rho)+\mathbf{i}S}(\frac{1}{2}\frac{\Delta\rho}{\rho} + \mathbf{i}\Delta S - \frac{1}{2}|\frac{\nabla\rho}{\rho}|^2) \\
&+ \mathbf{i}e^{\frac{1}{2}\log(\rho)+\mathbf{i}S}(\lambda f(\rho) + \partial_t W_\delta).
\end{aligned}
$$

This implies that on the support or $|u_t|$, it holds that

$$
\begin{aligned}
\partial_t \rho &= -2\nabla\cdot(\rho\nabla S), \\
\partial_t S &= -|\nabla S|^2 - \frac{1}{4}\frac{\delta}{\delta\rho}I(\rho) + \lambda f(\rho) + \dot{W}_\delta.
\end{aligned} \tag{4.4}
$$

Based on the above result, taking spatial gradient on the potential S, we get the following system with the conservation law

$$
\begin{aligned}
\partial_t \rho &= -\nabla\cdot(\rho v), \\
\partial_t v &= -\nabla_x v\cdot v - \nabla_x \frac{1}{2}\frac{\delta}{\delta\rho}I(\rho) + 2\lambda\nabla_x f(\rho) + 2\nabla_x \dot{W}_\delta,
\end{aligned} \tag{4.5}
$$

where $v(t,x) = 2\nabla S(t,x)$.

The following theorem indicates that the stochastic NLS equation is a stochastic Wasserstein Hamiltonian flow due to the convergence of the Wong–Zakai approximation. For convenience, let us assume that $\mathscr{M} = \mathbb{T}^d$ or $\mathbb{R}^d$ and consider the case that W consists of a finite combinations of independent Brownian motions, i.e., $W(t,x) = \sum_{k=1}^N q_k(x)\beta_k(t)$, with $q_k(x) \in \mathbb{H}^m(\mathscr{M}) \cap W^{k,\infty}(\mathscr{M})$ for some $m \in \mathbb{N}$ and $k \in \mathbb{N}^+$. Here $\mathbb{H}^m(\mathscr{M}), W^{k,\infty}(\mathscr{M})$ are the standard Sobolev space.

Theorem 4.1 *Let $m \in \mathbb{N}$ and $k \in \mathbb{N}^+$. Suppose that the initial value of* (4.2) *and* (4.1) *$u_0 \in \mathbb{H}^m$ is $\mathbb{F}_0$-measurable and has any finite p-moment, $p \in \mathbb{N}^+$, and that f is a real-valued continuous function satisfies*

$$
\begin{aligned}
\|f(|u|^2)u - f(|v|^2)v\| &\le L_f(R)\|u-v\|, \ \|u\|,\|v\| \le R, \\
\|f(|u|^2)u\|_{\mathbb{H}^1} &\le L_f(R)(1+\|u\|_{\mathbb{H}^1}), \ \|u\|_{\mathbb{H}^1} \le R,
\end{aligned}
$$

where $\lim_{R\to\infty} L_f(R) = \infty$. *The Wong–Zakai approximation* (4.2) *is convergent almost surely to the stochastic NLS equation* (4.1) *up to a subsequence.*

Proof Since the driving noise is real-valued, the skew-symmetry of the NLS equation leads to the mass conservation laws for both (4.2) and (4.1). By the local Lipschitz property of $f(|\cdot|^2)(\cdot)$, one can obtain the existence of the unique mild solutions for both (4.2) and (4.1) in $\mathscr{C}([0,T],L^2)$ by a standard argument in [24]. In order to study the converge in L^2, let us define an approximation sequence $u_0^{R_1} \in \mathbb{H}^1, R_1 \to \infty$ of the initial value u_0, which can be taken by using truncated Fourier series or spectral Galerkin method (see e.g. [19]). The growth condition of f in $\mathbb{H}^1$ and the uniform boundedness assumption of q_k lead to

$$\mathbb{E}\Big[\sup_{t\in[0,T]}\|u_t^{R_1}\|_{\mathbb{H}^1}^{2p}\Big]\le C(T,R_1,p)<\infty, \mathbb{E}\Big[\sup_{t\in[0,T]}\|u_t^{\delta,R_1}\|_{\mathbb{H}^1}^{2p}\Big]\le C(T,R_1,\delta,p)<\infty,$$

where $p\ge 1$, $\lim_{R_1\to\infty}C(T,R_1,p)=\infty$, $\lim_{R_1\to\infty}C(T,R_1,\delta,p)=\infty$. Meanwhile, $u_t^{R_1}, u_t^{\delta,R_1}$ are convergent to u_t, u_t^{δ}, a.s. in $\mathscr{C}([0,T];L^2)$ as $R_1\to\infty$, respectively up to a subsequnce. The continuity estimate of $u_t^{R_1}, u_t^{R_1,\delta}$,

$$\mathbb{E}\Big[\|u^{R_1}(t)-u^{R_1}(s)\|^{2p}\Big]\le C(T,R_1,p)|t-s|^p,$$
$$\mathbb{E}\Big[\|u^{R_1,\delta}(t)-u^{R_1,\delta}(s)\|^{2p}\Big]\le C(T,R_1,\delta,p)(|t-s|^p+|\delta|^p),$$

can be obtained due to the mass conservation law and the continuity of $e^{\mathbf{i}\Delta t}$. However, to get the convergence of (4.2), we need a priori estimate of $u^{R_1,\delta}$ which is independent of δ. To this end, we study the enegry of the Wong–Zakai approximation, $H(u)=\int_{\mathscr{M}}\frac{1}{2}|\nabla u|^2 dvol_{\mathscr{M}}-\frac{\lambda}{2}\int_{\mathscr{M}}\int_0^{|u|^2}f(s)dsdvol_{\mathscr{M}}$, and obtain

$$H(u^{\delta}(t))=H(u^{\delta}(0))+\int_0^t\langle\nabla u^{\delta}(s),\mathbf{i}u^{\delta}(s)\nabla dW^{\delta}(s)\rangle.$$

By taking expectation, we get that

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in[0,T]}H(u^{\delta}(t))\Big]\\
&\le\mathbb{E}\Big[H(u^{\delta}(0))\Big]+\mathbb{E}\Big[\sup_{t\in[0,T]}|\int_0^{[t]_\delta}\langle\nabla u^{\delta}([s]_\delta),\mathbf{i}u^{\delta}([s]_\delta)\nabla dW^{\delta}(s)\rangle|\Big]\\
&+\mathbb{E}\Big[\sup_{t\in[0,T]}|\int_{[t]_\delta}^t\langle\nabla u^{\delta}([s]_\delta),\mathbf{i}u^{\delta}([s]_\delta)\nabla dW^{\delta}(s)\rangle|\Big]\\
&+\mathbb{E}\Big[\sup_{t\in[0,T]}|\int_0^{[t]_\delta}\langle\nabla u^{\delta}([s]_\delta),\mathbf{i}(u^{\delta}(s)-u^{\delta}([s]_\delta))\nabla dW^{\delta}(s)\rangle|\Big]\\
&+\mathbb{E}\Big[\sup_{t\in[0,T]}|\int_{[t]_\delta}^t\langle\nabla u^{\delta}([s]_\delta),\mathbf{i}(u^{\delta}(s)-u^{\delta}([s]_\delta))\nabla dW^{\delta}(s)\rangle|\Big]\\
&+\mathbb{E}\Big[\sup_{t\in[0,T]}|\int_0^{[t]_\delta}\langle\nabla(u^{\delta}(s)-u^{\delta}([s]_\delta)),\mathbf{i}u^{\delta}(s)dW^{\delta}(s)\rangle|\Big]\\
&+\mathbb{E}\Big[\sup_{t\in[0,T]}|\int_{[t]_\delta}^t\langle\nabla(u^{\delta}(s)-u^{\delta}([s]_\delta)),\mathbf{i}u^{\delta}(s)dW^{\delta}(s)\rangle|\Big]\\
&=\mathbb{E}\Big[H(u^{\delta}(0))\Big]+V_1+V_2+V_3+V_4+V_5+V_6.
\end{aligned}$$

Below we show the estimates of V_i $(i=1,\cdots,6)$. The Burkholder's inequality and mass conservation law lead to

$$V_1\le\mathbb{E}\Big[\int_0^T C(H(u^{\delta}([t]_\delta))+C(\|u_0\|))ds\Big].$$

Applying the Burkholder and Minkowski inequalities, and the mass conservation law, we achieve that for $T = K\delta$,

$$
\begin{aligned}
V_2 &\le 1 + \mathbb{E}\Big[\sup_{t\in[0,T]} \Big|\int_{[t]_\delta}^t \langle \nabla u^\delta([s]_\delta), \mathbf{i}u^\delta([s]_\delta)\nabla dW^\delta(s)\rangle\Big|^2\Big] \\
&\le 1 + \sum_{k=0}^{K-1} \mathbb{E}\Big[\sup_{t\in[t_k,t_{k+1}]} \Big|\int_{t_k}^t \langle \nabla u^\delta(t_k), \mathbf{i}u^\delta(t_k)\nabla dW(s)\rangle\Big|^2\Big] \\
&\le 1 + C\sum_{k=0}^{K-1} \mathbb{E}\Big[\sum_{i=1}^N \int_{t_k}^{t_{k+1}} \langle \nabla u^\delta(t_k), \mathbf{i}u^\delta(t_k)\nabla q_i(x)\rangle^2 dt\Big] \\
&\le 1 + C\sum_{i=1}^N \mathbb{E}\Big[\|\nabla u^\delta([t]_\delta)\|^2 \|u^\delta([t]_\delta)\|^2 \|q_i\|_{W^{1,\infty}}^2 dt\Big] \\
&\le 1 + C\|u(0)\|^2 \sum_{i=1}^N \|q_i\|_{W^{1,\infty}}^2 \int_0^T \mathbb{E}\Big[\|\nabla u^\delta([t]_\delta)\|^2\Big] dt.
\end{aligned}
$$

The definition of H leads to that there exists a constant $C(\|u_0\|)$ depending on $\|u_0\|$ such that

