Dissertation

MCKEAN-VLASOV SDES WITH SINGULAR DRIFTS

zur Erlangung des akademischen Grades
Doktor der Mathematik (Dr. math)
Eingereicht von

Zimo Hao

Fakultät für Mathematik
Universität Bielefeld

April 2023

Acknowledgement

First and foremost, I am indebted to my supervisors Prof. Dr. Michael Röckner and Prof. Dr. Xicheng Zhang for their encouragement and support for study in Bielefeld. During my three years PhD program, Prof. Dr. Michael Röckner gave me a lot of help in research and life. Moreover, his rigorous academic attitude and serious writing habits benefited me a lot. It has always been a pleasure to work with and learn from such an extraordinary mathematician. My heartfelt thanks are due to him for his academic supervision and personal support. I owe a great deal to Prof. Dr. Xicheng Zhang for introducing me to this exciting project. He always encourages me to investigate a problem deeper and deeper. Many problems, with which I don't know how to deal, can always break through with his help. I am also very grateful for his recommendation, under which I got this precious opportunity to study in Bielefeld University. He has monitored my progress and offered advice and encouragement throughout.

I want to express my gratitude to Prof. Dr. Rongchan Zhu and Prof. Dr. Xiangchan Zhu who provided the help, guidance and support. My sincere thanks are due to their guidance for the road of scientific research. Moreover, I got a lot of valuable knowledge about singular stochastic partial differential equations after many times of discussion during the joint work with them.

I would like to thank Prof. Dr. Zhen-Qing Chen, Prof. Dr. Stephen Menozzi, Prof. Dr. Weijun Xu and Prof. Dr. Zhenfu Wang for their help and valuable suggestions for the research. Especially, my special thanks go to Prof. Dr. Zhenfu Wang for his invitation to give talks and his continued support about studying propagation of chaos.

I must thank the German Research Foundation (DFG) through the Collaborative Research Centre (CRC) 1283/2 2021-317210226 "Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications" for the award of the funding which enabled me to undertake.

I would like to thank Dr. Longjie Xie, Dr. Guohuan Zhao, Dr. Chengcheng Ling, Dr. Xuhui Peng, Dr. Mengyu Cheng, Xianliang Zhao, Zhengyan Wu, Yi Han, Kuang Huang and Yang Yu for their helpful discussions. In particular, I thank my colleague and joint worker Mengyu Cheng for my writing of this thesis, I also learned a lot from her about the averaging principle.

Thanks also to the Bielefeld University, for providing the data and office for my PhD study. Thanks to the staff of CRC1283 for their contributions to the research.

I also thank Dr. Andre Schenke for his help within the tutorial class and Monika Bokermann, Ole-Johann Thom, Dr. Jean Daniel Mukam and Dr. Sebastian Grube for their help in Bielefeld.

Finally, I owe my deepest and most sincere thanks to my wife Lu, my parents and my grandparents. They have always taken great pride in me and encouraged me to overcome difficulties in research and life. Words are powerless to express this gratitude.

Abstract

This thesis is mainly concentrated on McKean-Vlasov stochastic differential equations (also called distributional dependent stochastic differential equations, abbreviated as DDSDE) with singular drifts driven by Brownian motion. More specifically, we consider the following four aspects of it:

(i) Strong well-posedness. We use Zvonkin's transformation and entropy formula to obtain the strong well-posedness of DDSDE where the different components of the drift are in different mixed $L^{\boldsymbol{p}}$ spaces. To this end, we show the maximal mixed $L^{\boldsymbol{p}}$-regularity estimate for the related parabolic partial differential equation (PDE) by transferring a scalar solution to a vector solution for a new system. When the drift is independent of the distribution, this result shows the strong well-posedness of N-particle systems with L^p interaction kernels, which extends the results of Krylov and Röckner [67].

(ii) Propagation of chaos. We show strong convergence of the propagation of chaos for the particle approximation of DDSDEs with singular L^p interactions as well as for moderately interacting particle systems on the level of particle trajectories. Moreover, when the interaction kernel is bounded and measurable, we also obtain the optimal rate of strong convergence, which is partially based on Jabin's and Wang's entropy method [58] and Zvonkin's transformation.

(iii) Averaging principle. We study the averaging principle for DDSDEs with drift in localized L^p spaces. Using Zvonkin's transformation and estimates for solutions to Kolmogorov equations, we prove that the solution of the original system strongly and weakly converges to that of the averaged system as the time scale ε goes to zero. Moreover, we obtain rates of the strong and weak convergence that depend on p respectively.

(iv) Euler-Maruyama approximation. We use the Euler-Maruyama approximation to show the existence of a solution to a class of McKean-Vlasov SDEs of Nemytskii-type with bounded, measurable drift and any initial data. When the initial data have densities in L^q with respect to Lebesgue measure, based on the associated nonlinear Fokker-Planck equation and heat kernel estimates for the Euler-Maruyama scheme,

we show the uniqueness of solutions for the McKean-Vlasov SDEs of Nemytskii-type and obtain a convergence rate of the Euler-Maruyama approximation, which is the same as the rate for the SDE cases in [9].

Contents

Notations

Sets and Spaces

E	Denotes a Banach space
$\mathbb{R}^d$	d-dimensional real Euclidean space.
$\mathcal{P}(E)$	The set of probability measures on E
$C_b(E), C_b^k(E)$	Respectively the space of real-valued bounded continuous functions and the space of functions with k ($\geqslant 1$) bounded continuous derivatives on E. $C_b := C_b(\mathbb{R}^d)$, $C_b^k := C_b^k(\mathbb{R}^d)$ and $C_b^\infty := \cap_{k\in\mathbb{N}} C_b^k$.
C_0^k	The set of real-valued C_b^k functions with compact support on $\mathbb{R}^d$, $k \in \mathbb{N} \cup \{\infty\}$.
$L^p := L^p(\mathbb{R}^d)$	The set of measurable functions f defined almost everywhere on $\mathbb{R}^d$ with respect to Lebesgue measure such that $\|f\|^p$ is interable for $p \geqslant 1$. When $p = \infty$, this is the set of bounded measurable functions. We denote the space of locally integrable functions by L^1_{loc}.
$\\| \cdot \\|_p$	The norm of L^p.
$H^{\alpha,p}$	Bessel potential space on $\mathbb{R}^d$, see Section 2.1 below.
$\\| \cdot \\|_{\alpha,p}$	The norm of $H^{\alpha,p}$.
$\mathcal{C}^\alpha$	Hölder spaces on $\mathbb{R}^d$, see (2.4) below.
$\tilde{L}^p$, $\tilde{H}^{\alpha,p}$	Localized Bessel potential spaces and localized L^p spaces on $\mathbb{R}^d$ with $\tilde{L}^p = \tilde{H}^{0,p}$, see (2.2) below.
$\\| \cdot \\|_{\alpha,p}$	The norm of $\widetilde{H}^{\alpha,p}$, $\\| \cdot \\|_p := \\| \cdot \\|_{0,p}$.
S_d	$\{\boldsymbol{\pi} = (i_1, i_2, .., i_d)$: any permutation of $(1, 2, ..., d)\}$.
$\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}$	the space of mixed $\boldsymbol{p}$-summable functions with permutation $\boldsymbol{\pi}$, where $\boldsymbol{p} = (p_1, ..., p_d) \in [1, \infty]^d$, see (2.6) below.
$\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}$	Localized $\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}$-space, see (2.9) below.
$\mathbb{R}_+$	The set $[0, +\infty)$.
$\mathbb{L}^q_I(E) := L^q(I; E)$	The space of L^q functions from a time interval $I \subset \mathbb{R}_+$ to E. $\mathbb{L}^q_T(E) := \mathbb{L}^q_{[0,T]}(E)$ and $\mathbb{L}^p_T := \mathbb{L}^p_T(L^p)$.

$\mathbb{L}_q^p(I) := \mathbb{L}_I^q(L^p)$;	$\mathbb{H}_q^{\alpha,p}(I) := \mathbb{L}_I^q(H^{\alpha,p})$;
$\mathbb{L}_q^p(T) := \mathbb{L}_q^p([0,T])$;	$\mathbb{H}_q^{\alpha,p}(T) := \mathbb{H}_q^{\alpha,p}([0,T])$.
$\widetilde{\mathbb{H}}_q^{\alpha,p}(I)$, $\widetilde{\mathbb{L}}_I^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})$	The localized spaces of $\mathbb{H}_q^{\alpha,p}(I)$ and $\mathbb{L}_I^q(\mathbb{L}_{\boldsymbol{\pi}}^{\boldsymbol{p}})$, see (2.3) and (2.10) below respectively.
$\widetilde{\mathbb{L}}_q^p(I) := \widetilde{\mathbb{H}}_q^{0,p}(I)$;	$\widetilde{\mathbb{H}}_q^p(T) := \widetilde{\mathbb{H}}_q^{\alpha,p}([0,T])$;
$\widetilde{\mathbb{L}}_q^p(T) := \mathbb{L}_q^p([0,T])$;	$\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}}) := \widetilde{\mathbb{L}}_{[0,T]}^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})$;
$\widetilde{\mathbb{L}}^p(T) := \widetilde{\mathbb{L}}_\infty^p(T)$.	
$\mathscr{I}^o$	$\{(q,\boldsymbol{p}) \in (2,\infty)^{1+d} : \sum_{i=1}^d 1/p_i + 2/q < 1\}$.
$\mathscr{I}_m$	$\{(q,\boldsymbol{p}) \in (1,\infty)^{1+d} : \sum_{i=1}^d 1/p_i + 2/q < m\}, \quad m = 1,2$.
$C(I,E)$	The set of continuous functions from a time interval $I \subset \mathbb{R}_+$ to E, endowed with the uniform topology. $\mathbb{C}_T := C([0,T],\mathbb{R}^d)$.

Generic elements and operations

C	A generic non-negative constant, the value of which may change from line to line.
Θ	The set of parameters that a constant may depend on. It may have different parameters in different occasions, which should be clear from the context.
$C(\Theta), C_\Theta$	A generic non-negative constant which depends on some fixed parameters $\Theta = (a_1, a_2, .., a_n)$. Its value may change from line to line.
∂_t, ∂_i	The derivatives in the time variable t and in the x_i direction.
M^*	The transpose of the matrix M.
tr$M$	The trace of the matrix M.
M_{ij}	The (i,j)(respectively row and column indexes) component of a matrix M.
∇	The gradient operator $(\partial_1, \partial_2, .., \partial_d)^*$.
div	The divergence of a vector field $F : \mathbb{R}^d \to \mathbb{R}^d$ defined by $\mathrm{div}F := \sum_{i=1}^d \partial_i F_i$.
∇^2	The Hessian matrix $(\partial_i \partial_j)_{i,j=1,..,d}$.
$\mathbb{I}_{d\times d}$ or $\mathbb{I}$	The d-dimensional identity matrix. We do not specify the dimension d when no confusion is possible. $\delta_{ij} := \mathbb{I}_{ij}$.
$\langle x,y\rangle$ or $x\cdot y$	The Euclidean inner product of two vectors $x,y \in \mathbb{R}^d$ defined by $\langle x,y\rangle = x\cdot y := \sum_{i=1}^d x_i y_i$. We also use $M\cdot N$ to denote the Frobenius inner product of two matrices M,N defined by $M\cdot N := \sum_{i,j=1}^d M_{ij}N_{ij}$.
$\mathcal{M}$	The Hardy-Littlewood maximal function in $\mathbb{R}^d$, see (2.14) below.

$\mu \otimes \nu$	The product measure on $E \times F$ of two measures μ, ν on E and F respectively. $\mu^{\otimes 1} := \mu$ and $\mu^{\otimes k} := \mu \otimes \mu^{\otimes k-1}$ for $k \geqslant 2$.
$\mathbf{x}^N = (x^1, .., x^N)$	A generic element of a product space E^N.
$P_t f := g_t * f$	The semigroup related to the Laplace Δ (see (2.25) below).

Probability and measures

$(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathbb{P})$	A filtered probability space with the filtration satisfying the usual conditions. All the random variables are defined on this space unless otherwise stated. The expectation is denoted by $\mathbb{E}$.
w or $(w_t)_{t\in[0,T]}$	The canonical process on the path space $\mathbb{C}_T$ defined by $w_t(\omega) := \omega_t$.
$\mathbb{E}_{\mathbb{Q}}[\cdot]$	The expectation under another probability measure $\mathbb{Q}$.
$K \circledast \mu$	The convolution of a function $K : \mathbb{R}^{2d} \to E$ with a measure μ on $\mathbb{R}^d$ defined as $K * \mu(x) := \int_{\mathbb{R}^d} K(x,y)\mu(\mathrm{d}y) \in E$, $\forall x \in \mathbb{R}^d$.
δ_x	The Dirac measure at the point x.
$\eta_{\mathbf{x}^N}$	The empirical measure $\eta_{\mathbf{x}^N} := \frac{1}{N}\sum_{i=1}^N \delta_{x^i}$.
$\langle \mu, \varphi \rangle = \mu(\varphi)$	The integral of a μ-integrable function φ with respect to a measure μ.
$\mathscr{L}(X) = \mu_X$	The law of a random variable X as an element of $\mathcal{P}(E)$ where X takes its value in the space E.
$\|\cdot\|_{\mathrm{var}}$	The total variation norm for measures.
$\mathcal{H}(\mu\|\nu)$	The relative entropy between two measures μ and ν, see (2.57) below.
$X \overset{(d)}{=} \mu$	A random variable X has a law $\mathscr{L}(X) = \mu$.

Assumptions

$(\mathbf{H}_a)$	An assumpion on the second order coefficient a in a PDE, where a is uniformly θ-order Hölder continuous, bounded and elliptic (page 33).
$(\mathbf{H}^\sigma_{mix})$	An assumption on the diffusion coefficient σ in an SDE to obtain the strong well-posedness for SDEs with mixed $L^{\boldsymbol{p}}$ coeffcicents (page 49).
$(\mathbf{A}^\sigma)$	An assumption on the diffusion coefficient σ in an SDE, where σ is uniformly θ-order Hölder continuous, bounded and elliptic (page 50).
$(\mathbf{H}^b)$	An assumption of an interaction kernel b when we study propagation of chaos (page 100).
$(\mathbf{H}^1_b)$, $(\mathbf{H}^2_b)$, $(\mathbf{H}^\sigma)$	Assumptions of the coefficients in an SDE when we study the averaging principle (pages 126, 127).

Conventions:

- We use := to indicate a definition;
- $a \vee b := \max\{a, b\}$, $a \wedge b := \min\{a, b\}$ and $a^+ := 0 \vee a$;
- In this thesis, we will utilize Einstein summation notation, which implies that summation is performed over repeated indices.
- By $A \lesssim_C B$ and $A \asymp B$ or simply $A \lesssim B$ and $A \asymp B$, we mean that for some constant $C \geqslant 1$,

$$A \leqslant CB, C^{-1}B \leqslant A \leqslant CB, \quad \text{respectively;}$$

- We shall use the same notation Γ_ε to denote mollifiers in various dimensions N, i.e.,

$$\Gamma_\varepsilon(x) = \varepsilon^{-N}\Gamma(x/\varepsilon),\ \varepsilon \in (0,1), \tag{0.1}$$

  where Γ is a nonnegative smooth density function in $\mathbb{R}^N$ with compact support in the unit ball.
- For a function $f \in L^1_{\rm loc}(\mathbb{R}^N)$, the mollifying approximation of f is defined by

$$f_\varepsilon(x) := f * \mathit{\Gamma}_\varepsilon(x) = \int_{\mathbb{R}^N} f(x-y)\mathit{\Gamma}_\varepsilon(y)\mathrm{d}y \quad \text{or} \quad f_n := f * \Gamma_{1/n}, \quad n \in \mathbb{N}.$$

  The dimension N takes different values in different occasions, which should be clear from the respective context.

Chapter 1

Introduction

In this thesis, we consider the following McKean-Vlasov SDEs, which are also called distributional dependent SDEs (abbreviated as DDSDEs), with singular drift on $\mathbb{R}^d$:

$$\mathrm{d}X_t = b(t, X_t, \mu_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \tag{1.1}$$

where $(W_t)_{t\geqslant 0}$ is a standard d-dimensional Brownian motion, μ_t is the time marginal distribution of X_t, $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ is measurable and $b : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ is measurable (in particular not necessary continuous). By Itô's formula, one sees that μ_t solves the following nonlinear Fokker-Planck equation in the distributional sense:

$$\partial_t\mu_t = \frac{1}{2}\partial_i\partial_j\left(\sigma_{ik}\sigma_{jk}(t)\mu_t\right) - \mathrm{div}(b(t, \cdot, \mu_t)\mu_t) = 0. \tag{1.2}$$

The story of the McKean-Vlasov SDEs started with a stochastic toy model for the Vlasov equation of plasma proposed by Kac. The classical notion of propagation of (Kac's) chaos was formalized in [62] where Kac first derived the spactially homogeneous Boltzmann equation by use of N-particle systems. After that, Markov processes associated to (1.2) were first studied by McKean in [76].

Recently, since it naturally appears in the studies of the limiting behavior of interacting particle systems and mean-field games, there is a vast and growing interest in McKean-Vlasov SDEs (1.1), which describe stochastic systems at the microscopic scale and whose distributional-density satisfies the macroscopic description (1.2).

In the present paper, we mainly focus on some approximations to (1.1): N-particle systems approximation based on [45], averaging principle approximation based on [25] and Euler-Maruyama approximation based on [44]. Moreover, we also investigate the well-posedness (i.e. existence and uniqueness) of McKean-Vlasov SDEs and N-particle systems. In the following, we give an introduction for each of them.

1.1 McKean-Vlasov SDEs and large systems of particles

Large systems of interacting particles is now fairly common. They are usually models for large populations of individuals subject to mutual interaction and random dispersal. For example, in plasma physics particles can represent ions and electrons in the Vlasov-Poisson equation [103, 17]; in biosciences they characterize the collective behavior of individuals [94] and describe the growth of cancer [33]. We refer to the book [95] and recent reviews [55, 39, 57, 21, 22] for more details.

The most classical model is the Newton dynamics for N indistinguishable point particles driven by two-body interaction forces and noise. By $X_t^{N,i}$ and $V_t^{N,i}$ we denote the position and velocity of particle number $i = 1, 2, .., N$ at the time $t \in \mathbb{R}_+$ respectively. Based on Newton's second law, the evolution of the system is described by the following stochastic system:

$$\begin{cases} \mathrm{d}X_t^{N,i} = V_t^{N,i}\mathrm{d}t, \\ \mathrm{d}V_t^{N,i} = \frac{1}{N}\sum_{j\neq i} K(X_t^{N,i} - X_t^{N,j})\mathrm{d}t + \sigma \mathrm{d}W_t^i, \end{cases} \tag{1.3}$$

where $i = 1, 2, .., N$ and W^i are N independent standard d-dimensional Brownian motions, which model random influences. The critical scaling in (1.3) is the factor $\frac{1}{N}$ in front of the interaction term. This is called the mean field scaling and it preserves the conservation of the total strength of the interaction (see [55, Section 1.1] for more details). When $\sigma \equiv 0$, it reduces to the classical Newton dynamics. Here the vector valued kernel K stands for the interaction force between two particle. Such kind of a system is called a kinetic system and the infinitesimal generator of the system is related the kinetic opeqrator $\partial_t - \Delta_v + v \cdot \nabla_x$. We refer to [40, 79, 51, 56] for mathematical kinetic theory for particle system. See also [50, 49, 43] for the related kinetic SDEs and McKean-Vlasov SDEs with singular drifts and Lévy noise.