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in[0,T]} \Big|\int_{[t]_\delta}^t \langle \nabla u^\delta([s]_\delta), \mathbf{i}u^\delta([s]_\delta)\nabla dW^\delta(s)\rangle\Big|^2\Big] \\
&\le 2C\|u_0\|^2 \sum_{i=1}^N \|q_i\|_{W^{1,\infty}}^2 \int_0^T \mathbb{E}\Big[H(u^\delta([t]_\delta))\Big] dt + C(\|u_0\|).
\end{aligned}
$$

The mild form of $u^\delta(s) - u^\delta([s]_\delta)$,

$$
\begin{aligned}
&u^\delta(s) - u^\delta([s]_\delta) \\
&= (e^{\mathbf{i}\Delta(s-[s]_\delta)} - I)u^\delta([s]_\delta) + \int_{[s]_\delta}^s e^{\mathbf{i}\Delta(s-r)}\mathbf{i}\lambda f(|u^\delta(r)|^2)u^\delta(r)dr \\
&+ \int_{[s]_\delta}^s \mathbf{i}e^{\mathbf{i}\Delta(s-r)}u^\delta(r)dW^\delta(r),
\end{aligned}
$$

together with the mass conservation law and $\|e^{\mathbf{i}\Delta t} - I\|_{\mathscr{L}(\mathbb{H}^1,L^2)} \le Ct^{\frac{1}{2}}$ (see, e.g., [24]), yields that

$$
\begin{aligned}
\|u^\delta(s) - u^\delta([s]_\delta)\| \le C\|u^\delta([s]_\delta)\|_{\mathbb{H}^1}\delta^{\frac{1}{2}} + L_f(\|u_0\|)(1+\|u_0\|)\delta \\
+ C\|W([s]_\delta + \delta) - W([s]_\delta)\|\|u_0\|.
\end{aligned} \tag{4.6}
$$

By making use of (4.6) and the Burkholder's inequality, we obtain

$$
\begin{aligned}
V_3 &\le C(1 + \mathbb{E}\Big[\int_0^T \|\nabla u^\delta([s]_\delta)\|^2 ds\Big]) \\
&+ C(\|u_0\|)\mathbb{E}\Big[\int_0^T \|\nabla u^\delta([s]_\delta)\|(1+\|u_0\|)\Big(\frac{\|W([s]_\delta+\delta) - W([s]_\delta)\|_{L^\infty}^2}{\delta} \\
&+ \|W([s]_\delta+\delta) - W([s]_\delta)\|_{L^\infty}\Big) ds\Big] \\
&\le C(\|u_0\|)(1 + \mathbb{E}\Big[\int_0^T H(u^\delta([s]_\delta))ds\Big]).
\end{aligned}
$$

Similar arguments yield that

$$
\begin{aligned}
V_4 &\le C\mathbb{E}\Big[\sup_{t\in[0,T]}\int_{[t]_\delta}^t \|\nabla u^\delta([s]_\delta)\|^2\|W([s]_\delta+\delta)-W([s]_\delta)\|\delta^{-\frac{1}{2}}ds\Big]\\
&+C(\|u_0\|)\mathbb{E}\Big[\sup_{t\in[0,T]}\int_{[t]_\delta}^t \|\nabla u^\delta([s]_\delta)\|(1+\|u_0\|)\Big(\frac{\|W([s]_\delta+\delta)-W([s]_\delta)\|_{L^\infty}^2}{\delta}\\
&+\|W([s]_\delta+\delta)-W([s]_\delta)\|_{L^\infty}\Big)ds\Big]\\
&\le C\delta\mathbb{E}\Big[\sup_{s\in[0,T]}H(u^\delta([s]_\delta))\Big]+C(\|u_0\|)\delta.
\end{aligned}
$$

The estimates of V_5 and V_6 are omitted here since they are very similar to those of V_3 and V_4. We conclude that

$$
\begin{aligned}
&V_1+V_2+V_3+V_4+V_5+V_6\\
&\le C\delta\mathbb{E}\Big[\sup_{t\in[0,T]}H(u^\delta(t))\Big]+C\mathbb{E}\Big[\int_0^T(H(u^\delta([t]_\delta))dt\Big]+C(\|u_0\|).
\end{aligned}
$$

Thus, we obtain $\mathbb{E}\Big[\sup_{t\in[0,T]}H(u^\delta(t))\Big]\le C(T,R_1,\|u_0\|)$ by using Gronwall's inequality and taking δ small enough. Similarly, it holds that for any $p\ge 1$,

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in[0,T]}H^p(u^\delta(t))\Big]\le C(T,R_1,\|u_0\|,p),\\
&\mathbb{E}\Big[\|u^{R_1,\delta}(t)-u^{R_1,\delta}(s)\|^{2p}\Big]\le C(T,R_1,p)(|t-s|^p+|\delta|^p).
\end{aligned}
$$

Next, it suffices to prove the convergence of the Wong–Zakai approximation. To this end, we consider a stopping time $\tau=\inf\{t\in[0,T]|\|u^{R_1}(t)\|\ge R$ or $\|u^{\delta,R_1}([t]_\delta)\|\ge R\}$. In the following, we omit the supindex

R_1. Applying the chain rule, we obtain that for $t \le \tau$,

$$
\begin{aligned}
&\|u(t)-u^\delta(t)\|^2 = \|u(0)-u^\delta(0)\|^2 + 2\int_0^t \langle \mathbf{i}f(|u(s)|^2)u(s) - \mathbf{i}f(|u^\delta(s)|^2)u^\delta(s),\ u(s)-u^\delta(s)\rangle ds \\
&+2\int_0^t \langle u(s)-u^\delta(s), -\frac{1}{2}\sum_{k=1}^N |q_k|^2 u(s)\rangle ds \\
&+2\int_0^t \langle u(s)-u^\delta(s), \mathbf{i}u(s)dW(s) - \mathbf{i}u^\delta(s)dW_\delta(s)\rangle \\
&+\int_0^t \sum_{k=1}^N \int_{\mathscr{M}} |u(s)|^2|q_k|^2 dvol_M ds \\
&\le \int_0^t 2L_f(\|u(0)\|)\|u(s)-u^\delta(s)\|^2 ds + \int_0^t \langle u^\delta(s), \sum_{k=1}^N |q_k|^2 u(s)\rangle ds \\
&-2\int_0^t \langle u(s), \mathbf{i}u^\delta(s)dW^\delta(s)\rangle - 2\int_0^t \langle u^\delta(s), \mathbf{i}u(s)dW(s)\rangle \\
&\le \int_0^t 2L_f(\|u(0)\|)\|u(s)-u^\delta(s)\|^2 ds + \int_0^t \langle u^\delta(s), \sum_{k=1}^N |q_k|^2 u(s)\rangle ds \\
&-2\int_0^t \langle u(s), \mathbf{i}u^\delta([s]_\delta)dW^\delta(s)\rangle - 2\int_0^t \langle u(s), \mathbf{i}(u^\delta(s)-u^\delta([s]_\delta))dW^\delta(s)\rangle \\
&-2\int_0^t \langle u^\delta([s]_\delta), \mathbf{i}u(s)dW(s)\rangle - 2\int_0^t \langle u^\delta(s)-u^\delta([s]_\delta), \mathbf{i}u(s)dW(s)\rangle \\
&=: \int_0^t 2L_f(\|u(0)\|)\|u(s)-u^\delta(s)\|^2 ds + III_1 + III_2 + III_3 + III_4 + III_5.
\end{aligned}
$$

For the term III_2, the property of Wiener process, the mass conservation law, Hölder's and Young's inequality, as well as the property of the martingale, yield that

$$
\begin{aligned}
\mathbb{E}[III_2] &\le -2\int_0^{[t]_\delta} \mathbb{E}\Big[\langle u(s)-u([s]_\delta), \mathbf{i}u^\delta([s]_\delta)dW^\delta(s)\rangle\Big] \\
&-2\int_0^{[t]_\delta} \mathbb{E}\Big[\langle u([s]_\delta), \mathbf{i}u^\delta([s]_\delta)dW^\delta(s)\rangle\Big] + C\delta^{\frac{1}{2}} \\
&\le C(1+C_R)\delta^{\frac{1}{2}} - 2\int_0^{[t]_\delta} \mathbb{E}\Big[\langle \int_{[s]_\delta}^s \mathbf{i}u([r]_\delta))dW(r), \mathbf{i}u^\delta([s]_\delta)dW^\delta(s)\rangle\Big] \\
&-2\int_0^{[t]_\delta} \mathbb{E}\Big[\langle \int_{[s]_\delta}^s (\exp(\mathbf{i}\Delta(r-[s]_\delta))-I)\mathbf{i}u([r]_\delta))dW(r), \mathbf{i}u^\delta([s]_\delta)dW^\delta(s)\rangle\Big] \\
&\le -2\int_0^{[t]_\delta} \mathbb{E}\Big[\langle \int_{[s]_\delta}^s \mathbf{i}u([r]_\delta)dW(r), \mathbf{i}u^\delta([s]_\delta)dW^\delta(s)\rangle\Big] + C(1+C_R)\delta^{\frac{1}{2}}.
\end{aligned}
$$

Similar to III_2, we have that $\mathbb{E}[III_4] \le C(1+C_R)\delta^{\frac{1}{2}}$.

For the terms III_3 and III_5, by taking expectation and using the property $\|e^{\mathbf{i}\Delta t}-I\|_{\mathscr{L}(\mathbb{H}^1,L^2)}\le Ct^{\frac12}$, the continuity estimate of u and the property of martingale, we arrive at

$$\begin{aligned}
\mathbb{E}\Big[III_3\Big] &\le -\int_0^{[t]_\delta} 2\mathbb{E}\Big[\langle u(s)-u([s]_\delta),\mathbf{i}(u^\delta(s)-u^\delta([s]_\delta))dW^\delta(s)\rangle\Big]\\
&\quad -\int_0^{[t]_\delta} 2\mathbb{E}\Big[\langle u([s]_\delta),\mathbf{i}(u^\delta(s)-u^\delta([s]_\delta))dW^\delta(s)\rangle\Big]+C(1+C_R)\delta^{\frac12}.\\
&= -\int_0^{[t]_\delta} 2\mathbb{E}\Big[\langle u([s]_\delta),\mathbf{i}\Big(\int_{[s]_\delta}^s \mathbf{i}u^\delta([r]_\delta)dW^\delta(r)\Big)dW^\delta(s)\rangle\Big]+C(1+C_R)\delta^{\frac12},\\
\mathbb{E}[III_5] &\le -2\mathbb{E}\Big[\int_0^{[t]_\delta}\langle\int_{[s]_\delta}^s \mathbf{i}u^\delta([r]_\delta)dW^\delta(r),\mathbf{i}u([s]_\delta)dW(s)\rangle\Big]+C(1+C_R)\delta^{\frac12}.
\end{aligned}$$

Due to the independent increments of W and the property of conditional expectation, we obtain that

$$\begin{aligned}
&2\int_0^{[t]_\delta}\mathbb{E}\Big[\langle\int_{[s]_\delta}^s \mathbf{i}u([r]_\delta))dW(r),\mathbf{i}u^\delta([s]_\delta)dW^\delta(s)\rangle\Big]\\
&=2\sum_{k=0}^{\frac{[t]_\delta}{\delta}-1}\mathbb{E}\Big[\int_{t_k}^{t_{k+1}}\langle u(t_k)(W(s)-W(t_k)),u^\delta(t_k)(W(t_{k+1})-W(t_k))\rangle\delta^{-1}\Big]ds\\
&=2\sum_{k=0}^{\frac{[t]_\delta}{\delta}-1}\mathbb{E}\Big[\int_{t_k}^{t_{k+1}}\frac{s-t_k}{\delta}\sum_{i=1}^N\langle u(t_k),u^\delta(t_k)|q_i|^2\rangle\Big]ds\\
&=\int_0^{[t]_\delta}\mathbb{E}\Big[\langle u^\delta([s]_\delta),\sum_{k=1}^N|q_k|^2u([s]_\delta)\rangle\Big]ds.
\end{aligned}$$

On the other hand, $\int_{[t]_\delta}^t\mathbb{E}\Big[\langle u^\delta([s]_\delta),\sum_{i=1}^N|q_i|^2u([s]_\delta)\rangle\Big]ds\le C\delta$ due to the mass conservation law and assumption on q_i.

Combining the above estimates, we obtain that

$$\begin{aligned}
&\mathbb{E}\Big[\|u(t)-u^\delta(t)\|^2\Big]\\
&\le\int_0^t 2L_f(R)\mathbb{E}\Big[\|u(s)-u^\delta(s)\|^2\Big]+C(1+C_R)\delta^{\frac12}+\int_0^t\mathbb{E}\Big[\langle u^\delta(s),\sum_{i=1}^N|q_i|^2u(s)\rangle\Big]ds\\
&-2\int_0^{[t]_\delta}\mathbb{E}\Big[\langle\int_{[s]_\delta}^s \mathbf{i}u([r]_\delta))dW(r),\mathbf{i}u^\delta([s]_\delta)dW^\delta(s)\rangle\Big]\\
&\le\int_0^t 2L_f(\|u(0)\|)\mathbb{E}\Big[\|u(s)-u^\delta(s)\|^2\Big]+C(1+C_R)\delta^{\frac12}+\int_0^t\mathbb{E}\Big[\langle u^\delta(s),\sum_{i=1}^N|q_i|^2u(s)\rangle\Big]ds\\
&-\int_0^{[t]_\delta}\mathbb{E}\Big[\langle u^\delta([s]_\delta),\sum_{i=1}^N|q_i|^2u([s]_\delta)\rangle\Big]ds.
\end{aligned}$$

Applying the Gronwall's inequality and the continuity estimate of u and u^δ, we get

$$\mathbb{E}[\|u(t)-u^\delta(t)\|^2]\le C(1+C_R)\exp(2L_f(\|u(0)\|)T)\delta^{\frac12}.$$

It follows that

$$\begin{aligned}&\mathbb{P}(\|u(t)-u^{\delta}(t)\|>\varepsilon)\\&\le \mathbb{P}(\|u^{R_1}(t)-u(t)\|>\frac{\varepsilon}{3})+\mathbb{P}(\|u^{R_1,\delta}(t)-u^{\delta}(t)\|>\frac{\varepsilon}{3})\\&+\mathbb{P}(\|u^{R_1}(t)-u^{R_1,\delta}(t)\|>\frac{\varepsilon}{3},t\le\tau)+\mathbb{P}(\|u^{R_1}(t)-u^{R_1,\delta}(t)\|>\frac{\varepsilon}{3},t>\tau).\end{aligned}$$

Taking limit on $\delta\to 0$, $R,R_1\to\infty$, using the strong convergence estimate and Chebyshev's inequality, we obtain

$$\begin{aligned}&\lim_{\delta\to 0}\mathbb{P}(\|u(t)-u^{\delta}(t)\|>\varepsilon)\\&\le \lim_{\delta\to 0}\frac{9}{\varepsilon^2}C(1+C_R)\exp(2L_f(\|u_0\|)T)\delta^{\frac{1}{2}}\\&+\lim_{R\to\infty}\mathbb{P}(\sup_{s\in[0,t]}\|u(s)\|\ge R)+\lim_{R\to\infty}\mathbb{P}(\sup_{s\in[0,t]}\|u^{\delta}([s]_{\delta})\|\ge R)=0.\end{aligned}$$

Similarly, following the above arguments, we further obtain

$$\lim_{\delta\to 0}\mathbb{E}[\sup_{t\in[0,T]}\|u(t)-u^{\delta}(t)\|^2]=0,$$

which implies that

$$\lim_{\delta\to 0}\mathbb{P}(\sup_{t\in[0,T]}\|u(t)-u^{\delta}(t)\|>\varepsilon)=0.$$

Remark 4.1 *Similar to the stochastic Wasserstein Hamiltonian flow induced by classical Stochastic ODEs, one may expect the particle version of the stochastic nonlinear Schrödinger equation* (4.1)*, that is,*

$$\begin{aligned}&dX_t=v(t,X_t),\\&dv(t,X_t)=-\nabla_{X_t}\frac{1}{2}\frac{\delta}{\delta\rho}I(\rho(t,X_t))+2\lambda\nabla_{X_t}f(\rho(t,X_t))+2\nabla_{X_t}\circ dW(t).\end{aligned}\tag{4.7}$$

But we have not found a rigorous way to prove it. This will be studied in the future.

4.2 NLS equation with random dispersion

The dimensionless NLS equation with random dispersion is given by

$$du=\mathbf{i}\Delta u\frac{1}{\varepsilon}m(\frac{t}{\varepsilon^2})dt+\mathbf{i}\lambda f(|u|^2)udt,\tag{4.8}$$

where m is a real-valued centered stationary random process. Under ergodic assumptions on m, it is expected that the limiting model when $\varepsilon\to 0$ is the following stochastic NLS equation with white noise dispersion

$$du=\sigma_0\mathbf{i}\Delta u\circ dB_t+\mathbf{i}\lambda f(|u|^2)udt,\tag{4.9}$$

where $\sigma_0^2=2\int_0^\infty\mathbb{E}[m(0)m(t)]dt$ (see e.g. [25]). For simplicity, we set $\sigma_0=1$ in (4.9) throughout this subsection.

To see (4.9) as a stochastic Wasserstein Hamiltonian flow, let us use (4.8) instead of Wong–Zakai approximations. Assume that the real valued centered stationary process $m(t)$ is continuous and such that for any $T>0$, $t\mapsto \varepsilon\int_0^{\frac{t}{\varepsilon^2}} m(s)ds$ converges in distribution to a standard real-valued Brownian motion B in $\mathscr{C}([0,T])$ (see e.g. [25]).

First, using Madelung transform $u(t,x)=\sqrt{\rho(t,x)}e^{\mathbf{i}S(t,x)}$ gives

$$
\begin{aligned}
&e^{\frac12\log(\rho)+\mathbf{i}S}(\frac12\frac{\partial_t\rho}{\rho}+\mathbf{i}\partial_t S)\\
&=\mathbf{i}e^{\frac12\log(\rho)+\mathbf{i}S}\Big(\frac12\frac{\nabla\rho}{\rho}+\mathbf{i}\nabla S)^2+(\frac12\frac{\Delta\rho}{\rho}+\mathbf{i}\Delta S-\frac12|\frac{\nabla\rho}{\rho}|^2)\Big)\frac1\varepsilon m(\frac{t}{\varepsilon^2})\\
&+\mathbf{i}e^{\frac12\log(\rho)+\mathbf{i}S}\lambda f(\rho)\\
&=\mathbf{i}e^{\frac12\log(\rho)+\mathbf{i}S}\Big(\frac14(\frac{\nabla\rho}{\rho})^2-(\nabla S)^2+\mathbf{i}\frac{\nabla\rho}{\rho}\cdot\nabla S)+(\frac12\frac{\Delta\rho}{\rho}+\mathbf{i}\Delta S-\frac12|\frac{\nabla\rho}{\rho}|^2)\Big)\frac1\varepsilon m(\frac{t}{\varepsilon^2})\\
&+\mathbf{i}e^{\frac12\log(\rho)+\mathbf{i}S}\lambda f(\rho).
\end{aligned}
$$

We obtain that

$$
\begin{aligned}
\partial_t\rho&=-2\nabla\cdot(\rho\nabla S)\frac1\varepsilon m(\frac{t}{\varepsilon^2}),\\
\partial_t S&=(-|\nabla S|^2-\frac14\frac{\delta}{\delta\rho}I(\rho))\frac1\varepsilon m(\frac{t}{\varepsilon^2})+\lambda f(\rho),
\end{aligned}
\tag{4.10}
$$

which can be rewritten as

$$
\begin{aligned}
\partial_t\rho&=-\nabla\cdot(\rho v)\frac1\varepsilon m(\frac{t}{\varepsilon^2}),\\
\partial_t v&=(-\nabla_x v\cdot v-\nabla_x\frac12\frac{\delta}{\delta\rho}I(\rho))\frac1\varepsilon m(\frac{t}{\varepsilon^2})+2\lambda\nabla_x f(\rho).
\end{aligned}
$$

Based on the above calculations, following the similar steps in the proof of Proposition 4.1, we conclude the following result.