Moreover, upon regarding $V_t^{N,i}$ as the derivative of position $X_t^{N,i}$ with respect to time, we have

$$\mathrm{d}\dot{X}_t^{N,i} = \frac{1}{N}\sum_{j\neq i} K(X_t^{N,i} - X_t^{N,j})\mathrm{d}t + \sigma \mathrm{d}W_t^i.$$

Consequently, the system is commonly referred to as a second-order system.

In this thesis, we only consider the following first order N-particle system:

$$\mathrm{d}X_t^{N,i} = \frac{1}{N}\sum_{j=1}^{N} K(X^{N,i} - X^{N,j})\mathrm{d}t + \sigma \mathrm{d}W_t^i. \tag{1.4}$$

By Itô's formula, the time marginal distribution μ_t^N of $(X_t^{N,1},..,X_t^{N,N})$, satisfies the following Fokker-Planck equation:

$$\partial_t \mu_t^N = \frac{1}{2}\sigma^2 \sum_{i=1}^N \Delta_{x_i}\mu_t^N - \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \mathrm{div}_{x_i}\left(K(x_i - x_j)\mu_t^N\right). \tag{1.5}$$

Such system can model interacting particles in physics (for instance, the point vortex system in 2-dimensional fluids [35]), biological and sociological sciences phenomena of flocking, swarming, and aggregation (see [27, 16] for examples).

However, due to the large number of N, simulating the microscopic N-particle systems (1.4) directly is exceedingly complicated. Actually, the number N of particles can reach a scale of 10^{25} for most physical models and 10^9 in typical bioscience models (see [55, Section 1.2]). Even for $N = 4, 5$, the dynamics of ODE (1.4) (i.e. $\sigma = 0$) can be chaotic and it is difficult to follow orbits of particles (see for instance [111]). Fortunately, the N-particle system (1.4) can usually be approximated by a McKean-Vlasov SDE, thanks to the famous Laws of Large Numbers. More precisely, for large N, one expects to approximate the solution of (1.4) by the solution to the McKean-Vlasov SDE (1.1) with $\sigma(t, x) = \sigma$ and $b(t, x, \mu) = K * \mu$.

1.1.1 Propagation of chaos

In this subsection we recall some notions and well-known results about the propagation of chaos.

Classical framework introduced by Kac

Let E be a Polish space and $\mu \in \mathcal{P}(E)$ a probability measure on E. Let $(\mu^N)_{N\in\mathbb{N}}$ be a sequence of symmetric probability measures on the respective product space E^N, where symmetric means that for any permutation $(i_1, \cdots, i_N)$ of $(1, \cdots, N)$,

$$\mu^N(\mathrm{d}x_{i_1}, \cdots, \mathrm{d}x_{i_N}) = \mu^N(\mathrm{d}x_1, \cdots, x_N).$$

In particular, μ^N has a common 1-marginal distribution. One says that $(\mu^N)_{N\in\mathbb{N}}$ is μ-*chaotic* if for any $k \in \mathbb{N}$ (see [62]),

$$\mu^{N,k} \text{ weakly converges to } \mu^{\otimes k} \text{ as } k \leqslant N \to \infty, \tag{1.6}$$

where $\mu^{N,k}(\mathrm{d}x_1, \cdots, \mathrm{d}x_k) = \mu^N(\mathrm{d}x_1, \cdots, \mathrm{d}x_k, E, \cdots, E)$ is the k-fold marginal distribution of μ^N. It is well known that (1.6) holds if and only if (1.6) holds for only $k = 2$ (see [98, (i) of Proposition 2.2]). In the language of random variables, Kac's chaos can be restated as follows:

Let $\boldsymbol{\xi}^N := (\xi^{N,1}, \cdots, \xi^{N,N})$ be a family of E^N-valued random variables. If the law of $\boldsymbol{\xi}^N$ is symmetric and μ-chaotic, one says that $\boldsymbol{\xi}^N$ is μ-chaotic. It is also equivalent to (see [98, (ii) of Proposition 2.2])

$$\text{the empirical measure } \eta_{\boldsymbol{\xi}^N}(\mathrm{d}y) := \tfrac{1}{N}\textstyle\sum_{j=1}^N \delta_{\xi^{N,j}}(\mathrm{d}y) \in \mathcal{P}(E) \text{ converges to } \mu \text{ in law.} \tag{1.7}$$

Note that $\boldsymbol{\xi}^N$ can be regarded as N-random particles in the state space E. From this viewpoint, Kac's chaos means that if one observes the distribution of any k-particles, then they become statistically independent as N goes to infinity. Indeed, (1.7) is a law of large numbers, i.e., for any $\varphi \in C_b(E)$,

$$\eta_{\boldsymbol{\xi}^N}(\varphi) := \frac{1}{N}\sum_{j=1}^N \varphi(\xi^{N,j}) \to \mu(\varphi) := \int_E \varphi(x)\mu(\mathrm{d}x), \quad \text{in law.}$$

In Hauray and Mischler's work [51], various quantitative and qualitative estimates related to chaos are obtained for different notions such as Kac's chaos, entropy chaos and Fisher information chaos. More references about Kac's chaos can be also found in [51].

Propagation of chaos

If one considers Kac's chaos as a static version of chaos, then propagation of chaos is usually understood as a dynamical version of Kac's chaos. More precisely, let $(\boldsymbol{\xi}_t^N)_{t\geqslant 0} := (\xi_t^{N,1}, \cdots, \xi_t^{N,N})_{t\geqslant 0}$ be a family of E^N-valued continuous stochastic processes, which can be thought of as the evolution of N-particles. Let $(\xi_t)_{t\geqslant 0}$ be a limit E-valued continuous stochastic process defined on the same probability space. Let μ_t^N be the law of $\boldsymbol{\xi}_t^N$ in E^N and μ_t be the law of ξ_t in E. Suppose that μ_0^N is μ_0-chaotic at time 0. One says that *propagation of chaos* holds if for any time $t > 0$, μ_t^N is μ_t-chaotic. In this thesis, we call it the *weak convergence of the propagation of chaos*. Usually, as the evolution of particle distributions, the probability measures μ_t^N and μ_t satisfy some Fokker-Planck equation in the weak sense, like (1.5) and (1.2). Therefore, it can be studied by pure PDE methods. However, as stochastic processes, one would like to have the following *stronger convergence of the propagation of chaos*: for each $t > 0$,

$$\lim_{N\to\infty} \mathbb{E}|\xi_t^{N,1} - \xi_t| = 0,$$

or in the functional path sense

$$\lim_{N\to\infty} \mathbb{E}\left(\sup_{s\in[0,t]} |\xi_s^{N,1} - \xi_s|\right) = 0. \tag{1.8}$$

In fact, when K is globally Lipschitz continuous, McKean [77] firstly established the following result for (1.4): for any $T > 0$,

$$\mathbb{E}\left(\sup_{s\in[0,T]} |X_s^{N,1} - X_s|^2\right) \leqslant \frac{C(b,\sigma,T)}{N}, \tag{1.9}$$

where the constant $C(b,\sigma,T) > 0$ can be estimated explicitly and $(X_t)_{t\geqslant 0}$ is the solution to (1.1) with $b(t,x,\mu) = K * \mu$, $\sigma(t,x) = \sigma$ driven by Brownian motion W^1. We note that the power of convergence rate $1/N$ is sharp because of the central limit theorem. The above estimate was also reproven by Sznitman [98] by more direct synchronous coupling methods.

Singular kernel cases

Obviously, Lipschitz assumptions on the interaction kernel K is too strong in practice. In fact, most of the interesting physical models have bounded measurable or even singular interaction kernels. For examples, the rank-based interaction diffusion studied in [93, 69] has a discontinuous interaction kernel (see (5.11) below), and the Biot-Savart law appearing in the vortex description of 2d imcompressible Navier-Stokes equations has a singular kernel. We refer to [108, Section 1.3] for more examples of singular kernels. For this type of singular kernels, Osada [84] was the first to show the propagation of chaos for the point vortices associated with the 2d Navier-Stokes equation with large viscosity. Recently, in [38], Fournier, Hauray, and Mischler dropped the assumption of large viscosity by the classical martingale method.

More recently, Jabin and Wang [58] were the first to obtain a quantitative convergence rate for the relative entropy between the law of the particle system, $\mu^{N,k}$, and the tensorized limit law, $\mu^{\otimes k}$, where the key point is an estimate for the entropy and a large deviation type exponential functional. In fact, the results in [58] can be applied to a large class of singular kernels K in $W^{-1,\infty}$ with $K(x) = -K(-x)$, as well as to some nonlinear interactions with bounded measurable interaction kernel. More precisely, for the system

$$\mathrm{d}X_t^{N,i} = F\left(\frac{1}{N}\sum_{j\neq i} K(X_t^{N,i} - X_t^{N,j})\right)\mathrm{d}t + \sigma \mathrm{d}W_t^i, \tag{1.10}$$

where F is Lipschitz and K is bounded measurable, the author in [58] established a global relative entropy estimate of the form

$$\sup_{N\in\mathbb{N}} \mathcal{H}(\mu_t^{N,N}|\mu_t^{\otimes N}) < \infty,$$

as long as the same is true for $t = 0$, where the relative entropy $\mathcal{H}$ is defined by (2.57), $\mu_t^{N,N}$ and $\mu_t^{\otimes N}$ is the time marginal distribution of $\mu^{N,N}$ and $\mu^{\otimes N}$ respectively (see Section

5.2.2 for the proof). From the well-known subadditivity inequality (2.61) and Pinsker's inequality (2.58), one deduces the local estimates

$$\mathcal{H}(\mu_t^{N,k}|\mu_t^{\otimes k}) \leqslant C\frac{k}{N}, \quad \text{and} \quad \|\mu_t^{N,k} - \mu_t^{\otimes k}\|_{var} \leqslant C\sqrt{\frac{k}{N}},$$

where $\|\cdot\|_{var}$ stands for the total variation distance. We note that the proof in [58] for kernels in $W^{-1,\infty}$ strongly depends on the symmetry of the kernel $K(x)$, not valid for general L^p-singular kernels.

When $F(x) = x$ in (1.10), Lacker in [70] applies the BBGKY hierarchy to give the following optimal weak convergence rate for the total variation distance between $\mu^{N,k}$ and $\mu^{\otimes k}$ for general L^∞ kernels:

$$\|\mu_t^{N,k} - \mu_t^{\otimes k}\|_{var} \leqslant C\frac{k}{N}.$$

It should be noted that the linear assumption $F(x) = x$ can not be dropped there, since the linearity of conditional expectations are needed.

For general L^p-singular interaction kernels, in [99], Tomašević uses the partial Girsanov transform as in [59] to derive the weak convergence of the propagation of chaos under the extra assumption that the set of discontinuous points of the interaction kernel has Lebesgue measure zero. In [53], Hoeksema, Holding, Maurelli and Tse showed a large deviation result for a particle system with L^p-singular interaction kernels. As a byproduct, they also obtain the weak convergence of the propagation of chaos (see also [69]). However, in [99] and [53], both of them assume the initial distributions of the particle system are i.i.d, that is, the initial distributions are not really *chaotic*. This assumption is crucial for them to construct a weak solution for the interacting particle system by Girsanov's transform. In the present paper we overcome this difficulty by showing the existence of strong solutions for the particle system (see Lemma 5.6 below), and then obtain the both weak and strong convergence of the propagation of chaos for singular L^p interaction kernels with $p > d$ and the quantitative convergence (1.9) for bounded measurable kernels by Zvonkin's transformation, a method introduced in [118] by Zvonkin and further developed in [101, 67, 113]; see Section 3.1.1 for details. It should be noted that Bao and Huang [2] have already used the Zvonkin transformation to obtain propagation of chaos for Hölder interaction kernels with non-optimal rate $N^{-1/4}$.

Moderately interacting particle systems

When K is singular, say a Dirac measure or Poisson kernel $K(x) = \pm x/|x|^d$, the N-particle system (1.4) is not expected to have a solution in general. A usual way of tackling this problem is to mollify the kernel by $K_N := K * \phi_{\varepsilon_N}$, where ε_N goes to zero as $N \to \infty$, $\phi_\varepsilon(x) := \varepsilon^{-d}\phi(x/\varepsilon)$ and $\phi \in C_0^\infty$ is a smooth probability density function. For the choice of $\varepsilon_N = N^{-\beta/d}$ with $\beta \in [0,1]$, if $\beta = 0$ and 1, it respectively corresponds to weakly interacting and strongly interacting, while for $\beta \in (0,1)$, it is called moderately interacting by Oelschläger in [81], and the particle system (1.4) with $K = K_N$ is called a moderately interacting particle system.

Especially, when K is the Dirac measure, we are interested in the moderately interacting kernel $K_N = \phi_{\varepsilon_N}$, where $\varepsilon_N \to 0$ as $N \to \infty$ and the following N-particle systems

$$\mathrm{d}X_t^{N,i} = F\big(t, X_t^{N,i}, \frac{1}{N}\sum_{j\neq i} \phi_{\varepsilon_N}(X_t^{N,i} - X_t^{N,j})\big)\mathrm{d}t + \sigma \mathrm{d}W_t^i, \quad i = 1, \cdots, N. \tag{1.11}$$

Then, the solution $X^{N,i}$ is expected to converge to the solution of the following equation (see [81, 61]):

$$\mathrm{d}X_t = F(t, X_t, \rho_t(X_t))\mathrm{d}t + \sigma \mathrm{d}W_t, \tag{1.12}$$

where ρ_t stands for the density of the law of X_t with respect to Lebesgue measure. Here $\rho := (\rho_t)_{t\geqslant 0}$ solves the following nonlinear and *local (or Nemytskii-type)* Fokker-Planck equation:

$$\partial_t \rho = \frac{\sigma^2}{2}\Delta\rho + \mathrm{div}(F(\rho)\rho).$$

It should be kept in mind that for $d = 1$ and $F(\rho) = \rho$, this is a Burgers-type equation.

Equation (1.12) is called McKean-Vlasov SDE of Nemytskii-type (also called density dependent SDE, abbreviated as dDSDE). For the following more general cases,

$$\mathrm{d}X_t = b\Big(t, X_t, \rho_t(X_t)\Big)\mathrm{d}t + \sigma\Big(t, X_t, \rho_t(X_t)\Big)\mathrm{d}W_t, \quad X_0 \overset{(d)}{=} \nu_0, \tag{1.13}$$

they were first introduced and investigated in [5, Section 2] (see also [3]).

In [81] Oelschläger showed the weak convergence of the propagation of chaos for moderately interacting particle systems, when F and ϕ are smooth. Generally, the moderately interacting refers to any choice of ε_N with $\varepsilon_N \to 0$ and $\varepsilon_N^{-1}/N = o(1)$. When F, ϕ are smooth enough, Jourdain and Méléard in [61] showed the following strong convergence:

$$\mathbb{E}\left(\sup_{t\in[0,T]} |X_t^{N,1} - X_t|^2\right) \leqslant C\left(\varepsilon_N^4 + \frac{\varepsilon_N^2}{N} e^{C\varepsilon_N^{-2d-2}}\right), \tag{1.14}$$

with a constant C, where $(X_t)_{t\in[0,T]}$ is the solution to dDSDE (1.12) driven by Brownian motion W^1.

Recently, this type of moderately interacting systems has regained much attention after the semigroup approach developed by Flandoli, Leimbach and Olivera [33], see for instance [34] for a PDE-ODE system related to aggregation phenomena; [94] for non-local conservation laws; [35] for 2D Navier-Stokes equation and [82] for quantitative convergence results for a variety of singular kernels. However, all the above results does not cover the Nemytskii-type caes, i.e. $K = \delta_0$, and the mollifer function ϕ in their work is required to be smooth. In the current thesis, we shall investigate the strong convergence of the propagation of chaos for the moderately interacting particle systems when $K = \delta_0$ and the mollifier ϕ is only bounded measurable.

1.1.2 Well-posedness of particle systems and McKean-Vlasov SDEs

Before studying the propagation of chaos with singular drift (interacting kernel) K, one of the main obstacles is to establish the strong or weak well-posedness of the SDEs for both the particle systems (1.4) and the McKean-Vlasov SDE (1.1).

For the motion of a single particle, when $\phi \in L_t^q(L_x^p)$ with

$$\frac{d}{p} + \frac{2}{q} < 1,$$

Krylov and Röckner [67] showed the existence and uniqueness of strong solutions to the following SDE by Girsanov's and Zvonkin's transformation:

$$\mathrm{d}X_t = \phi_t(X_t)\mathrm{d}t + \mathrm{d}W_t.$$

Later, Zhang [113, 114] extended their result to the multiplicative noise case, again using Zvonkin's transformation from [118] (see also [116, 109]). However, for the N-particle system (1.4) with $\phi(x, y) = K(x - y)$, where K is in some L^p space, one cannot use these well-known results for $L_t^q(L_x^p)$ drifts to derive the well-posedness by considering (1.4) as an SDE in $\mathbb{R}^{Nd}$. For instance, when $N = 3$, consider the following SDE in $\mathbb{R}^{3d}$:

$$\begin{cases} \mathrm{d}X_t^1 = \Big[\phi\big(X_t^1, X_t^2\big) + \phi\big(X_t^1, X_t^3\big)\Big]\mathrm{d}t + \mathrm{d}W_t^1, \\ \mathrm{d}X_t^2 = \Big[\phi\big(X_t^2, X_t^1\big) + \phi\big(X_t^2, X_t^3\big)\Big]\mathrm{d}t + \mathrm{d}W_t^2, \\ \mathrm{d}X_t^3 = \Big[\phi\big(X_t^3, X_t^1\big) + \phi\big(X_t^3, X_t^2\big)\Big]\mathrm{d}t + \mathrm{d}W_t^3, \end{cases} \tag{1.15}$$

where $\phi(x, y) = K(x - y)$ and $K \in L^p$ with $p > d$. For $i = 1, 2, 3$, let $\phi_i(x_1, x_2, x_3) := \sum_{j \neq i} \phi(x_i, x_j)$. As a function of (x_1, x_2, x_3) in $\mathbb{R}^{3d}$, one only has

$$\phi_i \in L^\infty_{x_i^*} L^p_{x_i}, \quad i = 1, 2, 3, \tag{1.16}$$

where x_i^* stands for the remaining variables except for x_i. It does not satisfy the conditions in [67]. Note that in the same work [67], Krylov and Röckner also showed the strong well-posedness for a class of special stochastic particle system with singular gradient interaction

$\phi = \nabla V$, where V is continuously differentiable on $\mathbb{R}^d\backslash\{0\}$ and satisfies some other conditions (see Section 9 in [67]). Moreover, the strong well-posedness for particle system with Biot-Savart law interaction kernel $\phi(x) = (-x_2, x_1)/|x|^2$ was established in [83] and [37], which is related to the random point vortex approximation for two dimensional Navier-Stokes equations. In the above well-known papers, the key point of establishing the strong well-posedness is to prove that the process $X_t^i - X_t^j$ for $i \neq j$ does not touch the singular point 0, i.e. the state space is $\mathbb{R}^{Nd}$ "without diagonals". However, the strong well-posedness for particle systems as in (1.4) with general L^p-interaction kernels on all of $\mathbb{R}^{Nd}$ has still been open.

Therefore, our first task is to extend [67, 114] to the case of mixed L^p-spaces. We mention here that although Ling and Xie [74] have already considered singular SDEs in mixed L^p-spaces, their result cannot be applied to equation (1.15) due to the new feature that we need to consider the order of the integral in x_1, x_2, x_3 as well as the different integrability indices. Note that each ϕ_i belongs to a different mixed L^p-space.

Let us turn to the well-posedness of McKean-Vlasov SDEs (1.1). So far there are numerous literatures devoted to studying this problem. When b is bounded and $\mu \to b(t, x, \mu)$ is uniformly Lipschitz with respect to the Wasserstein distance, Li and Min [72] obtained existence and uniqueness for weak solutions. Under some one-side Lipschitz assumptions (also with respect to the Wasserstein distance), Wang [104] showed the strong well-posedness and also some functional inequalities for the solution. We want to emphsize that the Lipschitz assumption for $\mu \to b(t, x, \mu)$ with respect to the Wasserstein distance is not satisfied by the case mean-field limit SDEs with non-continuous kernels. More precisely, consider

$$b(x, \mu) = \int_{\mathbb{R}^d} K(x - y)\mu(\mathrm{d}y)$$

with some kernel K which is only bounded or in L^p. Then, $\mu \to b(x, \mu)$ is not even continuous with respect to the Wasserstein metric for every point x, but $\mu \to b(\cdot, \mu)$: $\mathcal{P}(\mathbb{R}^d) \to L^p(\mathbb{R}^d)$ is Lipschitz with respect to the total variation distance.

In the case, where b is only measurable, of at most linear growth and $\mu \to b(t, x, \mu)$ is uniformly Lipschitz continuous with respect to the total variation distance with σ uniformly non-degenerate and Lipschitz continuous, by using the classical Krylov estimates, Mishura and Veretenikov [80] obtained strong well-posedness for (1.1). After that, the strong well-posedness was extended to local $L_t^q L_x^p$ drift by Röckner and Zhang in [89]. Furthermore, by the relative entropy method and Girsanov's theorem, Lacker [70] also obtained some well-posedness results under linear growth assumptions (see also [69]). Then, in the special case $|b(t, x, \mu)| \leqslant h * \mu(t, x)$ with some $h \in L_t^q L_x^p$, Han obtained well-posedness for $L_t^q L_x^p$ drift based on the relative entropy method in [42]. In [117], by some heat kernel estimates and the Schauder-Tychonoff fixed point theorem, Zhao established well-posedness for DDSDE in a more general case.

Furthermore, weak solutions to the dDSDE (1.13), were constructed in [5], first solving the corresponding Fokker-Planck-Kolmogorov equation and using the superposition principle. In [3, 5], for a large class of time independent coefficients b, σ, Barbu and Röckner obtained the existence of weak solutions for such (possibly degenerate) density dependent SDEs (see [5, Section 2]). The strategy in [3] and [5] is to solve the associated nonlinear Fokker-Planck equation and then by the well-known superposition principle (cf. [100], generalizing [68] and [32]) to establish the existence of a weak solution to dDSDE (1.13). Later, in [4, 7, 8], the same authors prove the uniqueness of weak solutions to dDSDE (1.13), which is a consequence of the uniqueness of the corresponding nonlinear Fokker-Planck equation and its linearized version. Recently, in [6], they also consider the existence of solutions to a class of nonlinear Fokker-Planck equations with measure-valued initial data. It is natural to ask for a probabilistic method to construct the solution. In this thesis, we consider the following general distributional density-distributional dependent SDE (abbreviated as dDDSDE):

$$\mathrm{d}X_t = b(t, X_t, \rho_t(X_t), \mu_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \tag{1.17}$$

where μ_t and ρ_t is the time marginal distribution and density with respect to Lebesgue measure of X_t respectively. Here we assume that $r \to b(\cdot, r, \mu)$ and $\mu \to \sup_r |b(\cdot, r, \mu)|$ are both Lipschitz from $\mathbb{R}_+$ and $\mathcal{P}(\mathbb{R}^d)$, with respect to the total variation distance, to $L_t^q L^p$ respectively. For this model, we start directly from dDDSDE (1.17) and obtain the existence of solutions by using Picard iteration, entropy formula and heat kernel estimates. In other words, we don't use the superposition principle. Moreover, our assumptions on the drift are weaker compared to Barbu and Röckner's series of papers. Especially, there is no regularity assumption of b in x. On the other hand, we only consider the case when σ is independent of ρ_t (see Chapter 4 for more details).

1.2 Averaging principle for McKean-Vlasov SDEs

The averaging principle is one of the main methods in perturbation theory. It came into being by Clairaut, Laplace and Lagrange more than two centuries ago. The averaging principle was first established for the following deterministic systems by Krylov, Bogolybov and Mitropolsky [11, 65]:

$$X_t^\varepsilon = X_0 + \int_0^t b(\frac{s}{\varepsilon}, X_s^\varepsilon)\mathrm{d}s, \tag{1.18}$$

where $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ is a vector field, $0 < \varepsilon \ll 1$ is the time scale and s/ε is called the highly oscillating time component (s/ε is also called fast variable, while X_s^ε is called slow variable). Then it was extended to stochastic differential equations by Khasminskii [63]. After that, extensive work on the averaging principle for finite and infinite dimensional stochastic differential equations was done; see e.g. [1, 18, 19, 20, 26, 30, 36, 41, 64, 75, 85, 102, 107] and the references therein.

Usually, solving the original system (1.18) is relatively difficult because of the high oscillating time component. Therefore, it is desirable to find a simplified system which simulates and predicts the evolution of the original system over a long time scale. As is well known, the highly oscillating time component can be "averaged" out to produce such a simplified system under some suitable conditions, which is called averaging principle.

More exactly, consider the following averaged system:

$$X_t = X_0 + \int_0^t \bar{b}(X_s)\mathrm{d}s, \tag{1.19}$$

where

$$\bar{b}(x) := \lim_{T\to\infty} \frac{1}{T}\int_0^T b(t,x)\mathrm{d}t. \tag{1.20}$$

b is called a *KBM-vector feld* (KBM stands for Krylov, Bogolyubov and Mitropolsky) if the convergence (1.20) is uniformly with respect to x in any bounded subsets of $\mathbb{R}^d$ (see e.g. [90]). The averaging principle states that, as the time scale ε goes to zero, the solution of the original systems (1.18) converges to that of (1.19). We note that if b is bounded and periodic function with respect to time t, then

$$\bar{b}(x) = \frac{1}{T_b}\int_0^{T_b} b(t,x)\mathrm{d}t,$$

where T_b is a period of b, and

$$\omega(T) := \sup_{t,x}\left|\frac{1}{T}\int_t^{t+T}(b(t,x)-\bar{b}(x))\mathrm{d}s\right| \to 0 \quad \text{as } T\to\infty. \tag{1.21}$$

Let us give a brief proof of the averaging principle for ODE systems (1.18), here under the condition (1.21) and if b is bounded and Lipschitz. First of all, it is easy to see that

$$|X_t^\varepsilon - X_t| \leqslant \|b\|_{Lip}\int_0^t |X_s^\varepsilon - X_s|\mathrm{d}s + \left|\int_0^t \Big(b(\frac{s}{\varepsilon},X_s) - \bar{b}(X_s)\Big)\,\mathrm{d}s\right|, \tag{1.22}$$

where $\|b\|_{Lip}$ is the Lipschitz constant of b. For $h\in(0,1)$, we define $\pi_h(t) := t$ for $t\in[0,h)$ and

$$\pi_h(t) := [t/h]h, \quad t\geqslant h.$$

The reason why we define $\pi_h(t) = t$ for $t\in[0,h)$ is that the drift b considered in this paper is always in an L^p space. If the initial data do not have an L^q density, $\mathbb{E}b(X_{\pi_h(t)}) = \mathbb{E}b(X_0)$

will blow up for all $t < h$. Although here b is Lipschitz and bounded, we use this definition for $\pi_h(t)$ in the whole thesis.

Then, we note that

$$\begin{aligned}\left|b(\frac{s}{\varepsilon}, X_s) - \bar{b}(X_s)\right| \leqslant &\left|b(\frac{s}{\varepsilon}, X_s) - b(\frac{s}{\varepsilon}, X_{\pi_h(s)})\right| + \left|\bar{b}(X_s) - \bar{b}(X_{\pi_h(s)})\right| \\ &+ \left|b(\frac{s}{\varepsilon}, X_{\pi_h(s)}) - \bar{b}(X_{\pi_h(s)})\right| \\ \leqslant &2\|b\|_{Lip}|X_{\pi_h(s)} - X_s| + \left|b(\frac{s}{\varepsilon}, X_{\pi_h(s)}) - \bar{b}(X_{\pi_h(s)})\right|,\end{aligned}$$

and we have

$$\left|\int_0^t \left(b(\frac{s}{\varepsilon}, X_s) - \bar{b}(X_s)\right) \mathrm{d}s\right| \leqslant 2\|b\|_{Lip}\|b\|_\infty h + \left|\int_0^t \left(b(\frac{s}{\varepsilon}, X_{\pi_h(s)}) - \bar{b}(X_{\pi_h(s)})\right) \mathrm{d}s\right|.$$

We set $N := [t/h]$. Based on a change of variable, one sees that

$$\begin{aligned}&\left|\int_0^t \left(b(\frac{s}{\varepsilon}, X_{\pi_h(s)}) - \bar{b}(X_{\pi_h(s)})\right) \mathrm{d}s\right| \\ &\leqslant \left|\sum_{k=1}^{N-1} \int_{kh}^{(k+1)h} \left(b(\frac{s}{\varepsilon}, X_{kh}) - \bar{b}(X_{kh})\right) \mathrm{d}s\right| + 4h\|b\|_\infty \\ &\leqslant h \sum_{k=1}^{N-1} \left|\frac{\varepsilon}{h} \int_{kh/\varepsilon}^{kh/\varepsilon + h/\varepsilon} \left(b(\frac{s}{\varepsilon}, X_{kh}) - \bar{b}(X_{kh})\right) \mathrm{d}s\right| + 4h\|b\|_\infty \\ &\leqslant hN\omega(h\varepsilon) + 4h\|b\|_\infty \leqslant t\omega(h/\varepsilon) + 4h\|b\|_\infty.\end{aligned}$$

In view of (1.22) and Gronwall's inequality, we have

$$\sup_{t\in[0,T]} |X_t^\varepsilon - X_t| \leqslant C(T, b) \inf_{h\in(0,1)} (h + \omega(h/\varepsilon)) \to 0, \quad \text{as } \varepsilon \to 0.$$

This method is called technique of time discretization. And this kind of convergence is called strong convergence analogous to the one in the propagation of chaos.

In this thesis, we are interested in using the techniques of time discretization to investigate the averaging principle of the following DDSDE with highly oscillating time component

$$\mathrm{d}X_t^\varepsilon = b\left(\frac{t}{\varepsilon}, X_t^\varepsilon, \mu_t^\varepsilon\right)\mathrm{d}t + \sigma(X_t^\varepsilon)\mathrm{d}W_t, \quad X_0^\varepsilon = \xi, \tag{1.23}$$

where $\sigma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ is a measurable function, $\mu_t^\varepsilon := \mathcal{L}(X_t^\varepsilon)$ is the time marginal law of X_t^ε, $0 < \varepsilon \ll 1$ is the time scale and the drift b is only L^p integrable in x (see Chapter 6 for the concrete conditions). Then the averaged equation is

$$\mathrm{d}X_t = \bar{b}(X_t, \mu_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \quad X_0 = \xi, \tag{1.24}$$

where μ_t stands for the distribution of X_t and

$$\bar{b}(x,\mu) = \lim_{\varepsilon\to 0}\frac{1}{T}\int_0^T b(t,x,\mu)\mathrm{d}t$$

where the limit on the right hand side is assumed to exist (see condition $(\mathbf{H}_b^2)$ on page 126 for details).

Recall that the strong convergence rate of the averaging principle for slow-fast McKean-Vlasov SDE was established by the techniques of time discretization and Poisson equation in [88]. Furthermore, as discussed in [54], the strong convergence rate of the averaging principle for slow-fast McKean-Vlasov SPDE was studied, based on the variational approach and the technique of time discretization. Note that the coefficients of the slow equation with fast variables were assumed to be globally Lipschitz continuous with respect to the slow variable in the above results. Recently, the strong convergence without a rate for DDSDE with highly oscillating time component driven by fractional Brownian motion and standard Brownian motion was shown in [92], under the assumption that the drift term is continuous in the slow variable.

Recall that Lipschitz or mere continuity assumptions on b are too strong for some applications. There are a lot of interesting models from physics, only having bounded measurable or even singular L^p interaction kernels b. However, to the best of our knowledge, there is no result concerning the averaging principle both of DDSDE and SDE with L^p drift.

Following the above motivations, we consider the strong and weak convergence of the averaging principle for DDSDE with L^p drift in this thesis. Moreover, we obtain the rate of the strong and weak convergence, which is important for functional limit theorems in probability and homogenization in PDEs. To show this kind of results, we have to overcome the difficulty of the non-continuity of b. More precisely, we need to estimate the following difference for some L^p coefficient b:

$$\mathbb{E}\left|\int_0^t \left(b(\frac{s}{\varepsilon},X_s) - b(\frac{s}{\varepsilon},X_{\pi_h(s)})\right)\mathrm{d}s\right|. \tag{1.25}$$

To this end, partially inspired by [28] and [73], we use a different technique based on the Markov property and the time regularity of the semigroup to calculate (1.25) (see Lemma 3.24). After that, we show a distributional version of (1.25) in Lemma 3.30, which makes the classical time discretization method work again (see Section 6.1 for more details).

1.3 Euler-Maruyama scheme for dDSDEs

In this part, we would like to study the Euler-Maruyama scheme for the following dDSDE

$$X_t = X_0 + \int_0^t b(s,X_s,\rho_s(X_s))\mathrm{d}s + \sqrt{2}W_t, \tag{1.26}$$

where ρ_t is the distributional-density of X_t with respect to Lebesgue measure and b is uniformly bounded. More precisely, for $h \in (0,1)$, we consider the following Euler-Maruyama scheme:

$$X_t^h = X_0 + \int_0^t b(s, X^h_{\pi_h(s)}, \rho^h_{\pi_h(s)}(X^h_{\pi_h(x)}))\mathrm{d}s + \sqrt{2}W_t, \tag{1.27}$$

where ρ_s^h is the distributional-density of X_s^h. whose existence is easily seen from the construction. We note that there is no continuity assumed for the drift term b with respective to x, which is an obstacle for establishing the convergence. When $b = b(t,x)$ is bounded, Dini continuous and independent of the measure, the following strong convergence is obtained by Dareiotis and Gerencsér in [28] by elementary calculations: for any $\varepsilon \in (0,1)$

$$\sup_{t\in[0,T]} \mathbb{E}\left|X_t^h - X_t\right|^2 \leqslant Ch^{1-\varepsilon}.$$

It is extended to L^p cases by Lê and Ling in [73] based on the stochastic sewing lemma.

In the present thesis, we consider the weak convergence (i.e. the convergence between the time marginal law). Assuming that b is Lipschitz with respect to $r \to b(t,x,r)$, the strong convergence is directly deduced from the weak convergence by stability estimates with respect to drift b and results in [28, 73]. When b is only bounded and independent of the density ρ_t, for the weak convergence with respect to the total variation distance, Bencheikh and Jourdain obtain the following rate in [9]:

$$\|\rho_t^h - \rho_t\|_1 \lesssim \sqrt{h}. \tag{1.28}$$

They used Duhamel's formula for the Euler-Maruyama scheme and calculated each term carefully. In this thesis, we develop a technique based on estimates of the semigroup to deduce the weak convergence rate (1.28). For more well-known results and applications of Euler-Maruyama scheme for SDE we refer to papers mentioned above.

1.4 Main results

First we obtain the strong well-posedness to the following non-distribution dependent SDE:

$$\mathrm{d}X_t = b(t,X_t)\mathrm{d}t + \sigma(t,X_t)\mathrm{d}W_t, \tag{1.29}$$

where different component of the drift $b = (b_1, ..., b_d)$ is in different mixed L^p spaces, as e.g. in SDE (1.15). More precisely, in **Theorem 3.1** (page 49) we show that if σ is uniformly

Hölder continuous in x with respect to t and elliptic, and for some $q_i, p_{i,j} \in [2,\infty)$ and $\boldsymbol{\pi}_i \in S_d$, $i = 0, 1, .., d$ and $j = 1, 2, ..., d$ and every T,

$$\|\nabla\sigma\|_{\widetilde{\mathbb{L}}_T^{q_0}(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}_0}^{\boldsymbol{p}_0})} + \sum_{i=1}^{d} \|b_i\|_{\widetilde{\mathbb{L}}_T^{q_i}(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}_i}^{\boldsymbol{p}_i})} < \infty, \tag{1.30}$$

with

$$\sup_i \left(\frac{2}{q_i} + \frac{1}{p_{i1}} + \cdots + \frac{1}{p_{id}}\right) < 1,$$

where $\boldsymbol{p}_i := (p_{i1}, ..., p_{id})$, then there exists a unique strong solution to SDE (1.29). Here $\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})$ is the mixed L^p space with permutation (see (2.10) below). As an application, we have a unique strong solution to the N-particle system (1.4) for any L^p interaction kernel K with $p > d$ (see Remark 3.2 for more details). For this part we need to construct Zvonkin's transformation. To this end, in Theorem 2.19, we transfer the scalar solution u of PDE (2.33) to the vector solution $(u_0, u_1, ..., u_d)$ of a new system (2.48) and obtain maximal L^p-regularity.