Proposition 4.2 *The critical point of the variational problem*

$$
I_\varepsilon(\rho^0,\rho^T)=\inf\{\mathscr{S}(\rho_t,\Phi_t)|\Delta_{\rho_t}\Phi_t\in\mathscr{T}_{\rho_t}\mathscr{P}_o(\mathscr{M}),\rho(0)=\rho^0,\rho(T)=\rho^T\}
\tag{4.11}
$$

whose action functional is given by the dual coordinates,

$$
\mathscr{S}(\rho_t,\Phi_t)=-\int_0^T\langle\Phi(t),\partial_t\rho_t\rangle dt+\int_0^T\mathscr{H}_0(\rho_t,\Phi_t)dt+\int_0^T\mathscr{H}_1(\rho_t,\Phi_t)\frac1\varepsilon m(\frac{t}{\varepsilon^2})dt,
$$

satisfies (4.10). *Here* $\mathscr{H}_0(\rho_t,\Phi_t)=-\lambda\int_{\mathscr{M}}\int_0^\rho f(s)dsdvol_{\mathscr{M}}$ *with a smooth function* f, $\mathscr{H}_1(\rho_t,\Phi_t)=\int_{\mathscr{M}}|\nabla\Phi_t|^2\rho_t dvol_{\mathscr{M}}$ $+\frac14 I(\rho)$, *where* $I(\rho)=\int_{\mathscr{M}}|\nabla\log(\rho)|^2\rho dvol_{\mathscr{M}}$.

It has been shown in [25] that the limit of (4.10) is the NLS equation with white noise dispersion. Therefore, (4.10) is also a stochastic Wasserstein Hamiltonian flow on density manifold.

Remark 4.2 *The above system is also expected to have a particle version. By applying the push-forward map in section 3 on $\widetilde{\Omega}$, the particle version of* (4.9) *is expected to be*

$$dX_t = v(t,X_t)\circ dB_t$$
$$dv(t,X_t) = -\nabla_{X_t}\frac{1}{2}\frac{\delta}{\delta\rho}I(\rho(t,X_t))\circ dB_t + 2\lambda\nabla_{X_t}f(\rho(t,X_t)).$$

We plan to study the well-poseness of the above mean-field stochastic ODEs in the future.

4.3 Schrödinger Bridge Problem (SBP) with common noise

In this part, we indicate that the critical point of the Schrödinger bridge problem (SBP) with common noise may also be a stochastic Wasserstein Hamiltonian flow. The SBP with common noise is inspired by [9, 58] for the Schrödinger Bridge type problem in stochastic case, where the common noise is added into the classical Schrödinger Bridge type problem [42, 12]. This problem can be formulated as a stochastic control problem on Wasserstein manifold:

$$\min_{\{v_t\}_{t\in[0,T]}}\left[\int_0^T\int_{\mathbb{R}^d}\frac{1}{2}|v_t(x)|^2\rho_t(x,\omega)\,dx\,dt\right] \tag{4.12}$$

$$\text{Subject to: } \frac{\partial\rho_t(x,\omega)}{\partial t}+\nabla\cdot(\rho_t(x,\omega)(v_t+A(x,t)\dot{W}_t(\omega))) = \Delta\rho_t. \tag{4.13}$$

$$\text{and } \rho_0(\cdot,\omega)=\rho_a,\ \rho_T(\cdot,\omega)=\rho_b. \tag{4.14}$$

The continuity equation (4.13) can be viewed as an SDE on the Wasserstein manifold $\mathscr{P}_2(\mathbb{R}^d)$, which reads

$$dX_t = v(t,X_t)dt+\sqrt{2}dB(t)+A(t,X_t)dW(t).$$

Here B is the Brownian motion which corresponding to the diffusion effect in (4.13), and W is another Brownian motion which is independent of B and is called the common noise.

In the following, we consider the Wong–Zakai approximation of (4.12), i.e,

$$\min_{\{v_t\}_{t\in[0,T]}}\left[\int_0^T\int_{\mathbb{R}^d}\frac{1}{2}|v_t(x)|^2\rho_t(x,\omega)\,dx\,dt\right] \tag{4.15}$$

$$\text{Subject to: } \frac{\partial\rho_t(x,\omega)}{\partial t}+\nabla\cdot(\rho_t(x,\omega)(v_t+A(x,t)\dot{\xi}_\delta(t)) = \Delta\rho_t.$$

$$\text{and } \rho_0(\cdot,\omega)=\rho_a,\ \rho_T(\cdot,\omega)=\rho_b,$$

and show that its critical point is a stochastic Wasserstein Hamiltonian flow.

Proposition 4.3 *Assume that W is d-dimensional Brownian motion, ξ^δ is the piecewisely linear Wong–Zakai approximation of W. Let $A(\cdot,t)\in\mathscr{C}_b^1(\mathbb{R}^d)$, $\rho_a,\rho_b\in\mathscr{P}_o(\mathbb{R}^d)$ be smooth. Then the critical point of* (4.15) *satisfies*

$$\partial_t\rho_t = \frac{\delta}{\delta\Phi}\mathscr{H}_0(\rho_t,\Phi_t)+\sum_{i=1}^d\frac{\delta}{\delta\Phi}\mathscr{H}_i(\rho_t,\Phi_t)(\dot{\xi}_\delta)_i(t), \tag{4.16}$$

$$\partial_t\Phi_t = -\frac{\delta}{\delta\rho}\mathscr{H}_0(\rho_t,\Phi_t)-\sum_{i=1}^d\frac{\delta}{\delta\rho}\mathscr{H}_i(\rho_t,\Phi_t)(\dot{\xi}_\delta)_i(t),$$

where $\mathscr{H}_0(\rho,\Phi)=\frac{1}{2}\int_{\mathscr{M}}|\nabla\Phi|^2\rho dvol_{\mathscr{M}}-\frac{1}{8}I(\rho)$, $\mathscr{H}_i(\rho,\Phi)=\int_{\mathscr{M}}\rho A_t^i\partial_{x_i}\Phi dvol_{\mathscr{M}}$*. Here* A_t^i *denotes the i-th column of the matrix* A_t*.*

Proof By using the Lagrangian multiplier method, the critical point satisfies

$$\partial_t\rho_t+\nabla\cdot(\rho(\nabla S_t+A_t\dot{\xi}_\delta(t)))=\frac{1}{2}\Delta\rho_t, \tag{4.17}$$

$$\partial_t S_t+\frac{1}{2}|\nabla S_t|^2+\nabla S_t\cdot A_t\dot{\xi}_\delta(t)=-\frac{1}{2}\Delta S_t. \tag{4.18}$$

Applying the “Hopf-Cole” transform (see e.g. [41]) $\Phi_t=S_t-\frac{1}{2}\log(\rho_t)$, we obtain

$$\partial_t\rho_t+\nabla\cdot(\rho_t\nabla\Phi_t)+\nabla\cdot(\rho_t A_t\dot{\xi}_\delta(t))=0,$$
$$\partial_t\Phi_t+\frac{1}{2}|\nabla\Phi_t|^2+\nabla\Phi\cdot A_t\dot{\xi}_\delta(t)=\frac{1}{8}\frac{\delta}{\delta\rho}I(\rho),$$

which implies (4.16).

The above result also coincides with the generalized variational principle (3.11) with the action functional

$$\mathscr{S}(\rho_t,\Phi_t)=-\int_0^T\langle\Phi(t),\partial_t\rho_t\rangle dt+\int_0^T\mathscr{H}_0(\rho_t,\Phi_t)dt+\sum_{i=1}^d\int_0^T\mathscr{H}_i(\rho_t,\Phi_t)d\xi_\delta(t),$$

whose critical point is the stochastic Hamiltonian system (4.16). From the proof of Proposition 4.3, (4.16) is equivalent to the forward and backward system which contains the backward stochastic Hamilton-Jacobi equation (4.18) and a forward stochastic Kolmogorov equation (4.17), and plays the role of characteristics for the master equation [9]. The derivation of (4.16) may be extended to the mean-field game systems with common noise in [9,11] up to an Itô-Wentzell correction term [39]. If the Wong–Zakai approximation (4.15) is convergent to (4.12), then the critical point of (4.12) is expected to be a stochastic Wasserstein Hamiltonian flow. This will be our future research.

5 Conclusions

In this paper, we study the stochastic Wasserstein Hamiltonian flows, including the stochastic Euler–Lagrange equations and its Hamiltonian flows on density manifold. First, we show that the classical Hamiltonian motions with random perturbations and random initial data induce the stochastic Wasserstein Hamiltonian flows via Wong–Zakai approximation with Lagrangian formalism. Then we propose a generalized variational principle to derive and investigate the generalized stochastic Wasserstein Hamiltonian flows, including the stochastic nonlinear Schrödinger equation, Schrödinger equation with random dispersion and stochastic Schrödinger bridge problem. The study provides rigorous mathematical justification for the principle that the conditional probability density of stochastic Hamiltonian flow in sample space is stochastic Hamiltonian flow on density manifold.

6 Acknowledgements

The authors would like to thank the anonymous referees and the associate editor for their comments and suggestions.