Next, to prove well-posedness of dDDSDE (1.17), we show the following stability results for densities of the solutions to (1.29) with respect to the drift (see Lemma 4.4 below): Let b^0, b^1 be two Borel measurable functions satisfying (1.30) and $\rho_k(t,x)$, $k = 0, 1$, be the densities of the time marginal laws of two solutions to SDE (1.29) with $b = b^k$. Then for any $T > 0$, there is a constant $C > 0$ such that for all $t \in [0, T]$,

$$\|\rho_0(t) - \rho_1(t)\|_{\mathbb{L}^\infty} \leqslant C \sum_{i=1}^{d} \int_0^t (t-s)^{-\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}_i}|)} \|b_i^0(s) - b_i^1(s)\|_{\widetilde{\mathbb{L}}_{\boldsymbol{\pi}_i}^{\boldsymbol{p}_i}} \mathrm{d}s. \tag{1.31}$$

Following the proof of this stability, we proceed to formulate **Theorem 4.1** (page 91). In this theorem, if the initial datum admits a bounded density, σ satisfies assumption (1.30) and

$$\sup_{i,\mu} \|\sup_{r\geqslant 0} |b_i(\cdot,\cdot,r,\mu)|\|_{\widetilde{\mathbb{L}}_T^{q_i}(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}_i}^{\boldsymbol{p}_i})} < \infty,$$

by Picard iteration and the entropy difference (2.64), we show well-posedness for dDDSDE (1.17). Moreover, based on the strong well-posedness of non-distribution dependent SDE (1.29), we obtain the unique strong solution to the McKean-Vlasov SDE (1.17) in **Theorem 4.2** (page 91).

After the strong well-posedness is obtained for N-particle systems and limit McKean-Vlasov SDEs, we study the strong convergence of the propagation of chaos. We first obtain it in **Theorem 5.1** (page 100) for the following particle systems:

$$\mathrm{d}X_t^{N,i} = F(t, X_t^{N,i}, \frac{1}{N}\sum_{j\neq i}\phi_t(X_t^{N,i}, X_t^{N,j}))\mathrm{d}t + \sigma(t, X_t^{N,i})\mathrm{d}W_t^i,$$

where σ satisfies the condition (1.30) and for some measurable $h : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}_+$ and $\kappa_1 > 0$

$$|F(t,x,r)| \leqslant h(t,x) + \kappa_1|r|, \quad |F(t,x,r) - F(t,x,r')| \leqslant \kappa_1|r-r'|,$$

and for some $q > 2$, $\boldsymbol{p} = (p_1, .., p_d) \in [2,\infty]^d$ with $2/q + 1/p_1 + ... + 1/p_d < 1$ and $\boldsymbol{\pi} \in S_d$, $T > 0$,

$$\|h\|_{\mathbb{L}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})} + \left[\int_0^T \sup_{y\in\mathbb{R}^d}\left(\|\phi_t(\cdot,y)\|^q_{\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}}} + \|\phi_t(y,\cdot)\|^q_{\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}}}\right)\mathrm{d}t\right]^{\frac{1}{q}} \leqslant \kappa_1.$$

To show this strong convergence, we use the partial Girsanov transform as used in [59, 99] to show uniform Krylov's estimate for particle systems, which implies the weak convergence of the propagation of chaos by the classical martingale approach (see **Theorem 5.5** below, page 104). It should be noted that in **Theorem 5.5**, by the strong well-posedness of (1.29), the solution to the particle system is a measurable functional of Brownian motion. In contrast to [99, 53], we thus need not to assume that the initial data are independent.

For the strong convergence, we use Zvonkin's transformation and Lemma 5.12, which transfers weak convergence to strong convergence. When ϕ is uniformally bounded, combining with the entropy method developed in [58], we obtain the optimal convergence rate

$$\mathbb{E}\left(\sup_{t\in[0,T]}|X_t^{N,i} - X_t^i|\right) \lesssim N^{-1/2},$$

where X^i is the unique strong solution to limiting DDSDE driven by Brownian motion W^i.

Moreover, we also consider the moderately interacting particle systems (1.11) and obtain the strong convergence rate of the propagation of chaos in **Theorem 5.3** (page: 102) when F and ϕ are uniformly bounded. More precisely, we extend the result (1.14) from smooth to non-continuous cases. The proof is based on Theorem 5.1 and the stability estimate (1.31).

The results for well-posedness and the propagation of chaos are from the joint paper [45] with Röckner and Zhang.

For the averaging principle for DDSDE, in Chapter 6 we consider the systems (1.23) with highly oscillating time component and the averaged DDSDE (1.24). Assuming that

$\mu \to b(t,\cdot,\mu)$ from $\mathcal{P}(\mathbb{R}^d)$ to some localized L^p space is Lipschitz with respect to the total variation distance uniformly in $t \in \mathbb{R}_+$, under some time periodic condition like (1.21), for any $p > d$, we obtain the following weak and strong convergence rate in **Theorem 6.2** (page 127):

$$\sup_{t\in[0,T]} \|\mu_t^\varepsilon - \mu_t\|_{var} \leqslant C \inf_{h>0} \left(h^{\frac{1}{2}-\frac{d}{2p}} + \omega\left(\frac{h}{\varepsilon}\right)\right) \tag{1.32}$$

and

$$\mathbb{E}\left(\sup_{t\in[0,T]} |X_t^\varepsilon - X_t|^{2\ell}\right) \leqslant C \inf_{h>0} \left((\omega(h/\varepsilon))^2 + h^{1-\frac{d}{p}}\right)^\ell .$$

Moreover, we also obtain an analogous result for non-distribution dependent SDE in Theorem 6.3, where the convergence rate is independent of p. To show these results, we need a distribution dependent version for (1.25) (see Lemma 3.30 below) and the solution to the related PDE as in Lemma 3.29 and Lemma 3.32. After these estimates, we obtain the weak convergence rate. For the strong convergence, we use Zvonkin's transformation again, i.e. the same technique as in the part about the propagation of chaos.

The results for the averaging principle are based on the joint paper [25] with Cheng and Röckner.

Finally, we consider the Euler-Maruyame scheme (1.27) for dDSDE (1.13), where b is uniformly bounded and $\sigma = \sqrt{2}$. If $r \to b(t,x,r)$ is continuous, we obtain the existence of solution to dDSDE (1.13) and show that there is a sequence $h_k \to 0$, as $k \to \infty$, such that X^{h_k} converges to this solution of (1.13) in the weak sense. Moreover, if $r \to b(t,x,r)$ is Lipschitz uniformly in (t,x) and assuming that the initial datum admit s an L^q density with $q > d$, we obtain the well-posedness of dDSDE (1.13) and gain the following convergence rate for the Euler-Maruyame scheme:

$$\sup_{t\in[0,T]} \|\rho_t^h - \rho_t\|_1 \leqslant C_T\sqrt{h},$$

where the rate is the same as that in (1.28), obtained in [9] for the non-distribution dependent case. All the above results are included in **Theorem 7.2** (page 146). It should be noted that this is a different way of proof in comparison with that of Theorem 4.1 to show well-posedness of McKean-Vlasov SDEs. Furthermore, the assumptions for this result are also different from those in Theorem 4.1: Although b is bounded here and not in some localized L^p space, the initial datum condition is less restrictive than that in Theorem 4.1, where the time marginal law of the initial datum is required to have a bounded density with respect to Lebesgue measure. In particular, we obtain the existence for any initial datum $X_0 = x \in \mathbb{R}^d$, whose distribution is a Dirac measure and dose not have a density with respect to Lebesgue measure.

The well-posedness of the dDSDE (1.13) and the convergence properties of the Euler-Maruyama scheme have been established in a joint paper [44] with Röckner and Zhang. The convergence rate of the scheme is first rigorously obtained in this thesis.

1.5 Structure of the thesis

In Chapter 2 we first introduce the basic function spaces used in this thesis, localized Bessel potential spaces and localized mixed $L^{\boldsymbol{p}}$ spaces, and main properties of these spaces in Section 2.1. Then we study second order parabolic PDEs with mixed localized $L^{\boldsymbol{p}}$-drifts and show the unique existence of strong solutions. Before that, when $\boldsymbol{p} = (p, p, ..., p)$, we recall some results from [109] about second order parabolic PDEs with localized L^p drift. On the basis of these results for the parabolic PDEs, we are able to derive the corresponding results for the elliptic PDEs. We note that Lemma 2.8 is not a corollary of Theorem 2.19, since the conditions on f are different. In Theorem 2.19, since each component of the drift may be in a different mixed $L^{\boldsymbol{p}}$-space, the new point here is that the second order derivative of the solution shall stay in a direct sum space. Apart from these, we also introduce basic concepts about SDE in Section 2.3 and some well-known results about the relative entropy in Section 2.4 for later use.

In Chapter 3, we study the well-posedness and some properties of solution to SDEs with L^p drifts. We show the weak and strong well-posedness for SDE with mixed $L^{\boldsymbol{p}}$-drifts in Section 3.1. As usual, we need to prove a priori Krylov estimates based on the PDE estimates obtained in Section 2.2, and then show that we can perform the Zvonkin transformation. To this end, we first give a proof for the main result of [109] in Section 3.1.1 as an example how to use Zvonkin's transformation. Therein we will show what the Zvonkin transformation is and how to derive the Zvonkin transformation and strong well-posedness for SDE from Krylov estimates and results from PDEs. Moreover, in Section 3.2, we study the time regularity of solutions to SDE with L^p-drifts, which is used to show time discretization type estimates in averaging principle, like (1.25). For (1.25), we first assume that $\{X_t\}_{t\in[0,T]}$ is a solution to the SDE (1.29) with out drift, that is $b \equiv 0$, based on the assumption on σ. We obtain some time difference estimates of the corresponding heat kernel, which implies the estimate (1.25) for any localized L^p function b with $p > d \vee 2$ (see Lemma 3.24 below). Then it follows from the Girsanov transform that (1.25) holds for the solution to (1.29) with localized L^p drifts (see (3.79) below). Furthermore, we obtain the "distribution dependent version" of it, namely Lemma 3.30, which is based on Lemma 3.29, which in turn provides an estimation of the difference $\mu_t - \mu_s$ between the time marginal distributions. As a byproduct, we also obtain the time regularity of the gradient of the solutions for parabolic PDE in Section 3.2.3, which is used in Chapter 6.

In Chapter 4, by Picard's iteration, we show the weak and strong well-posedness for dDDSDEs (1.17) (see Section 2.3 for their definitions) with mixed $L^{\boldsymbol{p}}$-drifts. We use the entropy formula, Pinsker's inequality and the stability (1.31) to show that the density of the time marginal law of the Picard iteration is a Cauchy sequence in $L^1 \cap L^\infty$ (see Lemma 4.4 below).

In Chapter 5, we prove the strong converagence of the propagation of chaos. First, by the classical martingale method we show that the propagation of chaos for systems as in (1.4) with singular kernels holds in the weak sense, where the key point is to use the partial Girsanov transform used in [59, 99] to derive some uniform estimate for the exponential functional. In particular, the strong solution is used to treat the chaos of the initial distributions. Moreover, we also provide a detailed proof for Jabin and Wang's quantitative result [58] for bounded interaction kernels. This is not new and only for the readers' convenience. Then we give the proof of the strong convergence of the propagation of chaos and show how to use Zvonkin's transformation again to derive the strong convergence from the weak convergence, where the key point is Lemma 5.12.

In Chapter 6, based on the time regularity estimates obtained in Section 3.2, we use a Itô-Tanaka trick to give the weak convergence rate for the averaging principle of DDSDE with localized L^p drift (see Theorem 6.6 below). Moreover, based on Zvonkin's transformation again, we obtain the strong convergence rate from the weak convergence. Here, Lemma 6.5 is of crucial importance. This lemma is derived from the time periodic condition (1.21) and the "distribution dependent version" of the time discretization, as shown in Lemma 3.30. In Section 6.5, we also give some examples to illustrate our results.

In Chapter 7, we study the Euler-Maruyama approximation for dDSDE (1.26) with bounded drift and use Euler-Maruyama scheme to give a proof of well-posedness of (1.26). We first establish some estimates for the density of the time marginal distributions for the Euler-Maruyama scheme (1.27) with bounded measurable drifts, by which we prove the compactness of the time marginal distributional densities of the Euler-Maruyama scheme (1.27) as to $h \to 0$. Then we obtain the well-posedness by approximation from this Euler-Maruyama scheme. Moreover, by using the technique from Section 6.3, we obtain the weak convergence rate for the Euler-Maruyama scheme.

The Appendix contains technical lemmas, two types of Gronwall inequalities and Schauder estimates for the parabolic equations used in the proofs of our results.

Chapter 2

Preliminary

2.1 Localized Bessel potential spaces and mixed L^p spaces

In this section, we introduce the definition of localized Bessel potential space and localized mixed $\mathbb{L}^{\boldsymbol{p}}$ spaces for later use.

Let $d \in \mathbb{N}$. For any $(\alpha, p) \in \mathbb{R} \times [1, \infty]$, we write

$$H^{\alpha,p} := (\mathbb{I} - \Delta)^{-\alpha/2} \left(L^p(\mathbb{R}^d)\right)$$

for the usual Bessel potential space (see [96, Chapter V] for example) with the norm given by

$$\|f\|_{\alpha,p} := \|(\mathbb{I} - \Delta)^{\alpha/2} f\|_p,$$

where $\|\cdot\|_p$ is the usual $L^p(\mathbb{R}^d)$-norm. Here $(\mathbb{I} - \Delta)^{\alpha/2} f$ is defined through Fourier's transform

$$(\mathbb{I} - \Delta)^{\alpha/2} f := \mathcal{F}^{-1} \left((1 + |\cdot|^2)^{\alpha/2} \mathcal{F} f\right).$$

We note that if $\alpha = n \in \mathbb{N}$ and $p \in (1, \infty)$, an equivalent norm in $H^{n,p}$ is given by

$$\|f\|_{n,p} \asymp \|f\|_p + \|\nabla^n f\|_p.$$

For any $r > 0$, let B_z^r be the ball in $\mathbb{R}^d$ with radius r and center z. Let $\chi : \mathbb{R}^d \to [0, 1]$ be a smooth cutoff function with $\chi|_{B_1} = 1$ and $\chi|_{B_2^c} = 0$. For fixed $r > 0$, we set

$$\chi_z^r(x) := \chi((x - z)/r), \quad x, z \in \mathbb{R}^d. \tag{2.1}$$

Given $r > 0$, we introduce the following localized $H^{\alpha,p}$-space:

$$\widetilde{H}^{\alpha,p} := \left\{ f \in H^{\alpha,p}_{\mathrm{loc}}(\mathbb{R}^d) : \|\!|f|\!\|_{\alpha,p} := \sup_z \|\chi_r^z f\|_{\alpha,p} < \infty \right\}. \tag{2.2}$$

Clearly, this space does not depend on r and the corresponding norms are equivalent. When $\alpha = 0$, we simply write

$$\widetilde{L}^p := \widetilde{H}^{0,p} \quad \text{and} \quad |||f|||_p := |||f|||_{0,p}.$$

It follows from Hölder's inequality that for any $1 \leqslant p_2 \leqslant p_1 \leqslant \infty$

$$L^{p_1} \subset \widetilde{L}^{p_1} \subset \widetilde{L}^{p_2}.$$

This monotonic property is the main advantage of using localized spaces.

For $0 \leqslant t_0 < t_1$, $T > 0$, $p, q \in [1, \infty]$ and $\alpha \in \mathbb{R}$, we set

$$\mathbb{L}^p_q(t_0, t_1) := L^q\big([t_0, t_1]; L^p\big), \quad \mathbb{L}^p_q(T) := \mathbb{L}^p_q(0, T), \quad \mathbb{H}^{\alpha,p}_q(T) := L^q\big([0, T]; H^{\alpha,p}\big).$$

Now we introduce the localized space

$$\widetilde{\mathbb{H}}^{\alpha,p}_q(T) := \left\{ f \in \mathbb{H}^{\alpha,p}_q(T) : |||f|||_{\widetilde{\mathbb{H}}^{\alpha,p}_q(T)} := \sup_{z\in\mathbb{R}^d} \|\chi^z_r f\|_{\mathbb{H}^{\alpha,p}_q(T)} < \infty \right\}. \tag{2.3}$$

By a finite covering technique, it can be verified that also the definition of $\widetilde{\mathbb{H}}^{\alpha,p}_q$ does not depend on the choice of r (see [109, Section 2]). We note that all these spaces are Banach spaces and that

$$L^q\big([0, T]; \widetilde{H}^{\alpha,p}\big) \subset \widetilde{\mathbb{H}}^{\alpha,p}_q(T).$$

For $\alpha = 0$, set

$$\widetilde{\mathbb{L}}^p_q(T) := \widetilde{\mathbb{H}}^{0,p}_q(T).$$

If $q = \infty$, for simplicity, we define

$$\widetilde{\mathbb{H}}^{\alpha,p}(T) := L^\infty([0, T]; \widetilde{H}^{\alpha,p}), \quad \widetilde{\mathbb{L}}^p(T) := \widetilde{\mathbb{L}}^p_\infty(T), \quad \text{and} \quad \mathbb{L}^\infty_T := L^\infty([0, T] \times \mathbb{R}^d).$$

Moreover, for $\alpha \geqslant 0$, let $\mathcal{C}^\alpha$ be the usual Hölder space with norm:

$$\|f\|_{\mathcal{C}^\alpha} := \sum_{j=0}^{[\alpha]} \|\nabla^j f\|_\infty + \sup_{x\neq y\in\mathbb{R}^d} \frac{|\nabla^{[\alpha]} f(x) - \nabla^{[\alpha]} f(y)|}{|x - y|^{\alpha-[\alpha]}}, \tag{2.4}$$

where ∇^j stands for the j-order gradient and $[\alpha]$ stands for the integer part of α.

Lemma 2.1 (Embedding lemma)**.** *Let $1 < p < \infty$. Then we have*

$$\widetilde{H}^{\alpha,p} \subset \mathcal{C}^{\alpha-d/p}$$

and

$$\widetilde{\mathbb{H}}^{\alpha,p}(T) \subset L^\infty\left([0, T]; \mathcal{C}^{\alpha-d/p}\right)$$

provided $\alpha > d/p$.

Proof. It follows from Sobolev's embedding theorem that $H^{\alpha,p} \subset \mathcal{C}^{\alpha-d/p}$ if $\alpha > d/p$. Note that

$$\|g\|_{\mathcal{C}^{\alpha-d/p}} \leqslant \sup_z \|\chi_r^z g\|_{\mathcal{C}^{\alpha-d/p}}$$

for all $g \in \mathcal{C}^{\alpha-d/p}$ and $r > 0$. Therefore, we have

$$\|g\|_{\mathcal{C}^{\alpha-d/p}} \leqslant \sup_z \|\chi_r^z g\|_{\mathcal{C}^{\alpha-d/p}} \lesssim \sup_z \|\chi_r^z g\|_{H^{\alpha,p}} = \|g\|_{\widetilde{H}^{\alpha,p}}.$$

Moreover, for all $f \in \widetilde{\mathbb{H}}^{\alpha,p}(T)$ one sees that

$$\sup_{t\in[0,T]} \|f(t)\|_{\mathcal{C}^{\alpha-d/p}} \leqslant \sup_{t\in[0,T]} \sup_z \|\chi_r^z f(t)\|_{\mathcal{C}^{\alpha-d/p}} \lesssim \sup_{t\in[0,T]} \sup_z \|\chi_r^z f(t)\|_{H^{\alpha,p}} = \|f\|_{\widetilde{\mathbb{H}}^{\alpha,p}(T)}.$$

□

The following lemma is from [67].

Lemma 2.2. *Let $p, q \in (1, \infty]$, $T > 0$ and $u \in \widetilde{\mathbb{H}}_q^{2,p}(T)$ with $\partial_t u \in \widetilde{\mathbb{L}}_q^p(T)$. The following statements hold.*
(i) If $d/p + 2/q < 2$, then $u(t, x)$ is a bounded Hölder continuous function on $[0, T] \times \mathbb{R}^d$. More precisely, for any $R > 0$, $\varepsilon, \delta \in (0, 1)$ satisfying

$$\varepsilon + d/p + 2/q < 2, \quad 2\delta + d/p + 2/q < 2,$$

there is a constant $C = C(d, p, q, R, \varepsilon, \delta)$ such that for all $s, t \in [0, T]$ and $x, y \in \mathbb{R}^d$ with $|x|, |y| \leqslant R$

$$|u(t, x) - u(s, y)| \leqslant C \left(|t - s|^\delta + |x - y|^\varepsilon\right). \tag{2.5}$$

(ii) If $d/p+2/q < 1$, then $\nabla_x u$ is Hölder continuous in $[0, T]\times\mathbb{R}^d$, that is for any $\varepsilon \in (0, 1)$ satisfying

$$\varepsilon + d/p + 2/q < 1,$$

there is a constant $C = C(d, p, q, \varepsilon, T, R)$ such that for all $s, t \in [0, T]$ and $x, y \in B_R$, (2.5) holds with $\nabla_x u$ in place of u and $\varepsilon/2$ in place of δ.

Proof. We note that by definition, for any $R > 0$, $u1_{B_R} \in \mathbb{H}_q^{2,p}(T)$ with $\partial_t u1_{B_R} \in \mathbb{L}_q^p(T)$. Thus, it is direct from [67, Lemma 10.1]. □

Now, let's introduce the definition of localized mixed L^p-spaces, which was originally introduced in [10]. As we have seen in the introduction, these are very suitable for singular interacting particle systems (see also [53]).

We set

$$S_d := \{\boldsymbol{\pi} = (i_1, i_2, .., i_d) \ : \ \text{any permutation of } (1, 2, ..., d)\}.$$

For a multi-index $\boldsymbol{p} = (p_1, \cdots, p_d) \in (0, \infty]^d$ and any permutation $\boldsymbol{\pi} \in S_d$, the mixed $\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}$-space is defined by

$$\|f\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} := \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \cdots \left(\int_{\mathbb{R}} |f(x_1, \cdots, x_d)|^{p_d} \mathrm{d}x_{i_d} \right)^{\frac{p_{d-1}}{p_d}} \cdots \mathrm{d}x_{i_2} \right)^{\frac{p_1}{p_2}} \mathrm{d}x_{i_1} \right)^{\frac{1}{p_1}}. \tag{2.6}$$

When $\boldsymbol{p} = (p, \cdots, p) \in (0, \infty]^d$, the mixed $\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}$-space is the usual $L^p(\mathbb{R}^d)$-space. Note that for general $\boldsymbol{\pi} \neq \boldsymbol{\pi}'$ and $\boldsymbol{p} \neq \boldsymbol{p}'$,

$$\mathbb{L}^{\boldsymbol{p}'}_{\boldsymbol{\pi}} \neq \mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}} \neq \mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}'}.$$

For multi-indices $\boldsymbol{p}, \boldsymbol{q} \in (0, \infty]^d$, we shall use the following notations:

$$\frac{1}{\boldsymbol{p}} := \Big(\frac{1}{p_1}, \cdots, \frac{1}{p_d}\Big), \quad \boldsymbol{p} \cdot \boldsymbol{q} := \sum_{i=1}^d p_i q_i, \quad \Big|\frac{1}{\boldsymbol{p}}\Big| = \sum_{i=1}^d \frac{1}{p_i},$$

and

$$\boldsymbol{p} > \boldsymbol{q} \ \ (\text{resp. } \boldsymbol{p} \geqslant \boldsymbol{q}; \ \boldsymbol{p} = \boldsymbol{q}) \Longleftrightarrow p_i > q_i \ (\text{resp. } p_i \geqslant q_i; \ p_i = q_i) \ \text{ for all } i = 1, \cdots, d.$$

Moreover, we use bold numbers to denote constant vectors in $\mathbb{R}^d$, for example,

$$\mathbf{1} = (1, \cdots, 1), \quad \mathbf{2} = (2, \cdots, 2).$$

For multi-indices $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r} \in (0, \infty]^d$ with $\frac{1}{\boldsymbol{p}} + \frac{1}{\boldsymbol{r}} = \frac{1}{\boldsymbol{q}}$, the following Hölder inequality holds

$$\|fg\|_{\mathbb{L}^{\boldsymbol{q}}_{\boldsymbol{\pi}}} \leqslant \|f\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} \|g\|_{\mathbb{L}^{\boldsymbol{r}}_{\boldsymbol{\pi}}}. \tag{2.7}$$

For any multi-indices $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r} \in [1, \infty]^d$ with $\frac{1}{\boldsymbol{p}} + \frac{1}{\boldsymbol{r}} = \mathbf{1} + \frac{1}{\boldsymbol{q}}$, the following Young inequality holds

$$\|f * g\|_{\mathbb{L}^{\boldsymbol{q}}_{\boldsymbol{\pi}}} \leqslant \|f\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} \|g\|_{\mathbb{L}^{\boldsymbol{r}}_{\boldsymbol{\pi}}}. \tag{2.8}$$

For $\boldsymbol{p} \in [1, \infty]^d$, we introduce the following localized $\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}$-space:

$$\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}} := \Big\{ f \in L^1_{\mathrm{loc}}(\mathbb{R}^d), |\!|\!| f |\!|\!|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} := \sup_z \|\chi^r_z f\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} < \infty \Big\}, \tag{2.9}$$

and for a finite time interval $\mathrm{I} \subset \mathbb{R}$ and $q \in [1, \infty]$,

$$\widetilde{\mathbb{L}}^q_{\mathrm{I}}(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}) := \Big\{ f \in L^1_{\mathrm{loc}}(\mathrm{I} \times \mathbb{R}^d), |\!|\!| f |\!|\!|_{\widetilde{\mathbb{L}}^q_{\mathrm{I}}(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} := \sup_z \|\chi^r_z f\|_{\mathbb{L}^q_{\mathrm{I}}(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} < \infty \Big\}, \tag{2.10}$$

where for a Banach space $\mathbb{B}$ we set

$$\mathbb{L}^q_{\rm I}(\mathbb{B}) := L^q({\rm I};\mathbb{B}).$$

By the finitely covering technique again, it is easy to see that the definitions of $\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}$ and $\widetilde{\mathbb{L}}^q_{\rm I}(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})$ do not depend on the choice of r, and for any $1 \leqslant q_2 \leqslant q_1 \leqslant \infty$ and $1 \leqslant \boldsymbol{p}_2 \leqslant \boldsymbol{p}_1 \leqslant \infty$,

$$\widetilde{\mathbb{L}}^{\boldsymbol{p}_1}_{\boldsymbol{\pi}} \subset \widetilde{\mathbb{L}}^{\boldsymbol{p}_2}_{\boldsymbol{\pi}}, \quad \widetilde{\mathbb{L}}^{q_1}_{\rm I}(\widetilde{\mathbb{L}}^{\boldsymbol{p}_1}_{\boldsymbol{\pi}}) \subset \widetilde{\mathbb{L}}^{q_2}_{\rm I}(\widetilde{\mathbb{L}}^{\boldsymbol{p}_2}_{\boldsymbol{\pi}}). \tag{2.11}$$

Since the supremum z in the definition of $\widetilde{\mathbb{L}}^q_{\rm I}(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})$ is taken outside the time integral, we obviously have

$$\mathbb{L}^q_{\rm I}(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}) \subset \widetilde{\mathbb{L}}^q_{\rm I}(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}).$$

For simplicity we write

$$\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}) := \widetilde{\mathbb{L}}^q_{[0,T]}(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}), \quad \mathbb{L}^p_T := \mathbb{L}^p_{[0,T]}(\mathbb{L}^p), \ \mathbb{L}^\infty_T(\mathcal{C}^\alpha) := \mathbb{L}^\infty_{[0,T]}(\mathcal{C}^\alpha).$$

Example 2.3. For $i = 1, \cdots, d$ and $\alpha \in (0,1)$, let $f_i(x) = b(x)|x_i|^{-\alpha}$, where $b(x)$ is a bounded measurable function. It is easy to see that $f_i \in \widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}_i}$, where

$$\boldsymbol{\pi}_i = (1, \cdots, i-1, i+1, \cdots, d, i)$$

and $\boldsymbol{p} = (\infty, \cdots, \infty, p)$ with $p \in (1, \frac{1}{\alpha})$. From this example, one sees that for a C^1-diffeomorphism Φ from $\mathbb{R}^d$ to $\mathbb{R}^d$, say $\Phi(x) = (x_i, x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_d)$, it may happen that

$$f_i \circ \Phi \notin \widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}_i}.$$

The following lemma is obvious by the definitions.

Lemma 2.4. *For any $f \in \widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}$, there is a constant $C = C(\boldsymbol{p}) > 0$ such that for all $\varepsilon \in (0,1)$,*

$$|||f_\varepsilon|||_{\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} \leqslant C|||f|||_{\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}}, \tag{2.12}$$

and for any $R > 0$,

$$\lim_{\varepsilon \to 0} \|(f_\varepsilon - f)\chi^R_0\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} = 0. \tag{2.13}$$

The local Hardy-Littlewood maximal function in $\mathbb{R}^d$ is defined by

$$\mathcal{M}f(x) := \sup_{r \in (0,1)} \frac{1}{|B^r_0|} \int_{B^r_0} f(x+y){\rm d}y. \tag{2.14}$$

The following result is taken from [109, Lemma 2.1] and [52, Theorem 4.1].

Lemma 2.5. *(i) There is a constant $C = C(d) > 0$, such that for any $f \in L^\infty(\mathbb{R}^d)$ with $\nabla f \in L^1_{\text{loc}}(\mathbb{R}^d)$,*

$$|f(x) - f(y)| \leqslant C|x-y|\big(\mathcal{M}|\nabla f|(x) + \mathcal{M}|\nabla f|(y) + \|f\|_\infty\big) \tag{2.15}$$

for Lebesgue-almost all $x, y \in \mathbb{R}^d$.

(ii) For any $(q, \boldsymbol{p}) \in (1,\infty)^{1+d}$, there is a $C = C(d,p,q) > 0$ such that for all $f \in \widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})$,

$$\|\mathcal{M}f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \leqslant C\|f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{2.16}$$

2.2 PDE with singular coefficients

2.2.1 Localized L^p spaces

In order to study DDSDE, we consider the following second order parabolic PDE in $\mathbb{R}_+ \times \mathbb{R}^d$:

$$\partial_t u = a_{ij}\partial_i\partial_j u - \lambda u + b \cdot \nabla u + f, \quad u(0) = \varphi, \tag{2.17}$$

where $\lambda \geqslant 0$, $a = (a_{ij}) : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ is a symmetric matrix-valued Borel measurable function satisfying **(H$_a$)**, i.e.,

(H$_a$) there exist constants $c_0 > 0$ and $\theta \in (0,1)$ such that

$$c_0^{-1}|\xi| \leqslant |a(t,x)\xi| \leqslant c_0|\xi|, \quad \|a(t,x) - a(t,y)\|_{HS} \leqslant c_0|x-y|^\theta$$

for all $\xi \in \mathbb{R}^d$, $t \in \mathbb{R}_+$ and $x, y \in \mathbb{R}^d$,

and $b : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ is a vector-valued Borel measurable function. Firstly, we introduce the definition of a solution to PDE (2.17).

Definition 2.6. Let $T > 0$, $p, q \in (1,\infty)$, $\lambda \geqslant 0$, $b, f \in \widetilde{\mathbb{L}}^p_q(T)$ and $\varphi \in C^\infty_b$. We call a function u with $\partial_t u \in \widetilde{\mathbb{L}}^p_q(T)$ and $u \in \widetilde{\mathbb{H}}^{2,p}_q(T)$ a *solution* of PDE (2.17) if for Lebesgue almost all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$,

$$u(t,x) = \int_0^t \big(a_{ij}\partial_i\partial_j u(s,x) - \lambda u(s,x) + b \cdot \nabla u(s,x) + f(s,x)\big)\mathrm{d}s + \varphi(x).$$

Remark 2.7. For any $\chi \in C^\infty_0(\mathbb{R}^d)$ and $f \in \widetilde{\mathbb{H}}^{\alpha,p}_q(T)$, by the definition of the localized spaces $\widetilde{\mathbb{H}}^{\alpha,p}_q(T)$, we have $\chi f \in \mathbb{H}^{\alpha,p}_q(T)$. Hence for any solution u of PDE (2.17) in the sense of Definition 2.6, χu is Hölder continuous on $[0,T] \times \mathbb{R}^d$ if $d/p + 2/q < 2$ according to [67, Lemma 10.2]. Moreover, $\nabla(\chi u)$ is Hölder continuous on $[0,T] \times \mathbb{R}^d$ if $d/p + 2/q < 1$. In view of the arbitrariness of the cut-off function χ, u (respectively, ∇u) are locally Hölder continuous on $[0,T] \times \mathbb{R}^d$ if $d/p + 2/q < 2$ (respectively, $d/p + 2/q < 1$).

The following property of the solution u comes from [109, Theorem 3.2].

Lemma 2.8. *Let $T>0$, $\lambda\geqslant 0$ and $p,q\in(1,\infty)$. Assume* $(\mathbf{H}_a)$ *holds and $b\in\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)$ with $p_1\geqslant p$, $q_1\geqslant q$ and $2/q_1+d/p_1<1$. Set*

$$\Theta:=(d,T,p,q,\|b\|_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)},c_0,\theta).$$

Then there is a constant $\lambda_0=\lambda_0(\Theta)$ such that for all $\lambda\geqslant\lambda_0$ and $f\in\widetilde{\mathbb{L}}^p_q(T)$ and $\varphi\in\widetilde{H}^{2,p}$, there is a unique solution u to PDE (2.17) on $[0,T]$ in the sense of Definition 2.6 such that for any $\alpha\in[0,2)$, $p'\in[p,\infty]$, $q'\in[q,\infty]$ with

$$\beta:=2-\alpha+\frac{2}{q'}+\frac{d}{p'}-\big(\frac{2}{q}+\frac{d}{p}\big)>0, \tag{2.18}$$

there is a constant $C=C(\Theta,\alpha,p',q')>0$ such that for all $\lambda\geqslant\lambda_0$

$$\lambda^{\frac{\beta}{2}}\|u\|_{\widetilde{\mathbb{H}}^{\alpha,p'}_{q'}(T)}+\|\partial_t u\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\nabla^2 u\|_{\widetilde{\mathbb{L}}^p_q(T)}\leqslant C\big(\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\varphi\|_{2,p}\big). \tag{2.19}$$

Remark 2.9. By Lemma 2.1, we have $u\in L^\infty([0,T];\mathbf{C}^\gamma)$ for any $\gamma\in(1,2-2/q-d/p)$.

Proof. We note that u is a solution to PDE (2.17) with $u(0)=\varphi$ in the sense of Definition 2.6 if and only if $\bar u:=u-\varphi$ is a solution to PDE (2.17) with $\bar u(0)=0$ and $f=f+a_{ij}\partial_i\partial_j\varphi-\lambda\varphi-b\cdot\nabla\varphi$. Based on Lemma 2.1, we have

$$\begin{aligned}\|f+a_{ij}\partial_i\partial_j\varphi-\lambda\varphi-b\cdot\nabla\varphi\|_{\widetilde{\mathbb{L}}^p_q(T)}&\lesssim\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\varphi\|_{2,p}+\|b\|_{\widetilde{\mathbb{L}}^p_q(T)}\|\nabla\varphi\|_\infty\\&\lesssim\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+(1+\|b\|_{\widetilde{\mathbb{L}}^p_q(T)})\|\varphi\|_{2,p},\end{aligned}$$

and complete the proof by [109, Theorem 3.2]. □

With the help of a priori estimate (2.19), we obtain the well-posedness of PDE (2.17) for any $\lambda\geqslant 0$.

Proposition 2.10. *Let $T>0$, $p,q\in(1,\infty)$ with $2/q+d/p<1$, $\lambda\geqslant 0$, $b\in\widetilde{\mathbb{L}}^p_q(T)$. Then for all $f\in\widetilde{\mathbb{L}}^p_q(T)$ and $\varphi\in\widetilde{H}^{2,p}$ there is a unique solution u to PDE (2.17) on $[0,T]$ in the sense of Definition 2.6 such that*

$$\|\nabla u\|_{\widetilde{\mathbb{L}}^\infty_T}+\|\partial_t u\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|u\|_{\widetilde{\mathbb{H}}^{2,p}_q(T)}\leqslant C\left(\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\varphi\|_{\widetilde{H}^{2,p}}\right), \tag{2.20}$$

where $C=C(\Theta,\lambda)$.

Proof. By the standard continuity method, it suffices to show a priori estimate (2.20) for (2.17). To this end, we rewrite (2.17) as

$$\partial_t u=a_{ij}\partial_i\partial_j u-(\lambda+\lambda_0)u+b\cdot\nabla u+f+\lambda_0 u,$$

where λ_0 is as in Lemma 2.8. In view of Lemma 2.8, we have

$$(\lambda+\lambda_0)^{\frac{\beta}{2}}\|u\|_{\widetilde{\mathbb{H}}^{\alpha,p}_{q'}}+\|\partial_t u\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\nabla^2 u\|_{\widetilde{\mathbb{L}}^p_q(T)}\leqslant C\left(\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\lambda_0\|u\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\varphi\|_{2,p}\right), \tag{2.21}$$

where $C=C(\Theta,\alpha,q')>0$, $\beta=2-\alpha+\frac{2}{q'}-\frac{2}{q}>0$. Taking $q'=\infty$ in (2.21), then we have

$$(\lambda+\lambda_0)^{\frac{\beta}{2}}\sup_{t\in[0,T]}\|u(t)\|_p\leqslant C\left(\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\lambda_0\left(\int_0^T\|u(t)\|_p^q\mathrm{d}t\right)^{\frac{1}{q}}+\|\varphi\|_{2,p}\right).$$

Now it follows from the Gronwall lemma that

$$\sup_{t\in[0,T]}\|u(t)\|_p\leqslant C\left(\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\varphi\|_{2,p}\right), \tag{2.22}$$

where C depends on Θ,α,λ. Combining (2.21) and (2.22), we obtain for $1<\alpha<2-2/q$

$$\|u\|_{\widetilde{\mathbb{H}}^{\alpha,p}}+\|\partial_t u\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|u\|_{\widetilde{\mathbb{H}}^{2,p}_q(T)}\leqslant C\left(\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\varphi\|_{2,p}\right).$$

and complete the proof by Lemma 2.1. □

Next we consider the following second order elliptic PDE in $\mathbb{R}^d$:

$$a_{ij}\partial_i\partial_j u-\lambda u+b\cdot\nabla u=f, \tag{2.23}$$

where $\lambda\geqslant 0$, $a=(a_{ij}):\mathbb{R}^d\to\mathbb{R}^d\otimes\mathbb{R}^d$ is a symmetric matrix-valued Borel measurable function satisfying $(\mathbf{H}_a)$ and $b:\mathbb{R}^d\to\mathbb{R}^d$ is a vector-valued Borel measurable function. Firstly, we introduce the definition of a solution to PDE (2.23).