A Appendix

Proof of Lemma 2.2

The local existence of (2.4) and (2.1) is ensured thanks to the local Lipschitz condition of f and σ. To obtain a global solution, a priori bound on $H_0(x,p)$ is needed. Denote the solutions of (2.1) and (2.4) with same initial condition (x_0,p_0) by $(x_t^\delta,p_t^\delta),\delta>0$ and x_t^0,p_t^0, respectively. Applying the chain rule to $H_0(x_t^\delta,p_t^\delta)$ for (2.4) and (2.1), we get that

$$\begin{aligned}
H_0(x_t^\delta,p_t^\delta)&=H_0(x_0,p_0)+\int_0^t\eta\nabla_pH_0(x_s^\delta,p_s^\delta)\cdot\nabla_x\sigma(x_s)\dot{\xi}_\delta(s)ds\\
H_0(x_t,p_t)&=H_0(x_0,p_0)+\int_0^\tau\eta\nabla_pH_0(x_s,p_s)\cdot\nabla_x\sigma(x_s)dB_s\\
&\quad+\frac{1}{2}\int_0^\tau\eta^2\nabla_{pp}H_0(x_s,p_s)\cdot(\nabla_x\sigma(x_s),\nabla\sigma(x_s))ds.
\end{aligned}$$

By applying growth condition (2.3) and taking expectation on the second equation, we derive that

$$\begin{aligned}
H_0(x_t^\delta,p_t^\delta)&\le(H_0(x_0,p_0)+\eta C_1T)\exp(\int_0^t c_1\eta|\dot{\xi}_\delta(s)|ds),\\
\mathbb{E}\Big[H_0(x_t,p_t)\Big]&\le(\mathbb{E}\Big[H_0(x_0,p_0)\Big]+\frac{\eta^2}{2}C_1T)\exp(\int_0^\tau c_1\frac{\eta^2}{2}ds).
\end{aligned}$$

The first inequality leads to $H_0(x_t^\delta,p_t^\delta)<\infty$ since $\dot{\xi}_\delta(s)=\frac{B_{t_{k+1}}-B_{t_k}}{\delta}$, if $s\in[t_k,t_{k+1}]$. Furthermore, taking expectation on the first inequality, applying Fernique's theorem (see, e.g. [27]) for Gaussian variable and independent increments of B_t, we get that

$$\mathbb{E}\Big[H_0(x_t^\delta,p_t^\delta)\Big]\le C(T,\eta,c_1)(2^{[\frac{t}{\delta}]}(\mathbb{E}\Big[H_0(x_0,p_0)\Big]+1),$$

where $[w]$ is the integer part of the real number w. The second inequality yield that $H_0(x_t,p_t)<\infty,a.s$, and the global existence of the strong solution of (2.4). Similarly, for $p\ge2$, we have that

$$\begin{aligned}
\mathbb{E}\Big[H_0^p(x_t^\delta,p_t^\delta)\Big]&\le C(T,\eta,c_1,C_1,p)2^{p[\frac{t}{\delta}]}(\mathbb{E}\Big[H_0^p(x_0,p_0)\Big]+1),\\
\mathbb{E}\Big[H_0^p(x_t,p_t)\Big]&\le C(T,\eta,c_1,p)(\mathbb{E}\Big[H_0^p(x_0,p_0)\Big]+1).
\end{aligned}$$

Furthermore, applying the above bounded moment estimate, we obtain that for $s\le t$,

$$\begin{aligned}
\mathbb{E}\Big[|x(t)-x(s)|^{2p}+|p(t)-p(s)|^{2p}\Big]&\le C(T,\eta,c_1,C_1,c_0,C_1,p,x_0,p_0)|t-s|^p\\
\mathbb{E}\Big[|x^\delta(t)-x^\delta(s)|^{2p}+|p(t)-p(s)|^{2p}\Big]&\le C(T,\eta,c_1,C_1,c_0,C_1,p,x_0,p_0)2^{[\frac{t}{\delta}]}|t-s|^p.
\end{aligned}$$

However, the above estimate of x^δ is too rough and exponentially depending on $\frac{1}{\delta}$. As a consequence, we can not expect any convergence result. A delicate estimate of (x^δ,p^δ) is needed.

Assume that $t \in [t_k, t_{k+1}]$, $t_k = k\delta$. Then by using the expansion of (2.1), we have that

$$
\begin{aligned}
H_0(x_t^\delta, p_t^\delta) &= H_0(x_0,p_0) - \sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} \eta \nabla_p H_0(x_s^\delta, p_s^\delta)\cdot \nabla_x \sigma(x_s^\delta) d\xi_\delta(s) \\
&- \int_{t_k}^{t} \eta \nabla_p H_0(x_s^\delta, p_s^\delta)\cdot \nabla_x \sigma(x_s^\delta) d\xi_\delta(s) \\
&= H_0(x_0,p_0) - \sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} \eta \nabla_p H_0(x_{t_j}^\delta, p_{t_j}^\delta)\cdot \nabla_x \sigma(x_{t_j}^\delta) d\xi_\delta(s) \\
&- \int_{t_k}^{t} \eta \nabla_p H_0(x_{t_k}^\delta, p_{t_k}^\delta)\cdot \nabla_x \sigma(x_{t_k}^\delta) d\xi_\delta(s) \\
&- \sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} \eta\Big(\int_{t_j}^{s} \nabla_{pp} H_0(x_r^\delta, p_r^\delta)\cdot(\nabla_x\sigma(x_r^\delta), -\eta\nabla_x\sigma(x_r^\delta)\dot\xi_\delta(r)) dr \dot\xi_\delta(s) \\
&\quad + \int_{t_j}^{s} \nabla_{pp} H_0(x_r^\delta, p_r^\delta)\cdot(\nabla_x\sigma(x_r^\delta), -\frac{1}{2}(p_r^\delta)^\top d_x g^{-1}(x) p_r^\delta - \nabla_x f(x_s^\delta)) dr \dot\xi_\delta(s) \\
&\quad + \int_{t_j}^{s} \nabla_p H_0(x_r^\delta, p_r^\delta)\cdot \nabla_{xx}\sigma(x_r^\delta) g^{-1}(x_r^\delta) p_r^\delta dr \dot\xi_\delta(s) \\
&\quad + \int_{t_j}^{s} \nabla_{px} H_0(x_r^\delta, p^\delta)\cdot(\nabla_x\sigma(x_r^\delta)\dot\xi_\delta(s), g^{-1}(x_r^\delta) p_r^\delta) dr\Big) ds \\
&- \int_{t_k}^{t} \eta\Big(\int_{t_k}^{s} \nabla_{pp} H_0(x_r^\delta, p_r^\delta)\cdot(\nabla_x\sigma(x_r^\delta), -\eta\nabla_x\sigma(x_r^\delta)\dot\xi_\delta(r)) dr \dot\xi_\delta(s) \\
&\quad + \int_{t_k}^{s} \nabla_{pp} H_0(x_r^\delta, p_r^\delta)\cdot(\nabla_x\sigma(x_r^\delta), -\frac{1}{2}(p_r^\delta)^\top d_x g^{-1}(x_r^\delta) p_r^\delta - \nabla_x f(x_s^\delta)) dr \dot\xi_\delta(s) \\
&\quad + \int_{t_k}^{s} \nabla_p H_0(x_r^\delta, p_r^\delta)\cdot \nabla_{xx}\sigma(x_r^\delta) g^{-1}(x^{\delta_r}) p_r^\delta dr \dot\xi_\delta(s) \\
&\quad + \int_{t_k}^{s} \nabla_{px} H_0(x_r^\delta, p^\delta)\cdot(\nabla_x\sigma(x_r^\delta)\dot\xi_\delta(s)), g^{-1}(x_r^\delta) p_r^\delta) dr\Big) ds \\
&=: H_0(x_0,p_0) + \sum_{j=0}^{k-1} I_j^1 + I_k^1(t) \\
&+ \sum_{j=0}^{k-1}(I_j^{21} + I_j^{22} + I_j^{23} + I_j^{24}) + I_k^{21}(t) + I_k^{22}(t) + I_k^{23}(t) + I_k^{24}(t).
\end{aligned}
$$

Making use of the growth condition (2.3), we have that

$$
\begin{aligned}
&\sum_{j=0}^{k-1}(I_j^{21} + I_j^{22} + I_j^{23} + I_j^{24}) + I_k^{21}(t) + I_k^{22}(t) + I_k^{23}(t) + I_k^{24}(t) \\
&\le \sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} (C_1 + c_1 H_0(x_s^\delta, p_s^\delta))|\dot\xi_\delta(s)|^2 \delta ds + \sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} (C_1 + c_1 H_0(x_s^\delta, p_s^\delta))|\dot\xi_\delta(s)|\delta ds \\
&+ \int_{t_k}^{t} (C_1 + c_1 H_0(x_s^\delta, p_s^\delta))|\dot\xi_\delta(s)|^2 \delta ds + \int_{t_k}^{t} (C_1 + c_1 H_0(x_s^\delta, p_s^\delta))|\dot\xi_\delta(s)|\delta ds \\
&= \int_0^t (C_1 + c_1 H_0(x_s^\delta, p_s^\delta))|\dot\xi_\delta(s)|^2 \delta ds + \int_0^t (C_1 + c_1 H_0(x_s^\delta, p_s^\delta))|\dot\xi_\delta(s)|\delta ds.
\end{aligned}
$$

By using the Gronwall–Bellman inequality, we obtain that

$$
H_0(x_t^\delta, p_t^\delta) \le \exp\Big(\int_0^t c_1(|\dot\xi_\delta(s)|^2 + |\dot\xi_\delta(s)|)\delta ds\Big)\Big(H_0(x_0,p_0) + CT + \Big|\sum_{j=0}^{k-1} I_j^1 + I_k^1(t)\Big|\Big).
$$

For simplicity, assume that $T = K\delta$. Denote $[t]_\delta = t_k = k\delta$ if $t \in [t_k, t_{k+1})$. The definition of $\xi_\delta(s)$ yields that $s \in [t_j, t_{j+1}]$

$$
|\dot\xi_\delta(s)|^2\delta + |\dot\xi_\delta(s)|\delta = \Big|\frac{B(t_{j+1}) - B(t_j)}{\delta}\Big|^2\delta + |B(t_{j+1}) - B(t_j)|.
$$