Definition 2.11. Let $p\in(1,\infty)$, $\lambda,T\geqslant 0$ and $b,f\in\widetilde{L}^p$. We call $u\in\widetilde{H}^{2,p}$ a *solution* of PDE (2.23) if for Lebesgue almost all $x\in\mathbb{R}^d$,

$$a_{ij}\partial_i\partial_j u(x)-\lambda u(x)+b(x)\cdot\nabla u(x)=f(x).$$

As a corollary of Lemma 2.8, we have the following results.

Lemma 2.12. *Assume $b\in\widetilde{L}^p$ for some $p>d$. Then there are constants $\lambda_0=\lambda_0(d,\|b\|_p,p,c_0,\theta)$ and $C=C(d,\|b\|_p,p,p',c_0,\theta)$ such that for any $\lambda\geqslant\lambda_0$ and $f\in\widetilde{L}^p$, there exists a unique solution u to PDE (2.23) in the sense of Definition 2.16 such that*

$$\lambda^{\frac{\beta}{2}}\|u\|_{\widetilde{H}^{\alpha,p'}}+\|\nabla^2 u\|_p\leqslant C\|f\|_p, \tag{2.24}$$

where $\alpha\in[0,2),p\in[p,\infty],p'\in[p,\infty]$ with $\beta:=2-\alpha+\frac{d}{p'}-\frac{d}{p}>0$.

Proof. As usual, it suffices to show the a priori estimate (2.24). Let $T>0$, u be a solution to (2.23) and ϕ be a nonnegative and nonzero smooth function on $[0,\infty)$ with $\phi(0)=0$. Define $\tilde u(t,x):=\phi(t)u(x)$. Then, one sees that $\tilde u$ is a solution to the following parabolic equation in the sense of Definition 2.6:

$$\partial_t\tilde u = a_{ij}\partial_i\partial_j\tilde u-\lambda\tilde u+b\cdot\nabla\tilde u-\phi f+\phi' u,\quad \tilde u(0)=0.$$

By (2.19), we have for any $\alpha\in[0,2), p'\in[p,\infty]$ with $\beta:=2-\alpha+\frac{d}{p'}-\frac{d}{p}>0$,

$$\lambda^{\frac{\beta}{2}}\|\tilde u\|_{\widetilde{\mathbb{H}}^{\alpha,p'}(T)}+\|\nabla^2\tilde u\|_{\widetilde{\mathbb{L}}^p(T)}\leqslant C\|\phi' u-\phi f\|_{\widetilde{\mathbb{L}}^p(T)},$$

which implies that

$$\lambda^{\frac{\beta}{2}}\|u\|_{\widetilde{H}^{\alpha,p'}}+\|\nabla^2 u\|_p\leqslant C\|\phi\|_\infty^{-1}\left(\|\phi'\|_\infty\|u\|_p+\|\phi\|_\infty\|f\|_p\right),$$

where $\|\phi\|_\infty:=\sup_{t\in[0,T]}|\phi(t)|$. Noting that $\|u\|_p\leqslant\|u\|_{p'}$, we obtain (2.24) and complete the proof. □

2.2.2 Localized mixed L^p spaces

In this part we show the well-posedness to the PDE (2.17) with drifts in mixed $\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}$-space. For $t>0$, let $P_tf(x)=\mathbb{E}f(x+\sqrt{2}W_t)$ be the Gaussian heat semigroup, i.e.,

$$P_tf(x)=\int_{\mathbb{R}^d} g_t(x-y)f(y)\mathrm{d}y, \tag{2.25}$$

where

$$g_t(x):=(4\pi t)^{-\frac{d}{2}}\mathrm{e}^{-\frac{|x|^2}{4t}}.$$

First of all, we establish the following easy estimates about P_t.

Lemma 2.13. *(i) For any $\boldsymbol{p}\in(1,\infty)^d$, $T>0$ and $\beta\geqslant0$, there is a constant $C=C(T,\boldsymbol{p},\beta,d)>0$ such that for all $f\in\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}$ and $t\in(0,T]$,*

$$\|P_tf\|_{\mathcal{C}^\beta}\leqslant Ct^{-\frac{1}{2}(\beta+|\frac{1}{\boldsymbol{p}}|)}\|f\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}}. \tag{2.26}$$

(ii) For any $\boldsymbol{q}\geqslant\boldsymbol{p}$, there is a constant $C=C(\boldsymbol{q},\boldsymbol{p},d)>0$ such that for all $f\in\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}$ and $t>0$,

$$\|\nabla P_tf\|_{\mathbb{L}^{\boldsymbol{q}}_{\boldsymbol{\pi}}}\leqslant Ct^{-\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}}|-|\frac{1}{\boldsymbol{q}}|)}\|f\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}}. \tag{2.27}$$

Proof. (i) Note that for $m = 0, 1, \cdots$,

$$\nabla^m P_t f(x) = \int_{\mathbb{R}^d} \nabla^m g_t(x-y) f(y) \mathrm{d}y.$$

For $\frac{1}{\boldsymbol{q}} + \frac{1}{\boldsymbol{p}} = \mathbf{1}$, by Hölder's inequality (2.7) and the scaling, we have

$$\|\nabla^m P_t f\|_\infty \leqslant \|\nabla^m g_t\|_{\mathbb{L}^{\boldsymbol{q}}_{\boldsymbol{\pi}}} \|f\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} = t^{-\frac{1}{2}(m+|\frac{1}{\boldsymbol{p}}|)} \|\nabla^m g_1\|_{\mathbb{L}^{\boldsymbol{q}}_{\boldsymbol{\pi}}} \|f\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}},$$

where $\|\nabla^m g_1\|_{\mathbb{L}^{\boldsymbol{q}}_{\boldsymbol{\pi}}} < \infty$. Then estimate (2.26) follows by the interpolation theorem for Hölder spaces.

(ii) For $\boldsymbol{r} \in [1,\infty]^d$ with $\frac{1}{\boldsymbol{p}} + \frac{1}{\boldsymbol{r}} = \mathbf{1} + \frac{1}{\boldsymbol{q}}$, by Young's inequality (2.8) and the scaling, we have

$$\|\nabla P_t f\|_{\mathbb{L}^{\boldsymbol{q}}_{\boldsymbol{\pi}}} \leqslant \|\nabla g_t\|_{\mathbb{L}^{\boldsymbol{r}}_{\boldsymbol{\pi}}} \|f\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} = t^{-\frac{1}{2}(1+d-|\frac{1}{\boldsymbol{r}}|)} \|\nabla g_1\|_{\mathbb{L}^{\boldsymbol{r}}_{\boldsymbol{\pi}}} \|f\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}}.$$

Then estimate (2.27) follows because $\|\nabla g_1\|_{\mathbb{L}^{\boldsymbol{r}}_{\boldsymbol{\pi}}} < \infty$. □

We introduce the following index sets for later use:

$$\mathscr{I}^o := \Big\{(q,\boldsymbol{p}) \in (2,\infty)^{1+d} : \ |\tfrac{1}{\boldsymbol{p}}| + \tfrac{2}{q} < 1\Big\}, \tag{2.28}$$

and

$$\mathscr{I}_m := \Big\{(q,\boldsymbol{p}) \in (1,\infty)^{1+d} : \ |\tfrac{1}{\boldsymbol{p}}| + \tfrac{2}{q} < m\Big\}, \ m = 1, 2. \tag{2.29}$$

Remark 2.14. We note that $\mathscr{I}^o \subset \mathscr{I}_1$. For $(q,\boldsymbol{p}) \in \mathscr{I}^o$, it holds that $(\frac{q}{2}, \frac{\boldsymbol{p}}{2}) \in \mathscr{I}_2$.

For $\lambda \geqslant 0$ and $f \in \mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})$, we define

$$u(t,x) := \int_0^t \mathrm{e}^{-\lambda(t-s)} P_{t-s} f(s,x) \mathrm{d}s, \ t > 0,$$

which solves the following non-homeogenous heat equation

$$\partial_t u = \tfrac{1}{2}\Delta u - \lambda u + f, \quad u(0) = 0.$$

Lemma 2.15. *(i) For any $T > 0$, $(q,\boldsymbol{p}) \in \mathscr{I}_2$ and $\beta \in [0, 2 - |\frac{1}{\boldsymbol{p}}| - \frac{2}{q})$, there is a constant $C = C(T,d,q,\boldsymbol{p},\beta) > 0$ such that for all $\lambda \geqslant 0$,*

$$(1 \vee \lambda)^{\frac{1}{2}(2-\beta-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})} \|u\|_{\mathbb{L}^\infty_T(\mathcal{C}^\beta)} \leqslant C\|f\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{2.30}$$

(ii) For any $T > 0$, $(q,\boldsymbol{p}) \in \mathscr{I}_2$ and $(q',\boldsymbol{p}') \geqslant (q,\boldsymbol{p})$ with $|\frac{1}{\boldsymbol{p}}| + \frac{2}{q} < |\frac{1}{\boldsymbol{p}'}| + \frac{2}{q'} + 1$, there is a constant $C = C(T,d,q,\boldsymbol{p},q',\boldsymbol{p}') > 0$ such that for all $\lambda \geqslant 0$,

$$(1 \vee \lambda)^{\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}'}|+\frac{2}{q'}-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})} \|\nabla u\|_{\mathbb{L}^{q'}_T(\mathbb{L}^{\boldsymbol{p}'}_{\boldsymbol{\pi}})} \leqslant C\|f\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{2.31}$$

(iii) *For any $T > 0$, $(q, \boldsymbol{p}) \in \mathscr{I}_1$ and $\lambda \geqslant 0$, there is a constant $C = C(\lambda, T, d, q, \boldsymbol{p}) > 0$ such that for all $0 \leqslant t_0 < t_1 \leqslant T$,*

$$\|u(t_1) - u(t_0)\|_\infty \leqslant C(t_1 - t_0)^{\frac{1}{2}} \|f\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{2.32}$$

Proof. (i) For $\beta \in [0, 2 - |\frac{1}{\boldsymbol{p}}| - \frac{2}{q})$, by (2.26) and Hölder's inequality in the time variable, we have

$$\begin{aligned}
\|u(t)\|_{\mathcal{C}^\beta} &\lesssim \int_0^t \mathrm{e}^{-\lambda(t-s)}(t-s)^{-\frac{1}{2}(\beta+|\frac{1}{\boldsymbol{p}}|)} \|f(s)\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} \mathrm{d}s \\
&\leqslant \left(\int_0^t \left(\mathrm{e}^{-\lambda s} s^{-\frac{1}{2}(\beta+|\frac{1}{\boldsymbol{p}}|)}\right)^{\frac{q}{q-1}} \mathrm{d}s \right)^{1-\frac{1}{q}} \|f\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \\
&\lesssim (1 \vee \lambda)^{-\frac{1}{2}(2-\beta-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})} \|f\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}.
\end{aligned}$$

(ii) For $(q', \boldsymbol{p}') \geqslant (q, \boldsymbol{p})$ with $|\frac{1}{\boldsymbol{p}}| + \frac{2}{q} < |\frac{1}{\boldsymbol{p}'}| + \frac{2}{q'} + 1$, by (2.27) we have

$$\|\nabla u(t)\|_{\mathbb{L}^{\boldsymbol{p}'}_{\boldsymbol{\pi}}} \lesssim \int_0^t \mathrm{e}^{-\lambda(t-s)}(t-s)^{-\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}}|-|\frac{1}{\boldsymbol{p}'}|)} \|f(s)\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} \mathrm{d}s.$$

Let $r \geqslant 1$ be defined by $\frac{1}{r} = \frac{1}{q'} + 1 - \frac{1}{q}$. By Young's inequality we further have

$$\begin{aligned}
\|\nabla u\|_{\mathbb{L}^{q'}_T(\mathbb{L}^{\boldsymbol{p}'}_{\boldsymbol{\pi}})} &\lesssim \left(\int_0^T \mathrm{e}^{-r\lambda s} s^{-\frac{r}{2}(1+|\frac{1}{\boldsymbol{p}}|-|\frac{1}{\boldsymbol{p}'}|)} \mathrm{d}s \right)^{1/r} \|f\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \\
&\lesssim (1 \wedge \lambda)^{\frac{1}{r}-\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}}|-|\frac{1}{\boldsymbol{p}'}|)} \|f\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}.
\end{aligned}$$

(iii) For $0 \leqslant t_0 < t_1 \leqslant T$, by definition we have

$$\begin{aligned}
u(t_1) - u(t_0) &= \int_0^{t_0} \mathrm{e}^{-\lambda(t_1-s)}(P_{t_1-s} - P_{t_0-s}) f(s,x) \mathrm{d}s \\
&\quad + (\mathrm{e}^{-\lambda(t_1-t_0)} - 1) \int_0^{t_0} \mathrm{e}^{-\lambda(t_0-s)} P_{t_0-s} f(s,x) \mathrm{d}s \\
&\quad + \int_{t_0}^{t_1} \mathrm{e}^{-\lambda(t_1-s)} P_{t_1-s} f(s,x) \mathrm{d}s \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

For I_1, noting that

$$\|P_t f - f\|_\infty \leqslant \frac{1}{2} \int_0^t \|\Delta P_s f\|_\infty \mathrm{d}s \lesssim \left(\int_0^t s^{-\frac{1}{2}} \mathrm{d}s \right) \|\nabla f\|_\infty \lesssim t^{\frac{1}{2}} \|\nabla f\|_\infty,$$

by (2.26) and Hölder's inequality, we have

$$\begin{aligned}\|I_1\|_\infty &\lesssim (t_1-t_0)^{\frac{1}{2}}\int_0^{t_0}\|\nabla P_{t_0-s}f(s)\|_\infty \mathrm{d}s\\ &\lesssim (t_1-t_0)^{\frac{1}{2}}\int_0^{t_0}(t_0-s)^{-\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}}|)}\|f(s)\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}}\mathrm{d}s\\ &\lesssim (t_1-t_0)^{\frac{1}{2}}t_0^{\frac{1}{2}(1-\frac{2}{q}-|\frac{1}{\boldsymbol{p}}|)}\|f\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})},\end{aligned}$$

and because $1-\mathrm{e}^{-\lambda(t_1-t_0)}\leqslant \lambda(t_1-t_0)$,

$$\|I_2\|_\infty \lesssim \lambda(t_1-t_0)\|f\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}.$$

For I_3, as above, by (2.26) and Hölder's inequality, we have

$$\|I_3\|_\infty \lesssim \left(\int_0^{t_1-t_0}\left(\mathrm{e}^{-\lambda s}s^{-\frac{1}{2}|\frac{1}{\boldsymbol{p}}|}\right)^{\frac{q}{q-1}}\mathrm{d}s\right)^{1-\frac{1}{q}}\|f\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}\leqslant (t_1-t_0)^{1-\frac{1}{2}(\frac{2}{q}+|\frac{1}{\boldsymbol{p}}|)}\|f\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}.$$

Combining the above estimates and because $\frac{2}{q}+|\frac{1}{\boldsymbol{p}}|<1$, we obtain (2.32). □

Now we shall study the following second order parabolic PDE in $\mathbb{R}_+\times\mathbb{R}^d$:

$$\partial_t u=\mathrm{tr}(a\cdot\nabla^2 u)+b\cdot\nabla u-\lambda u+f,\quad u(0)=0,\tag{2.33}$$

where $\lambda\geqslant 0$, $a:=\sigma\sigma^*/2$, σ satisfies $(\mathbf{A}^\sigma)$ and

$$b,f\in L^1_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R}^d).$$

We introduce the following notion of solutions to PDE (2.33).

Definition 2.16. *Let $T>0$ and $\mathscr{U}_T\subset L^1_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R}^d)$ be some subclass of locally integrable functions. We call $u\in\mathscr{U}_T$ a solution of PDE (2.33) if for all $t\in[0,T]$ and $\varphi\in C_c(\mathbb{R}^d)$,*

$$\langle u(t),\varphi\rangle=\int_0^t\langle \mathrm{tr}(a\cdot\nabla^2 u)+b\cdot\nabla u,\varphi\rangle\mathrm{d}s-\lambda\int_0^t\langle u,\varphi\rangle\mathrm{d}s+\int_0^t\langle f,\varphi\rangle\mathrm{d}s,$$

where we have implicitly assumed that $\nabla^2 u\in L^1_{\mathrm{loc}}$ and $\nabla u\in L^\infty_{\mathrm{loc}}$ so that the terms on the right hand side are well defined. Here $\mathscr{U}_T$ will be specified below in the respective cases.

We first show the following result for bounded drift b (see [74, Theorem 2.1]).

Theorem 2.17. *Let $T>0$ and $(q,\boldsymbol{p})\in(1,\infty)^{1+d}$. Suppose that $(\mathbf{A}^\sigma)$ holds and b is bounded measurable. Then for any $f\in\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})$ and $\beta\in[0,2-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})$, there exists a unique solution $u\in\mathscr{U}_T$ in the sense of Definition 2.16, where $\mathscr{U}_T$ consists of all u with*

$$(1\vee\lambda)^{\frac{1}{2}(2-\beta-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})}\|u\|_{\mathbb{L}^\infty_T(\mathcal{C}^\beta)}+\|\nabla^2 u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}\leqslant C\|f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}.\tag{2.34}$$

Here and below, the constant $C = C(T, \kappa_0, d, \boldsymbol{p}, q, \beta, \|b\|_{\mathbb{L}^\infty_T}) > 0$ *is independent of* λ. *Moreover, for any* $(q', \boldsymbol{p}') \geqslant (q, \boldsymbol{p})$ *with* $|\frac{1}{\boldsymbol{p}}| + \frac{2}{q} < |\frac{1}{\boldsymbol{p}'}| + \frac{2}{q'} + 1$, *we also have*

$$(1 \vee \lambda)^{\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}'}|+\frac{2}{q'}-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})} \|\nabla u\|_{\widetilde{\mathbb{L}}^{q'}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}'}_{\boldsymbol{\pi}})} \leqslant C \|f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}, \tag{2.35}$$

and for all $0 \leqslant t_0 \leqslant t_1 \leqslant T$,

$$\|u(t_1) - u(t_0)\|_\infty \leqslant C(\lambda)(t_1 - t_0)^{\frac{1}{2}} \|f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{2.36}$$

Proof. We only prove the a priori estimates (2.34), (2.35) and (2.36). The existence is then standard by mollifying the coefficients and a compactness argument. Fix $r > 0$. Let χ^r_z be the cutoff function in (2.1) and $w_z := u\chi^r_z$. It is easy to see that

$$\partial_t w_z = \mathrm{tr}(a \cdot \nabla^2 w_z) - \lambda w_z + g_z, \quad w_z(0) = 0, \tag{2.37}$$

where

$$g_z := \mathrm{tr}(a \cdot \nabla^2 u)\chi^r_z - \mathrm{tr}(a \cdot \nabla^2 w_z) + (b \cdot \nabla u)\chi^r_z + f\chi^r_z.$$

Let $(q, \boldsymbol{p}) \in (1, \infty)^{1+d}$. By [74, Theorem 2.1], there is a constant $C = C(T, \kappa_0, d, \boldsymbol{p}, q) > 0$ such that

$$\|w_z\|_{\mathbb{L}^\infty_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|\nabla^2 w_z\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \lesssim_C \|g_z\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{2.38}$$

On the other hand, we can write (2.37) as

$$\partial_t w_z = \Delta w_z - \lambda w_z + \mathrm{tr}((a - \mathbb{I}) \cdot \nabla^2 w_z) + g_z, \quad w_z(0) = 0,$$

and by Duhamel's formula,

$$w_z(t, x) = \int_0^t \mathrm{e}^{-\lambda(t-s)} P_{t-s}(\mathrm{tr}((a - \mathbb{I}) \cdot \nabla^2 w_z) + g_z)(s, x)\mathrm{d}s.$$

Note that by (2.38),

$$\|\mathrm{tr}((a - \mathbb{I}) \cdot \nabla^2 w_z) + g_z\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \lesssim \|g_z\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{2.39}$$

For $\beta \in [0, 2 - |\frac{1}{\boldsymbol{p}}| - \frac{2}{q})$, by (2.30) and (2.39) we have

$$(1 \vee \lambda)^{\frac{1}{2}(2-\beta-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})} \|w_z\|_{\mathbb{L}^\infty_T(\mathcal{C}^\beta)} \lesssim \|g_z\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{2.40}$$

For $(q', \boldsymbol{p}') \geqslant (q, \boldsymbol{p})$ with $|\frac{1}{\boldsymbol{p}}| + \frac{2}{q} < |\frac{1}{\boldsymbol{p}'}| + \frac{2}{q'} + 1$, by (2.31) and (2.39) we have

$$(1 \vee \lambda)^{\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}'}|+\frac{2}{q'}-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})} \|\nabla w_z\|_{\mathbb{L}^{q'}_T(\mathbb{L}^{\boldsymbol{p}'}_{\boldsymbol{\pi}})} \lesssim \|g_z\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{2.41}$$

For $0 \leqslant t_0 < t_1 \leqslant T$, by (2.32) and (2.39) we have

$$\|w_z(t_1) - w_z(t_0)\|_\infty \lesssim (t_1 - t_0)^{\frac{1}{2}} \|g_z\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{2.42}$$

Since $\chi_z^{2r} \nabla^j \chi_z^r = \nabla^j \chi_z^r$ for $j = 0, 1, 2$, we have

$$\begin{aligned}\|g_z\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} &\lesssim \|\nabla u \nabla \chi_z^r\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|u \nabla^2 \chi_z^r\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|b\|_{\mathbb{L}^\infty_T} \|\nabla u \chi_z^r\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \\ &\leqslant (\|\nabla \chi_z^r\|_\infty + \|b\|_{\mathbb{L}^\infty_T}) \|\nabla u \chi_z^{2r}\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|\nabla^2 \chi_z^r\|_\infty \|u \chi_z^{2r}\|_{\mathbb{L}^q_T(\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}.\end{aligned}$$

Substituting this into (2.38), (2.40), (2.41) and (2.42) and taking supremum in $z \in \mathbb{R}^d$, we obtain

$$\|u\|_{\widetilde{\mathbb{L}}^\infty_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|\nabla^2 u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \lesssim \|f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|\nabla u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}, \tag{2.43}$$

and for $\beta \in [0, 2 - |\frac{1}{\boldsymbol{p}}| - \frac{2}{q})$,

$$(1 \wedge \lambda)^{\frac{1}{2}(2-\beta-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})} \|u\|_{\mathbb{L}^\infty_T(\mathcal{C}^\beta)} \lesssim \|f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|\nabla u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}, \tag{2.44}$$

and for $(q', \boldsymbol{p}') \geqslant (q, \boldsymbol{p})$ with $|\frac{1}{\boldsymbol{p}}| + \frac{2}{q} < |\frac{1}{\boldsymbol{p}'}| + \frac{2}{q'} + 1$,

$$(1 \vee \lambda)^{\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}'}|+\frac{2}{q'}-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})} \|\nabla u\|_{\widetilde{\mathbb{L}}^{q'}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}'}_{\boldsymbol{\pi}})} \lesssim \|f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|\nabla u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}, \tag{2.45}$$

and for all $0 \leqslant t_0 < t_1 \leqslant T$,

$$\|u(t_1) - u(t_0)\|_\infty \lesssim (t_1 - t_0)^{\frac{1}{2}} \Big(\|f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|\nabla u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \Big). \tag{2.46}$$

Note that by the interpolation inequality, for any $\varepsilon \in (0, 1)$,

$$\|\nabla u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \leqslant \varepsilon \|\nabla^2 u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + C_\varepsilon \|u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}.$$

Substituting this into (2.43) and choosing ε small enough, we derive that for any $t \in [0, T]$,

$$\|u(t)\|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} + \|\nabla^2 u\|_{\widetilde{\mathbb{L}}^q_t(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \lesssim \|f\|_{\widetilde{\mathbb{L}}^q_t(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \left(\int_0^t \|u(s)\|^q_{\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} \mathrm{d}s \right)^{1/q}.$$

By Gronwall's inequality, we get

$$\|u\|_{\mathbb{L}^\infty_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \|\nabla^2 u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \leqslant C \|f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})},$$

which together with (2.44), (2.45) and (2.46) yields (2.34), (2.35) and (2.36). □

Remark 2.18. For any $T,\gamma>0$ and $(q,\boldsymbol{p})\in\mathscr{I}_2$, there is a $C=C(T,\gamma,d,q,\boldsymbol{p})>0$ such that

$$\sup_x \mathbb{E}\left(\int_0^T h(s,x+W_{\gamma s})\mathrm{d}s\right)\leqslant C\|h\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{2.47}$$

Indeed, let $a=\sqrt{\gamma/2}\mathbb{I}$, $b=0$, $\lambda=0$ and $f(s,x)=h(T-s,x)$ in PDE (2.33). By (2.34) we have

$$\mathbb{E}\left(\int_0^T h(s,x+W_{\gamma s})\mathrm{d}s\right)=\int_0^T P_{\gamma(T-s)}f(s,x)\mathrm{d}s=u(T,x)\lesssim\|f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}=\|h\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}.$$

In particular, once we have the Gaussian type density estimate for SDEs, then by (2.47), we immediately have the Krylov estimate as we shall see in Theorem 3.14 below.

Next we consider the drift b being in the mixed L^p-space, where each component b_i may lie in a different mixed L^p-space. Thus the second order generalized derivative of u stays in a direct sum space of mixed L^p-spaces. To solve (2.33), we consider the followng systems:

$$\begin{cases}\partial_t u_0=\mathrm{tr}(a\cdot\nabla^2 u_0)-\lambda u_0+f, \quad u_0(0)=0,\\ \partial_t u_i=\mathrm{tr}(a\cdot\nabla^2 u_i)+b_i\cdot\partial_i(\sum_{j=0}^d u_j)-\lambda u_i, \quad u_i(0)=0, \quad i=1,..,d.\end{cases} \tag{2.48}$$

Clearly, $u:=\sum_{j=0}^d u_j$ solves the original scalar PDE (2.33). The following result seems to be new and is the cornerstone of studying SDEs with singular mixed L^p-coefficients.

Theorem 2.19. *Let $T>0$. Suppose $(\mathbf{H}_a)$ and for some $(q_i,\boldsymbol{p}_i)\in\mathscr{I}_1$ and $\boldsymbol{\pi}_i\in S_d$, $i=1,\cdots,d$,*

$$\|b_1\|_{\widetilde{\mathbb{L}}^{q_1}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_1}_{\boldsymbol{\pi}_1})}+\cdots+\|b_d\|_{\widetilde{\mathbb{L}}^{q_d}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_d}_{\boldsymbol{\pi}_d})}\leqslant\kappa_1<\infty. \tag{2.49}$$

Let $\boldsymbol{\pi}_0\in S_d$ and $(q_0,\boldsymbol{p}_0)\in\mathscr{I}_1$. Define

$$\vartheta:=1-\max_{i=0,\cdots,d}(|\tfrac{1}{\boldsymbol{p}_i}|+\tfrac{2}{q_i}). \tag{2.50}$$

For any $f\in\widetilde{\mathbb{L}}^{q_0}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_0}_{\boldsymbol{\pi}_0})$ and $\beta\in[0,\vartheta)$, there is a constant $C_0=C_0(T,\kappa_0,d,\boldsymbol{p}_i,q_i,\beta)\geqslant1$ so that for all $\lambda\geqslant C_0\kappa_1^{2/\vartheta}$, there exists a unique solution $u\in\mathscr{U}_T$ to PDE (2.33) in the sense of Definition 2.16, where $\mathscr{U}_T$ consists of all $u=u_0+u_1+\cdots+u_d$ with $(u_0,u_1,...,u_d)$ solves (2.48) and

$$\lambda^{\frac12(\vartheta-\beta)}\|u\|_{\mathbb{L}^\infty_T(\mathcal{C}^{1+\beta})}+\sum_{i=0}^d\|\nabla^2 u_i\|_{\widetilde{\mathbb{L}}^{q_i}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i})}\leqslant C_1\|f\|_{\widetilde{\mathbb{L}}^{q_0}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_0}_{\boldsymbol{\pi}_0})}, \tag{2.51}$$

where $C_1=C_1(T,\kappa_0,d,\boldsymbol{p}_i,q_i,\beta)>0$ is independent of λ and κ_1. Moreover, for all $0\leqslant t_0\leqslant t_1\leqslant T$,

$$\|u(t_1)-u(t_0)\|_\infty\leqslant C(\lambda)(t_1-t_0)^{\frac12}\|f\|_{\widetilde{\mathbb{L}}^{q_0}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_0}_{\boldsymbol{\pi}_0})}. \tag{2.52}$$

Proof. Again we only show the a priori estimate (2.51) since then the existence can be shown by a compactness argument. Let $u = u_0 + u_1 + \cdots + u_d$, where $(u_0, u_1, ..., u_d)$ solves (2.48). Let $\lambda \geqslant 1$ and $\beta \in [0, \vartheta)$ with ϑ being defined by (2.50). By Theorem 2.17 with $b = 0$, we have

$$\lambda^{\frac{1}{2}(1-|\frac{1}{\boldsymbol{p}_0}|-\frac{2}{q_0}-\beta)}\|u_0\|_{\mathbb{L}^\infty_T(\mathcal{C}^{1+\beta})} + \|\nabla^2 u_0\|_{\widetilde{\mathbb{L}}^{q_0}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_0}_{\boldsymbol{\pi}_0})} \lesssim \|f\|_{\widetilde{\mathbb{L}}^{q_0}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_0}_{\boldsymbol{\pi}_0})},$$

and

$$\|u_0(t_1) - u_0(t_0)\|_\infty \leqslant C(\lambda)(t_1 - t_0)^{\frac{1}{2}}\|f\|_{\widetilde{\mathbb{L}}^{q_0}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_0}_{\boldsymbol{\pi}_0})},$$

and for each $i = 1, \cdots, d$,

$$\lambda^{\frac{1}{2}(1-|\frac{1}{\boldsymbol{p}_i}|-\frac{2}{q_i}-\beta)}\|u_i\|_{\mathbb{L}^\infty_T(\mathcal{C}^{1+\beta})} + \|\nabla^2 u_i\|_{\widetilde{\mathbb{L}}^{q_i}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i})} \lesssim \|b_i \cdot \partial_i u\|_{\widetilde{\mathbb{L}}^{q_i}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i})} \lesssim \|b_i\|_{\widetilde{\mathbb{L}}^{q_i}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i})}\|\partial_i u\|_{\mathbb{L}^\infty_T},$$

and

$$\|u_i(t_1) - u_i(t_0)\|_\infty \leqslant C(\lambda)(t_1 - t_0)^{\frac{1}{2}}\|b_i\|_{\widetilde{\mathbb{L}}^{q_i}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i})}\|\partial_i u\|_{\mathbb{L}^\infty_T}.$$

Summing up the above inequalities for i from 0 to d, we obtain

$$\lambda^{\frac{1}{2}(\vartheta-\beta)}\|u\|_{\mathbb{L}^\infty_T(\mathcal{C}^{1+\beta})} + \sum_{i=0}^{d}\|\nabla^2 u_i\|_{\widetilde{\mathbb{L}}^{q_i}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i})} \leqslant C_1\|f\|_{\widetilde{\mathbb{L}}^{q_0}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_0}_{\boldsymbol{\pi}_0})} + C_2\kappa_1\|\nabla u\|_{\mathbb{L}^\infty_T},$$

where C_1, C_2 only depend on $T, \kappa_0, d, \boldsymbol{p}_i, q_i, \beta$, and

$$\|u(t_1) - u(t_0)\|_\infty \leqslant C(\lambda)(t_1 - t_0)^{\frac{1}{2}}\Big(\kappa_1\|\nabla u\|_{\mathbb{L}^\infty_T} + \|f\|_{\widetilde{\mathbb{L}}^{q_0}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_0}_{\boldsymbol{\pi}_0})}\Big).$$

Choosing $C_0 = (C_2/2)^{2/\vartheta} \vee 1$, we obtain (2.51) and (2.52) for all $\lambda \geqslant C_0\kappa_1^{2/\vartheta}$. □

2.3 Strong solutions, weak solutions and martingale solutions to SDEs and DDSDEs

In order to make the definition of solutions to the SDEs clear, we recall some classical terminology. Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathbb{P})$ be a probability space equipped with a filtration. We consider the following SDE:

$$X_t = X_0 + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}W_s, \tag{2.53}$$

where $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ are measurable and $(W_t)_{t\geqslant 0}$ is a standard d-dimensional Brownian motion. Then we call $(X_t)_{t\geqslant 0}$

- a *weak solution* to SDE (2.53), if one can construct a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathbb{P})$ and a adapted d-dimensional Brownian motion $(W_t)_{t\geqslant 0}$ under $\mathbb{P}$ such that (2.53) holds $\mathbb{P}$-a.s. and

$$\int_0^t |b(s, X_s)|\mathrm{d}s + \int_0^t \|\sigma(s, X_s)\|_{HS}^2 \mathrm{d}s < \infty \quad \mathbb{P}\text{-a.s.} \tag{2.54}$$

- a *strong solution* if it is a weak solution and $(X_t)_{t\geqslant 0}$ is adapted to the Brownian fitration $\mathscr{F}_t^W := \sigma(W_s, s \in [0,t])$.

Moreover, we call

- *uniqueness in law* holds for SDE (2.53) if every weak solution X to SDE (2.53), possibly on different probability spaces, has the same law in $\mathcal{P}(\mathbb{C}_T)$ for any $T > 0$;

- *pathwise uniqueness* holds for SDE (2.53), if on any given filtered probability space with any given Brownian motion, any two weak solutions to SDE (2.53) with the same initial data $X_0 \in \mathscr{F}_0$ coincide $\mathbb{P}$-a.s. in the path space $C(\mathbb{R}_+; \mathbb{R}^d)$.

We also call a

- *weak well-posedness* holds for SDE (2.53), if there is a unique weak solution in the sense of uniqueness in law;

- *strong well-posedness* holds for SDE (2.53), if there is a unique strong solution in the sense of pathwise uniqueness.

It is well-known that

- when weak well-posedness holds for SDE (2.53) with any initial data x, then weak well-posedness holds for SDE (2.53) with any initial data $X_0 \in \mathscr{F}_0$ (see [86, Proposition 1.4, page 367]). If we denote by $\mathbb{P}(x)$ and $\mathbb{P}$ the law of the unique solution to SDE (2.53) in $\mathcal{P}(C(\mathbb{R}_+; \mathbb{R}^d))$ when $X_0 = x$ and $X_0 = \xi \in \mathscr{F}_0$ respectively, then

$$\mathbb{P} = \int_{\mathbb{R}^d} \mathbb{P}(x)\mathbb{P} \circ (\xi)^{-1}(\mathrm{d}x).$$

- (Yamada-Watanabe's theorem) if pathwise uniqueness holds, then uniqueness in law holds and every weak solution is strong (see [86, Theorem 1.7, page 368]). Also, see [86, Exercise 1.20, page 375] for a converse version.

- when strong well-posedness holds for SDE (2.53) with any initial data x. There is a measurable function Φ on $\mathbb{R}^d \times C(\mathbb{R}_+; \mathbb{R}^d)$ such that for any filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathbb{P})$ thereon a $X_0 \in \mathscr{F}_0$ and a adapted d-dimensional Brownian motion $(W_t)_{t\geqslant 0}$,

$$X_t := \Phi(X_0, W_t)$$

is the unique strong solution to SDE (2.53) (see [86, Remark 2, page 369]).

Next, we consider the following distribution density-distribution dependent SDE (abbreviated as dDDSDE):

$$\mathrm{d}X_t = b(t, X_t, \rho_t(X_t), \mu_{X_t})\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \tag{2.55}$$

where $\rho_t(x)$ is the density of X_t and $b(t, x, r, \mu) : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ is a measurable function. The following is the definition of weak solution to dDDSDE (2.55).

Definition 2.20. *Let $\mathfrak{U} := (\Omega, \mathscr{F}, \mathbb{P}, (\mathscr{F}_t)_{t\geqslant 0})$ be a stochastic basis and (X, W) be a pair of continuous $\mathscr{F}_t$-adapted processes. Let $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$. We call $(X, W, \mathfrak{U})$ a solution of dDDSDE (2.55) with initial distribution μ_0 if*

(i) $\mu_0 = \mathbb{P} \circ X_0^{-1}$ and W is a standard Brownian motion on $\mathfrak{U}$.