Define a stopping time $\tau_R = \inf\{t \in [0,T] | \int_0^{[t]\delta} |\dot{\xi}_\delta|^2 \delta ds \ge R\}$. The stopping time is well-defined since the quadratic variation process of Brownian motion is bounded in $[0,T]$. Then taking $t \le \tau_R$ and using Hölder's inequality, then it holds that

$$\begin{aligned} H_0(x_t^\delta, p_t^\delta) &\le \exp(\int_{[t]}^t c_1(|\dot{\xi}_\delta(s)|^2 + |\dot{\xi}_\delta|ds) \exp(C(R+T))(H_0(x_0,p_0) + CT + |\sum_{j=0}^{k-1} I_j^1 + I_k^1(t)|) \quad (A.1)\\ &\le \exp(\int_{[t]}^t c_1(\frac{3}{2}|\dot{\xi}_\delta(s)|^2)ds) \exp(C(R+T))H_0(x_0,p_0)\\ &+ \exp(C(R+T))\exp(\int_{[t]}^t c_1 \frac{3}{2}|\dot{\xi}_\delta(s)|^2 ds) \left| \int_0^{[t]} -\eta \nabla_p H_0(x_{[s]_\delta}^\delta, p_{[s]_\delta}^\delta) \cdot \nabla_x \sigma(x_{[s]_\delta}) dB(s) \right|\\ &+ \exp(C(R+T))\exp(\int_{[t]}^t (c_1 \frac{3}{2}|\dot{\xi}_\delta(s)|^2 ds) \left| \int_{[t]}^t -\eta \nabla_p H_0(x_{[s]_\delta}^\delta, p_{[s]_\delta}^\delta) \cdot \nabla_x \sigma(x_{[s]_\delta}) \dot{\xi}_\delta(s) ds \right|. \end{aligned}$$

Similarly, one could obtain a analogous estimate of (A.1) with the integral over $[t_{k-1}, t_k]$, where t_k, $k \le K$, $t_K \le \tau_R$. By the Cauchy inequality and taking expectation on both sides of (A.1), applying the Burkholder–Davis–Gundy inequality (see e.g, [35]) and using the independent increments of Brownian motion, we get

$$\begin{aligned} &\mathbb{E}[H_0^2(x_{t_k}^\delta, p_{t_k}^\delta)]\\ &\le 3\mathbb{E}\Big[\exp(\int_{t_{k-1}}^{t_k} (3c_1|\dot{\xi}_\delta(s)|^2 ds)\Big] \exp(2C(R+T))\mathbb{E}\Big[H_0^2(x_0,p_0)\Big]\\ &+ 3\exp(2C(R+T))\mathbb{E}[\exp(\int_{t_{k-1}}^{t_k} 3c_1|\dot{\xi}_\delta(s)|^2 ds)]\mathbb{E}\Big[\Big|\int_0^{t_{k-1}} -\eta \nabla_p H_0(x_{[s]_\delta}^\delta, p_{[s]_\delta}^\delta) \cdot \nabla_x \sigma(x_{[s]_\delta}) dB(s)\Big|^2\Big]\\ &+ 3\exp(2C(R+T))\mathbb{E}\Big[\exp(\int_{t_{k-1}}^{t_k} 3c_1|\dot{\xi}_\delta(s)|^2 ds)|B(t_k) - B(t_{k-1})|^2\\ &\times \big|\eta \nabla_p H_0(x_{t_{k-1}}^\delta, p_{t_{k-1}}^\delta) \cdot \nabla_x \sigma(x_{t_{k-1}})\big|^2\Big]\\ &\le 3\mathbb{E}\Big[\exp(\int_{t_{k-1}}^{t_k} (3c_1|\dot{\xi}_\delta(s)|^2 ds)\Big] \exp(2C(R+T))\mathbb{E}\Big[H_0^2(x_0,p_0)\Big]\\ &+ 3\exp(2C(R+T))\mathbb{E}[\exp(\int_{t_{k-1}}^{t_k} 3c_1|\dot{\xi}_\delta(s)|^2 ds)]\mathbb{E}\Big[\int_0^{t_{k-1}} (C_1 + c_1 H_0(x_{[s]_\delta}^\delta, p_{[s]_\delta}^\delta))^2 ds\Big]\\ &+ 3\exp(2C(R+T))\mathbb{E}\Big[\exp(\int_{t_{k-1}}^{t_k} 3c_1|\dot{\xi}_\delta(s)|^2 ds)|B(t_k) - B(t_{k-1})|^2\Big]\\ &\times \mathbb{E}\Big[(C_1 + c_1 H_0^2(x_{t_{k-1}}^\delta, p_{t_{k-1}}^\delta))\Big]. \end{aligned}$$

Applying the Fernique theorem and choosing sufficient small δ such that $12c_1\delta < 1$, then we have that

$$\begin{aligned} &\mathbb{E}[\exp(\int_{t_{k-1}}^{t_k} 3c_1|\dot{\xi}_\delta(s)|^2 ds)] \le C,\\ &\mathbb{E}\Big[\exp(\int_{t_{k-1}}^{t_k} 3c_1|\dot{\xi}_\delta(s)|^2 ds)|B(t_k) - B(t_{k-1})|^2\Big]\\ &\le \sqrt{\mathbb{E}[\exp(\int_{t_{k-1}}^{t_k} 6c_1|\dot{\xi}_\delta(s)|^2 ds)]}\sqrt{\mathbb{E}\Big[|B(t_k) - B(t_{k-1})|^4\Big]} \le C\delta. \end{aligned}$$

The above estimation gives

$$\begin{aligned} \mathbb{E}[H_0^2(x_{t_k}^\delta, p_{t_k}^\delta)] &\le 3\exp(2C(R+T))C\mathbb{E}[H_0^2(x_0,p_0)]\\ &+ 6\exp(2C(R+T))C\int_0^{t_{k-1}} \mathbb{E}\Big[(C_1^2 + c_1^2 H_0^2(x_{[s]_\delta}^\delta, p_{[s]_\delta}^\delta))\Big]ds\\ &+ 6\exp(2C(R+T)C\delta\mathbb{E}\Big[C_1^2 + c_1^2 H_0^2(x_{t_{k-1}}^\delta, p_{t_{k-1}}^\delta)\Big]. \end{aligned}$$

Then the Grownall's inequality yield that

$$\begin{aligned}\mathbb{E}[H_0^2(x_{t_k}^\delta, p_{t_k}^\delta)] &\leq \exp(6TCc_1^2 \exp(2C(R+T))) \\ &\quad \times \Big(3\exp(2C(R+T))C\mathbb{E}[H_0^2(x_0,p_0)] + 6C_1^2 TC\exp(2C(R+T))\Big)\end{aligned}$$

Combining the above estimates with (A.1) and the Burkholder–Davis–Gundy inequality, we conclude that

$$\begin{aligned}\sup_{t\in[0,\tau^R)} \mathbb{E}[H_0^2(x_t^\delta, p_t^\delta)] &\leq (\exp(6TCc_1^2\exp(2C(R+T))) + C) \\ &\quad \times \Big(3\exp(2C(R+T))C\mathbb{E}[H_0^2(x_0,p_0)] + 6C_1^2 TC\exp(2C(R+T))\Big) \\ &=: C_R.\end{aligned}$$

Similarly, by choosing sufficient small δ, we have that for $t \in [0,\tau^R)$,

$$\mathbb{E}[H_0^p(x_t^\delta, p_t^\delta)] \leq C_{R,p} < \infty.$$

As a consequence, by again using (A.1), we obtain that

$$\mathbb{E}\Big[\sup_{t\in[0,\tau^R)} H_0^p(x_t^\delta, p_t^\delta)\Big] \leq C_{R,p} < \infty.$$

Next we show the convergence in probability of the solution of (2.1) to that of (2.4). Introduce another stopping time $\tau_{R_1} := \inf\{t \in [0,T] | |x_t| + |p_t| \geq R_1, |x^\delta_{[t]_\delta}| + |p^\delta_{[t]_\delta}| \geq R_1\}$. Let $t \in [0, \tau_R \wedge \tau_{R_1})$. By using the polynomial growth condition of f, σ and the fact that σ is independent of p, we obtain that

$$\begin{aligned}&|x^\delta(t) - x(t)|^2 \\ &= |x^\delta(0) - x(0)|^2 + \int_0^t 2\langle x^\delta(s) - x(s), g^{-1}(x^\delta(s))p^\delta(s) - g^{-1}(x(s))p(s)\rangle ds \\ &\leq |x^\delta(0) - x(0)|^2 + \int_0^t C_g(1+|p(s)|)(|x^\delta(s)-x(s)|^2 + |p^\delta(s)-p(s)|^2)ds, \\ &|p^\delta(t) - p(t)|^2 \\ &= \int_0^t \langle -(p^\delta(s))^\top d_x g^{-1}(x^\delta(s))p^\delta(s) + p(s)^\top d_x g^{-1}(x(s))p(s), p^\delta(s) - p(s)\rangle ds \\ &+ \int_0^t 2\langle -\nabla_x f(x^\delta(s)) + \nabla_x f(x(s)), p^\delta(s) - p(s)\rangle ds \\ &- \int_0^t 2\eta\langle p^\delta(s) - p(s), \nabla_x\sigma(x^\delta(s))d\xi_\delta(s) - \nabla_x\sigma(x(s))dB_t\rangle \\ &\leq C_g\int_0^t (1+|x^\delta(s)|)(|p^\delta(s)|^2 + |p(s)|^2)(|p^\delta(s)-p(s)|^2 + |x^\delta(s)-x(s)|^2)ds \\ &+ C_f\int_0^t (1+|x(s)|^{p_f} + |x^\delta|^{p_f})(|p^\delta(s)-p(s)|^2 + |x^\delta(s)-x(s)|^2)ds \\ &- \int_0^t 2\eta\langle p^\delta(s) - p(s), \nabla_x\sigma(x^\delta(s))d\xi_\delta(s) - \nabla_x\sigma(x(s))dB_s\rangle,\end{aligned}$$

where C_g and C_f are constants depending on f and g. To deal with the last term, we split it as follows,

$$
\begin{aligned}
&\int_0^t 2\eta\langle p^\delta(s)-p(s),\nabla_x\sigma(x^\delta(s))d\xi_\delta(s)-\nabla_x\sigma(x(s))dB_s\rangle\\
&=2\eta\int_0^t\langle p^\delta([s]_\delta)-p([s]_\delta),\nabla_x\sigma(x^\delta(s))d\xi_\delta(s)-\nabla_x\sigma(x(s))dB_s\rangle\\
&+2\eta\int_0^t\langle p^\delta(s)-p(s)-p^\delta([s]_\delta)+p([s]_\delta),\nabla_x\sigma(x^\delta(s))d\xi_\delta(s)-\nabla_x\sigma(x(s))dB_s\rangle\\
&=2\eta\int_0^t\langle p^\delta([s]_\delta)-p([s]_\delta),\nabla_x\sigma(x^\delta([s]_\delta))d\xi_\delta([s]_\delta)-\nabla_x\sigma(x([s]_\delta))dB_s\rangle\\
&+2\eta\int_0^t\langle p^\delta([s]_\delta)-p([s]_\delta),(\nabla_x\sigma(x^\delta(s))-\nabla_x\sigma(x^\delta([s]_\delta)))d\xi_\delta(s)-(\nabla_x\sigma(x(s))-\nabla_x\sigma(x([s]_\delta)))dB_s\rangle\\
&+2\eta\int_0^t\langle p^\delta(s)-p(s)-p^\delta([s]_\delta)+p([s]_\delta),\nabla_x\sigma(x^\delta([s]_\delta))d\xi_\delta(s)-\nabla_x\sigma(x([s]_\delta)dB_s\rangle\\
&+2\eta\int_0^t\langle p^\delta(s)-p(s)-p^\delta([s]_\delta)+p([s]_\delta),(\nabla_x\sigma(x^\delta(s)-\nabla_x\sigma(x^\delta([s]_\delta)))d\xi_\delta(s)\\
&\quad-(\nabla_x\sigma(x(s)-\nabla_x\sigma(x([s]_\delta))dB_s\rangle\\
&=: II^1+II^2+II^3+II^4.
\end{aligned}
$$

Taking expectation on II^1, using the property of the discrete martingale, the a prior estimates for $H_0(x_t,p_t)$ and $H_0(x_t^\delta,p_t^\delta)$ and Hölder's inequality, we have that

$$
\begin{aligned}
&\mathbb{E}[II^1]=0,\\
&\mathbb{E}[II^2]\le 2\eta\int_0^t\mathbb{E}\Big[\langle p^\delta([s]_\delta)-p([s]_\delta),\int_{[s]_\delta}^s(\nabla_{xx}\sigma(x^\delta(r))\cdot(g^{-1}(x^\delta(r))p^\delta(r))drd\xi_\delta(s)\rangle\Big]\\
&\quad-2\eta\int_0^t\mathbb{E}\Big[\langle p^\delta([s]_\delta)-p([s]_\delta),\int_{[s]_\delta}^s(\nabla_{xx}\sigma(x(r))\cdot(g^{-1}(x^\delta(r)p^\delta(r))drdB_s\rangle\Big]\\
&\quad\le C(R_1)\delta^{\frac12}.
\end{aligned}
$$

Similar arguments lead to $\mathbb{E}[II^4]\le C(R_1)\delta^{\frac12}$. For the term II^3, applying the continuity estimate of x_t and x_t^δ, as well as independent increments of the Brownian motion, we get

$$
\begin{aligned}
&\mathbb{E}[II^3]\\
&\le C(R_1)\delta^{\frac12}+2\eta^2\mathbb{E}\Big[\int_0^{[t]_\delta}\langle\int_{[s]_\delta}^s\nabla_x\sigma(x^\delta_{[r]_\delta})d\xi_\delta(r)-\int_{[s]_\delta}^s\nabla_x\sigma(x_{[r]_\delta})dB_r,\\
&\quad\nabla_x\sigma(x^\delta([s]_\delta))d\xi_\delta(s)-\nabla_x\sigma(x([s]_\delta)dB_s\rangle\Big]\\
&\le C(R_1)\delta^{\frac12}+2\eta^2\mathbb{E}\Big[\int_0^{[t]_\delta}|\nabla_x\sigma(x^\delta_{[s]_\delta})|^2\frac{s-[s]_\delta}{\delta^2}(B([s]_\delta+\delta)-B([s]_\delta))^2ds\Big]\\
&-2\eta^2\mathbb{E}\Big[\int_0^{[t]_\delta}\langle\nabla_x\sigma(x^\delta_{[s]_\delta}),\nabla_x\sigma(x_{[s]_\delta})\rangle\frac{s-[s]_\delta}{\delta^2}(B([s]_\delta+\delta)-B([s]_\delta))^2ds\Big]\\
&-2\eta^2\mathbb{E}\Big[\int_0^{[t]_\delta}\langle\nabla_x\sigma(x^\delta_{[s]_\delta}),\nabla_x\sigma(x_{[s]_\delta})\rangle\frac{B([s]_\delta+\delta)-B([s]_\delta)}{\delta}(B(s)-B([s]_\delta))ds\Big]\\
&+2\eta^2\mathbb{E}\Big[\int_0^{[t]_\delta}\langle\nabla_x\sigma(x_{[s]_\delta}),\nabla_x\sigma(x_{[s]_\delta})\rangle\frac{B([s]_\delta+\delta)-B([s]_\delta)}{\delta}(B(s)-B([s]_\delta))ds\Big]\\
&\le C(R_1)\delta^{\frac12}+2\eta^2\int_0^{[t]_\delta}\mathbb{E}\Big[|\nabla_x\sigma(x^\delta_{[s]_\delta})-\nabla_x\sigma(x_{[s]_\delta})|^2\Big]ds\\
&\le C(R_1)\delta^{\frac12}+\int_0^t C(R_1)\mathbb{E}\Big[|x_s^\delta-x_s|^2\Big]ds,
\end{aligned}
$$

where $C(R_1) > 0$ is monotone with R_1. Combining the above estimates, we achieve that

$$\mathbb{E}[|x^\delta(t)-x(t)|^2] \le \int_0^t C_g(1+C_{R_1})(\mathbb{E}[|x^\delta(s)-x(s)|^2]+\mathbb{E}[|p^\delta(s)-p(s)|^2])ds$$
$$\mathbb{E}[|p^\delta(t)-p(t)|^2] \le \int_0^t (C_g+C_f)(1+C_{R_1})(\mathbb{E}[|x^\delta(s)-x(s)|^2]+\mathbb{E}[|p^\delta(s)-p(s)|^2])ds+C(R_1)\delta^{\frac{1}{2}}.$$

Then the Gronwall's inequality implies that

$$\mathbb{E}[|x^\delta(t)-x(t)|^2]+\mathbb{E}[|p^\delta(t)-p(t)|^2] \le \exp(2(C_g+C_f)(1+C_{R_1})T)C(R_1)\delta^{\frac{1}{2}}. \tag{A.2}$$

By making use of (A.2) and Chebshev's inequality, we conclude that

$$\begin{aligned}
&\mathbb{P}(|x^\delta(t)-x(t)|+|p^\delta(t)-x(t)|\ge\varepsilon)\\
&\le \mathbb{P}(\{|x^\delta(t)-x(t)|+|p^\delta(t)-x(t)|\ge\varepsilon\}\cap\{t<\tau_R\}\cap\{t<\tau_{R_1}\})\\
&+\mathbb{P}(\{|x^\delta(t)-x(t)|+|p^\delta(t)-x(t)|\ge\varepsilon\}\cap\{t\ge\tau_R\})\\
&+\mathbb{P}(\{|x^\delta(t)-x(t)|+|p^\delta(t)-x(t)|\ge\varepsilon\}\cap\{t<\tau_R\}\cap\{t\ge\tau_{R_1}\})\\
&\le 2\frac{\mathbb{E}\Big[|x^\delta(t)-x(t)|^2+|p^\delta(t)-x(t)|^2\Big]}{\varepsilon^2}\\
&+\frac{\mathbb{E}\Big[\int_0^t|\dot\xi_\delta(s)|^2\delta ds\Big]}{R}+\frac{\mathbb{E}\Big[|x(t)|+|p(t)|+|x^\delta(t)|+|p^\delta(t)|\Big]}{R_1}\\
&\le \frac{2}{\varepsilon^2}\exp(2(C_g+C_f)(1+C_{R_1})T)C(R_1)\delta^{\frac{1}{2}}+\frac{C}{R}+C\frac{1+C_R}{R_1}.
\end{aligned}$$

Here, $\mathbb{E}[|x(t)|+|p(t)|+|x^\delta(t)|+|p^\delta(t)|] < C(1+C_R)$ is ensured by $\mathbb{E}[\sup_{t\in[0,\tau^R)} H_0^2(x_t^\delta,p_t^\delta)] \le C_R$. Taking limit on $\delta\to0$, $R_1\to\infty$, and $R\to\infty$ leads to

$$\lim_{\delta\to0}\mathbb{P}(|x^\delta(t)-x(t)|+|p^\delta(t)-p(t)|>\varepsilon)=0.$$

Similarly, one could utilize the properties of martingale and obtain the following estimate, for large enough $q>0$,

$$\mathbb{E}[|x^\delta(t)-x(t)|^q]+\mathbb{E}[|p^\delta(t)-p(t)|^q] \le C_q\exp(C_q(C_g+C_f)(1+C_{R_1})T)C(R_1)\delta^{\frac{q}{2}-1}.$$

This implies that for large enough $q>4$,

$$\begin{aligned}
&\mathbb{E}[\sup_{k\le K}\sup_{t\in[t_{k-1},t_k]}|x^\delta(t)-x(t)|^q]+\mathbb{E}[\sup_{k\le K}\sup_{t\in[t_{k-1},t_k]}|p^\delta(t)-p(t)|^q]\\
&\le \sum_{k=0}^{K-1}\mathbb{E}[\sup_{t\in[t_{k-1},t_k]}|x^\delta(t)-x(t)|^q]+\mathbb{E}[\sup_{t\in[t_{k-1},t_k]}|p^\delta(t)-p(t)|^q]\\
&\le C_qK\exp(C_q(C_g+C_f)(1+C_{R_1})T)C(R_1)\delta^{\frac{q}{2}-1}\\
&\le C_q\exp(C_q(C_g+C_f)(1+C_{R_1})T)C(R_1)\delta^{\frac{q}{2}-2}.
\end{aligned}$$

Combining the above estimate and applying the Chebshev's inequality, we further obtain

$$\lim_{\delta\to0}\mathbb{P}(\sup_{t\in[0,T]}|x^\delta(t)-x(t)|+\sup_{t\in[0,T]}|p^\delta(t)-p(t)|>\varepsilon)=0.$$

□

2 Declarations

Ethical Approval: this is not applicable.
Authors' contributions: Authors are listed in alphabetical order of the surnames and have equal contributions.
Competing interests: this work has no conflict of interest.
Funding: the research is partially supported by Georgia Tech Mathematics Application Portal (GT-MAP) and by research grants NSF DMS-1830225, and ONR N00014-21-1-2891, the start-up funds P0039016 and internal grants (P0041274,P0045336) from Hong Kong Polytechnic University, the CAS AMSS-PolyU Joint Laboratory of Applied Mathematics and the grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (PolyU25302822 for ECS project).
Availability of data and materials: this manuscript has no associated data.

References

1. G. P. Agrawal. Applications of Nonlinear Fiber Optics. Academic Press, San Diego, 2001.
2. G. P. Agrawal. Nonlinear Fiber Optics, 3rd ed. Academic Press, San Diego, 2001.
3. S. Albeverio, K. Yasue, and J.-C. Zambrini. Euclidean quantum mechanics: analytical approach. Ann. Inst. H. Poincaré Phys. Théor., 50(3):259–308, 1989.
4. L. Ambrosio and W. Gangbo. Hamiltonian ODEs in the Wasserstein space of probability measures. Comm. Pure Appl. Math., 61(1):18–53, 2008.
5. O. Bang, P. L. Christiansen, F. If, K. Ø. Rasmussen, and Y. B. Gaididei. Temperature effects in a nonlinear model of monolayer scheibe aggregates. Phys. Rev. E, 49:4627–4636, May 1994.
6. S. Bernstein. Sur les liaisons entre les grandeurs aléatoires. Verh. des intern. Math., 1932.
7. J.-M. Bismut. Mécanique aléatoire, volume 866 of Lecture Notes in Mathematics. Springer-Verlag, Berlin-New York, 1981. With an English summary.
8. Z. Brzeźniak and F. Flandoli. Almost sure approximation of Wong-Zakai type for stochastic partial differential equations. Stochastic Process. Appl., 55(2):329–358, 1995.
9. P. Cardaliaguet, F. Delarue, J. M. Lasry, and P. L. Lions. The master equation and the convergence problem in mean field games, volume 201 of Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, 2019.
10. R. Carmona and F. Delarue. Probabilistic theory of mean field games with applications. II, volume 84 of Probability Theory and Stochastic Modelling. Springer, Cham, 2018. Mean field games with common noise and master equations.
11. R. Carmona, F. Delarue, and D. Lacker. Mean field games with common noise. Ann. Probab., 44(6):3740–3803, 2016.
12. Y. Chen, T. T. Georgiou, and M. Pavon. On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint. J. Optim. Theory Appl., 169(2):671–691, 2016.
13. S. Chow, W. Huang, Y. Li, and H. Zhou. Fokker-Planck equations for a free energy functional or Markov process on a graph. Arch. Ration. Mech. Anal., 203(3):969–1008, 2012.
14. S. Chow, W. Li, and H. Zhou. A discrete Schrödinger equation via optimal transport on graphs. J. Funct. Anal., 276(8):2440–2469, 2019.
15. S. Chow, W. Li, and H. Zhou. Wasserstein Hamiltonian flows. J. Differential Equations, 268(3):1205–1219, 2020.
16. K. L. Chung and J.-C. Zambrini. Introduction to random time and quantum randomness, volume 1 of Monographs of the Portuguese Mathematical Society. World Scientific Publishing Co., Inc., River Edge, NJ, new edition, 2003.
17. C. Conforti and M. Pavon. Extremal flows on Wasserstein space. arXiv:1712.02257, 2017.
18. J. Cui, L. Dieci, and H. Zhou. Time discretizations of Wasserstein-Hamiltonian flows. Math. Comp., 91(335):1019–1075, 2022.
19. J. Cui, J. Hong, and Z. Liu. Strong convergence rate of finite difference approximations for stochastic cubic Schrödinger equations. J. Differential Equations, 263(7):3687–3713, 2017.
20. J. Cui, S. Liu, and H. Zhou. What is a stochastic Hamiltonian process on finite graph? An optimal transport answer. J. Differential Equations, 305:428–457, 2021.
21. J. Cui, S. Liu, and H. Zhou. Optimal control for stochastic nonlinear schrodinger equation on graph. arXiv:2209.05346, to appear in SIAM. J. Control Optim., 2022.
22. J. Cui, S. Liu, and H. Zhou. Wasserstein hamiltonian flow with common noise on graph. arXiv:2204.01185, to appear in SIAM. J. Appl. Math., 2022.
23. J. Cui and L. Sun. Stochastic logarithmic schrodinger equations: energy regularized approach. arXiv:2102.12607, to appear in SIAM. J. Math. Anal., 2022.

24. A. de Bouard and A. Debussche. A stochastic nonlinear Schrödinger equation with multiplicative noise. Comm. Math. Phys., 205(1):161–181, 1999.
25. A. de Bouard and A. Debussche. The nonlinear Schrödinger equation with white noise dispersion. J. Funct. Anal., 259(5):1300–1321, 2010.
26. G. E. Falkovich, I. Kolokolov, V. Lebedev, and S. K. Turitsyn. Statistics of soliton-bearing systems with additive noise. Phys. Rev. E, 63:025601, Jan 2001.
27. X. Fernique. Intégrabilité des vecteurs gaussiens. C. R. Acad. Sci. Paris Sér. A-B, 270:A1698–A1699, 1970.
28. M. Furi. Second order differential equations on manifolds and forced oscillations. In Topological methods in differential equations and inclusions (Montreal, PQ, 1994), volume 472 of NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., pages 89–127. Kluwer Acad. Publ., Dordrecht, 1995.
29. W. Gangbo, H. Kim, and T. Pacini. Differential forms on Wasserstein space and infinite-dimensional Hamiltonian systems. Mem. Amer. Math. Soc., 211(993):vi+77, 2011.
30. D. A. Gomes and J. Saúde. Mean field games models—a brief survey. Dyn. Games Appl., 4(2):110–154, 2014.
31. E. Hille and R. S. Phillips. Functional analysis and semi-groups. American Mathematical Society Colloquium Publications, Vol. XXXI. American Mathematical Society, Providence, R. I., 1974. Third printing of the revised edition of 1957.
32. E. P. Hsu. Stochastic analysis on manifolds, volume 38 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2002.
33. Q. Huang and J.-C. Zambrini. From second-order differential geometry to stochastic geometric mechanics. arXiv:2201.03706, 2022.
34. N. Ikeda and S. Watanabe. Stochastic differential equations and diffusion processes, volume 24 of North-Holland Mathematical Library. North-Holland Publishing Co., Amsterdam-New York; Kodansha, Ltd., Tokyo, 1981.
35. I. Karatzas and S. E. Shreve. Brownian motion and stochastic calculus, volume 113 of Graduate Texts in Mathematics. Springer-Verlag, New York, second edition, 1991.
36. G. Khesin, B.and Misioł ek and K. Modin. Geometric hydrodynamics and infinite-dimensional Newton's equations. Bull. Amer. Math. Soc. (N.S.), 58(3):377–442, 2021.
37. P. E. Kloeden and E. Platen. Numerical solution of stochastic differential equations, volume 23 of Applications of Mathematics (New York). Springer-Verlag, Berlin, 1992.
38. V. V. Konotop and L. Vázquez. Nonlinear random waves. World Scientific Publishing Co., Inc., River Edge, NJ, 1994.
39. N. V. Krylov. On the Itô-Wentzell formula for distribution-valued processes and related topics. Probab. Theory Related Fields, 150(1-2):295–319, 2011.
40. J. D. Lafferty. The density manifold and configuration space quantization. Trans. Amer. Math. Soc., 305(2):699–741, 1988.
41. F. Léger and W. Li. Hopf-Cole transformation via generalized Schrödinger bridge problem. J. Differential Equations, 274:788–827, 2021.
42. C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. Discrete Contin. Dyn. Syst., 34(4):1533–1574, 2014.
43. J. Lott. Some geometric calculations on Wasserstein space. Comm. Math. Phys., 277(2):423–437, 2008.
44. E. Madelung. Quanten theorie in hydrodynamischer form. Zeitschrift für Physik, 40(3-4):322–326, 1927. cited By 1026.
45. T. Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. Probab. Theory Related Fields, 129(2):245–260, 2004.
46. T. Mikami. Stochastic optimal transportation—stochastic control with fixed marginals. SpringerBriefs in Mathematics. Springer, Singapore, [2021] ©2021.
47. E. Nelson. Derivation of the Schrödinger equation from Newtonian mechanics. Physical Review, 150(4):1079–1085, 1966.
48. E. Nelson. The free Markoff field. J. Functional Analysis, 12:211–227, 1973.
49. E. Nelson. Quantum fluctuations. Princeton Series in Physics. Princeton University Press, Princeton, NJ, 1985.
50. D. Nualart. The Malliavin calculus and related topics. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.
51. E. Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. Ann. Inst. H. Poincaré, 2(4):269–310, 1932.
52. D. W. Stroock and S. R. S. Varadhan. On the support of diffusion processes with applications to the strong maximum principle. In Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III: Probability theory, pages 333–359, 1972.
53. U. Tetsuji and L. K. William. Dynamics of optical pulses in randomly birefringent fibers. Physica D: Nonlinear Phenomena, 55(1):166–181, 1992.
54. C. Villani. Optimal transport, volume 338 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 2009. Old and new.
55. L. Wang, J. Hong, R. Scherer, and F. Bai. Dynamics and variational integrators of stochastic Hamiltonian systems. Int. J. Numer. Anal. Model., 6(4):586–602, 2009.
56. X. Wang, K. Lu, and B. Wang. Wong-Zakai approximations and attractors for stochastic reaction-diffusion equations on unbounded domains. J. Differential Equations, 264(1):378–424, 2018.

57. E. Wong and M. Zakai. On the convergence of ordinary integrals to stochastic integrals. Ann. Math. Statist., 36:1560–1564, 1965.
58. C. Wu and J. Zhang. Viscosity solutions to parabolic master equations and Mckean-Vlasov sdes with closed-loop controls. arXiv:1805.02639.