(ii) For each $t > 0$, the distribution μ_{X_t} of X_t admits a density ρ_t,

$$\int_0^t |b(s, X_s, \rho_s(X_s), \mu_{X_s})|\mathrm{d}s + \int_0^t |\sigma(s, X_s)|^2\mathrm{d}s < \infty, \ \ a.s.,$$

and

$$X_t = X_0 + \int_0^t b(s, X_s, \rho_s(X_s), \mu_{X_s})\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}W_s, \ \ a.s.$$

We call $(X_t)_{t\geqslant 0}$ a strong solution to dDDSDE (2.55), if it is a weak solution to dDDSDE (2.55), and for $b(t, x) := b(t, x, \rho_t(x), \mu_{X_t})$, $(X_t)_{t\geqslant 0}$ is a strong solution to SDE (2.53). We also say *weak well-posedness* and *strong well-posedness* for dDDSDE (2.55) when we fix $b(t, x) := b(t, x, \rho_t(x), \mu_{X_t})$ and regard it as SDE (2.53).

Let $T > 0$ and $\mathbb{C}_T$ be the space of all continuous functions from $[0, T]$ to $\mathbb{R}^d$. We use ω to denote a path in $\mathbb{C}_T$ and by $w_t(\omega) = \omega_t$ to denote the coordinate process. Let $\mathcal{B}_t := \sigma\{w_s, s \leqslant t\}$ be the natural filtration. We also introduce the following notion of martingale solutions to dDDSDE (2.55).

Definition 2.21. *Let $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$. A probability measure $\mathbb{P} \in \mathcal{P}(\mathbb{C}_T)$ is called a martingale solution of dDDSDE (2.55) with initial distribution μ_0 if $\mathbb{P} \circ w_0^{-1} = \mu_0$ and for any $f \in C_0^2(\mathbb{R}^d)$, the process*

$$M_t^f(\omega) := f(w_t) - f(w_0) - \int_0^t \Big(\tfrac{1}{2}\mathrm{tr}((\sigma\sigma^*)(s, w_s) \cdot \nabla^2) + b(s, w_s, \rho_s(w_s), \mu_s) \cdot \nabla\Big) f(w_s)\mathrm{d}s \tag{2.56}$$

is a $\mathcal{B}_t$-martingale, where $\mu_s := \mathbb{P} \circ w_s^{-1}$ has a density $\rho_s(x)$. We shall use $\mathscr{M}_{\mu_0}^{\sigma,b}$ to denote the set of all martingale solutions of dDDSDE (2.55) associated with σ, b and initial distribution μ_0.

Remark 2.22. It is well known that weak solutions are equivalent to the martingale solutions (see [97]). If we consider the classical SDE, i.e., b only depends on (t, x), and if for each starting point (s, x), there is a unique martingale solution starting from (s, x), then as usual, we say the martingale problem is well-posed.

2.4 Some well-known results about relative entropy

In this section we recall the notion and some basic facts about the relative entropy. Let E be a Polish space and μ, ν be two probability measures on E. The relative entropy $\mathcal{H}(\mu|\nu)$ is defined by

$$\mathcal{H}(\mu|\nu) := \begin{cases} \int_E \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right) \mathrm{d}\nu, & \mu \ll \nu, \\ \infty, & \text{otherwise.} \end{cases} \tag{2.57}$$

Since $x \mapsto x \log x$ is convex on $[0, \infty)$, by Jensen's inequality, we have $\mathcal{H}(\mu|\nu) \geqslant 0$. Here we give a brief proof to the following data processing inequality.

Lemma 2.23. *Let E, F be two Polish spaces. Let $\mu, \nu \in \mathcal{P}(E)$ and $X : E \to F$ be measurable. Then, we have*

$$\mathcal{H}(\nu \circ X^{-1} | \mu \circ X^{-1}) \leqslant \mathcal{H}(\nu|\mu).$$

Proof. Without loss of generality, we assume that $\mathcal{H}(\nu|\mu) = \infty$ and $\nu \ll \mu$. First of all, it is easy to see that $\nu \ll \mu$ implies $\nu \circ X^{-1} \ll \mu \circ X^{-1}$. We set

$$f(x) := \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x), \ x \in E \quad \text{and} \quad g(y) := \frac{\mathrm{d}\nu \circ X^{-1}}{\mathrm{d}\mu \circ X^{-1}}(y), \ y \in F.$$

By definition, for any bounded measurable funtion φ on F, we have

$$\begin{aligned} \int_E \varphi(X(x)) g(X(x)) \mu(\mathrm{d}x) &= \int_F \varphi(y) g(y) \mu \circ X^{-1}(\mathrm{d}y) = \int_F \varphi(y) \nu \circ X^{-1}(\mathrm{d}y) \\ &= \int_E \varphi(X(x)) \nu(\mathrm{d}x) = \int_E \varphi(X(x)) f(x) \mu(\mathrm{d}x), \end{aligned}$$

which implies that $g \circ X = \mathbb{E}_\mu[f|\sigma(X)]$ and by Jensen's inequality for conditional expectation that

$$g(X(x)) \log g(X(x)) \leqslant \mathbb{E}_\mu[f \log f|\sigma(X)](x), \quad \mu\text{-a.e.}$$

Therefore, we have

$$\begin{aligned} \mathcal{H}(\nu \circ X^{-1}|\mu \circ X^{-1}) &= \int_F g(y) \log g(y) \mu \circ X^{-1}(\mathrm{d}y) = \int_E g(X(x)) \log g(X(x)) \mu(\mathrm{d}x) \\ &\leqslant \int_E f(x) \log f(x) \mu(\mathrm{d}x) = \mathcal{H}(\mu|\nu) \end{aligned}$$

and complete the proof. □

The following theorem contains the main tools used below (see [12, Theorem 2.1(ii)], [31, Lemma 1.4.3(a)] and [29, Lemma 3.9]).

Theorem 2.24. *(i) (Pinsker's inequality) For any two probability measures μ, ν,*

$$\|\mu-\nu\|_{\mathrm{var}}^2 \leqslant 2\mathcal{H}(\mu|\nu). \tag{2.58}$$

(ii) (The weighted Pinsker inequality) For any $\mu, \nu \in \mathcal{P}(E)$ and Borel measurable function f,

$$|\langle \mu-\nu, f\rangle|^2 \leqslant 2\left(1+\log\int_E \mathrm{e}^{f(x)^2}\nu(\mathrm{d}x)\right)\mathcal{H}(\mu|\nu). \tag{2.59}$$

(iii) (Variational representation of the relative entropy) For any $\mu, \nu \in \mathcal{P}(E)$,

$$\mathcal{H}(\mu|\nu) = \sup_{\psi\in\mathcal{B}_b(E)}\left(\int_E \psi \mathrm{d}\mu - \log\int_E \mathrm{e}^{\psi}\mathrm{d}\nu\right), \tag{2.60}$$

where $\mathcal{B}_b(E)$ is the set of all bounded Borel measurable functions.

(iv) (Dimensional bounds on entropy) Let μ^N be a symmetric probability measure on E^N and $\mu \in \mathcal{P}(E)$. Then for any $k \leqslant N$,

$$\mathcal{H}\big(\mu^{N,k}|\mu^{\otimes k}\big) \leqslant \frac{2k}{N}\mathcal{H}\big(\mu^N|\mu^{\otimes N}\big), \tag{2.61}$$

where $\mu^{N,k}$ is the marginal distribution of the first k-component of μ^N.

At the end of this section, we introduce the following entropy formula for the weak solutions of SDEs, which is a consequence of Girsanov's theorem (see [70, Lemma 4.4 and Remark 4.5] for the most general form).

Lemma 2.25. *For $i = 1, 2$, let $b^i : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ be two measurable functions. Suppose that for each $i = 1, 2$ that the SDE*

$$\mathrm{d}X_t^i = b^i(t, X_t^i)\mathrm{d}t + \sigma(t, X_t^i)\mathrm{d}W_t^i, \tag{2.62}$$

admits a unique weak solution on $[0, T]$ and for any $\kappa > 0$ and initial data X_0^i,

$$\mathbb{E}\{\exp\big(\kappa\int_0^T |(\sigma^{-1}b^i)(t, X_t^j)|^2\mathrm{d}t\big)\} < \infty, \quad \forall i, j = 1, 2. \tag{2.63}$$

Let $\mathbb{P}_i \in \mathcal{P}(\mathbb{C}_T)$ denotes the law of $(X_t^i)_{t\in[0,T]}$ and $\mu_t^i := \mathbb{P}_i \circ (\omega_t)^{-1}$, $i = 1, 2$. Then for all $t \in [0, T]$ we have

$$\mathcal{H}(\mu_t^1|\mu_t^2) \leqslant \mathcal{H}(\mu_0^1|\mu_0^2) + \frac{1}{2}\mathbb{E}_{\mathbb{P}_1}\left(\int_0^t |\sigma^{-1}(s, w_s)(b^1(s, w_s) - b^2(s, w_s))|^2\mathrm{d}s\right). \tag{2.64}$$

Proof. Without loss of generality, we only prove for $t = T$. Let $\mathbb{P}_i(x) \in \mathcal{P}(\mathbb{C}_T)$ be the law of the unique weak solution to SDE (2.62) when $X_0^i = x$. Recall that $\omega \in \mathbb{C}_T$ is the canonical process. Under $\mathbb{P}_i(x)$,

$$W_t^i := \int_0^t \sigma^{-1}(s,\omega_s)\mathrm{d}\omega_s - \int_0^t (\sigma^{-1}b^i)(s,\omega_s)\mathrm{d}s$$

is a Brownian motion. Then, under assumption (2.63), by Novikov's criterion and Girsanov's theorem,

$$M_t := \exp\left(\int_0^t \big(\sigma^{-1}(b^2 - b^2)\big)((s,\omega_s))\mathrm{d}W_s^1 - \frac{1}{2}\int_0^t |\sigma^{-1}(s,w_s)(b^1(s,w_s) - b^2(s,w_s))|^2\mathrm{d}s\right),$$

$t \in [0,T]$, is a $\mathbb{P}_1(x)$-martingale, and process

$$\widetilde{W}_t := W_t^1 + \int_0^t (\sigma^{-1}(b^1 - b^2))(s,\omega_s)\mathrm{d}s, \quad t \in [0,T],$$

is a Brownian motion under the probability measure defined by $\mathrm{d}\mathbb{Q}(x) := M_T\mathrm{d}\mathbb{P}_1(x)$. We note that $\omega_0 = x$ under both $\mathbb{P}_2(x)$ and $\mathbb{Q}(x)$. By the well-posedness of SDE (2.62) for $i = 2$, we have $M_T\mathrm{d}\mathbb{P}_1(x) = \mathrm{d}\mathbb{P}_2(x)$, which implies that

$$\begin{aligned}\mathcal{H}(\mathbb{P}_1(x)|\mathbb{P}_2(x)) &= -\mathbb{E}_{\mathbb{P}_1(x)}[\log M_T] \\ &= \frac{1}{2}\mathbb{E}_{\mathbb{P}_1(x)}\left(\int_0^t |\sigma^{-1}(s,w_s)(b^1(s,w_s) - b^2(s,w_s))|^2\mathrm{d}s\right).\end{aligned} \tag{2.65}$$

On the other hand, by the weak uniqueness of SDE (2.62) for $i = 1,2$, we have

$$\mathbb{P}_i = \int_{\mathbb{R}^d} \mathbb{P}_i(x)\mu_0^i(\mathrm{d}x). \tag{2.66}$$

Moreover, by [14, Theorem 2.6], we have

$$\mathcal{H}(\mathbb{P}_1 \otimes \mu_0^1|\mathbb{P}_2 \otimes \mu_0^2) = \mathcal{H}(\mu_0^1|\mu_0^2) + \int_{\mathbb{R}^d} \mathcal{H}(\mathbb{P}_1(x)|\mathbb{P}_2(x))\mu_0^1(\mathrm{d}x), \tag{2.67}$$

where $\mathbb{P}_1 \otimes \mu_0^1 := \mathbb{P}_i(x,\mathrm{d}\omega)\mu_0^i(\mathrm{d}x) \in \mathcal{P}(\mathbb{C}_T \times \mathbb{R}^d)$. Now, we let $Y : \mathbb{C}_T \times \mathbb{R}^d \to \mathbb{C}_T$ be $Y(\omega,x) := \omega$. Then, in light of (2.66), one sees that

$$\mathbb{P}_i = \big(\mathbb{P}_i \otimes \mu_0^i\big) \circ Y^{-1}.$$

Thus, by Lemma 2.23, (2.67), (2.65) and (2.66), we have

$$\mathcal{H}(\mathbb{P}_1|\mathbb{P}_2) \leqslant \mathcal{H}(\mu_0^1|\mu_0^2) + \frac{1}{2}\mathbb{E}_{\mathbb{P}_1}\left(\int_0^t |\sigma^{-1}(s,w_s)(b^1(s,w_s) - b^2(s,w_s))|^2\mathrm{d}s\right).$$

In view of the definition, $\mu_t^i = \mathbb{P}_i \circ (\omega_t)^{-1}$, by Lemma 2.23 again, we complete the proof. $\square$

Chapter 3

SDEs with singular coefficients

3.1 Strong well-posedness of SDEs with localized mixed L^p drifts

In this section, we will introduce the main method applied in this thesis-Zvonkin's transformation. In Section 3.1.1, we give a brife proof of the results in [109] so that one can grasp the main points of Zvonkin's argument. Combining the Zvonkin's transformation and the results of PDE (2.33) in Theorem 2.19, we give the proof to following strong well-posedness results in Section 3.1.2:

Theorem 3.1. *Let $T > 0$ and $x \in \mathbb{R}^d$. Assume that the following* $(\mathbf{H}^\sigma_{\mathbf{mix}})$ *holds and for some $(q_i, \boldsymbol{p}_i) \in \mathscr{I}^o$, $\boldsymbol{\pi}_i \in S_d$, $i = 1, 2, .., d$, (2.49) holds.*

$(\mathbf{H}^\sigma_{\mathbf{mix}})$ *There are $\kappa_0 \geqslant 1$ and $\theta \in (0, 1]$ such that for all $t \geqslant 0$ and $x, x', \xi \in \mathbb{R}^d$,*

$$\kappa_0^{-1}|\xi| \leqslant |\sigma(t,x)\xi| \leqslant \kappa_0|\xi|, \ \|\sigma(t,x) - \sigma(t,x')\|_{HS} \leqslant \kappa_0|x - x'|^\theta, \tag{3.1}$$

where $\|\cdot\|_{HS}$ is the usual Hilbert-Schmidt norm of a matrix. Moreover, for some $(q_0, \boldsymbol{p}_0) \in \mathscr{I}^o$ and $\boldsymbol{\pi}_0 \in S_d$ and any $T > 0$,

$$\|\nabla\sigma\|_{\mathbb{L}^{q_0}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_0}_{\boldsymbol{\pi}_0})} \leqslant \kappa_0. \tag{3.2}$$

Then there is a unique strong solution to the following SDE

$$\mathrm{d}X_t = b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \quad X_0 = x. \tag{3.3}$$

Remark 3.2. If we take $d = Nd$ and for $\mathbf{x}^N = (x^1, .., x^N)$ with $x^k \in \mathbb{R}^d$, $k = 1, .., N$, we define

$$b_i(\mathbf{x}^N) := \frac{1}{N}\sum_{j=1}^N K(x^i - x^j)$$

with some $K \in \widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}(\mathbb{R}^d)$, $\boldsymbol{\pi} \in S_d$ and $\boldsymbol{p} = (p_1, ..p_d) \in (2, \infty]^d$, then we have $b_i \in \widetilde{\mathbb{L}}^{\bar{\boldsymbol{p}}_i}_{\boldsymbol{\pi}_i}(\mathbb{R}^{Nd})$ with $\bar{\boldsymbol{p}}_i = (\infty, .., \infty, \bar{p}_{N(d-1)+1}, ., \bar{p}_{Nd})$ and $\boldsymbol{\pi}_i = (d(i-1)+1, .., di, 1, .., d(i-1), di+1, ..., Nd)$. Here $\bar{p}_{N(d-1)+i} = p_i$, $i = 1, .., d$. Moreover, $|\frac{1}{\bar{\boldsymbol{p}}_i}| = |\frac{1}{\boldsymbol{p}}|$. Thus, when $|\frac{1}{\boldsymbol{p}}| < 1$, we obtain the strong well-posedness for the following N-particle systems

$$\mathrm{d}X^{N,i}_t = \frac{1}{N}\sum_{j=1}^{N} K(X^{N,i}_t - X^{N,j}_t)\mathrm{d}t + \mathrm{d}W^i_t.$$

In particular, when $K \in \widetilde{L}^p(\mathbb{R}^d)$, $|\frac{1}{\boldsymbol{p}}| = \frac{d}{p}$, the condition is $p > d \vee 2$.

3.1.1 Strong well-posedness of SDE with localized L^p drifts- An example for Zvonkin's transformation

In this part, we will follow [116, 109] and prove the strong well-posedness of the SDE (3.3) by using Zvonkin's method, where $b \in \widetilde{\mathbb{L}}^{p_1}_{q_1}$ and $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ satisfies the following assumption

($\mathbf{A}^\sigma$) there exist constants $c_0 > 0$ and $\theta \in (0, 1)$ such that for all $(t, x, \xi) \in \mathbb{R}_+ \times \mathbb{R}^{2d}$,

$$c_0^{-1}|\xi|^2 \leqslant |\sigma^*(t, x)\xi|^2 \leqslant c_0|\xi|^2, \quad \|\sigma(t, x) - \sigma(t, y)\|_{HS} \leqslant c_0|x - y|^\theta,$$

where $\|\cdot\|_{HS}$ stands for the Hilbert-Schmidt norm of a matrix.

In the sequel, we set $a_{ij} := \frac{1}{2}\sum_k \sigma_{ik}\sigma_{jk}$. We note that the condition **($\mathbf{A}^\sigma$)** implies the condition **($\mathbf{H}_a$)**. In this part, we will prove the following result

Theorem 3.3. *Assume $b \in \widetilde{\mathbb{L}}^{p_1}_{q_1}$ and $\nabla\sigma \in \widetilde{\mathbb{L}}^{p_2}_{q_2}$ for some $p_i, q_i \in [2, \infty)$ with $d/p_i + 2/q_i < 1$, $i = 1, 2$. Then, for each $x \in \mathbb{R}^d$, there is a unique strong solution X_t to SDE (3.3).*

To this ends, the following Krylov type estimate will play a crucial role.

Lemma 3.4 (Krylov's estimate). *Let $T > 0$, $b \in \widetilde{\mathbb{L}}^{p_1}_{q_1}(T)$ for some $p_1, q_1 > 1$ with $d/p_1 + 2/q_1 < 1$. Assume* **($\mathbf{A}^\sigma$)** *holds. Then, there is a unique weak solution $(X_t)_{t\in[0,T]}$ to SDE (3.3). MOreover, for any $p, q \in (1, \infty]$ with $d/p + 2/q < 2$, there is a constant $C > 0$ such that for any $f \in \widetilde{\mathbb{L}}^p_q(t_0, t_1)$ and all $0 \leqslant t_0 < t_1 \leqslant T$,*

$$\mathbb{E}\left(\int_{t_0}^{t_1} |f(s, X_s)|\mathrm{d}s\Big|\mathscr{F}_{t_0}\right) \leqslant C\|f\|_{\widetilde{\mathbb{L}}^p_q(t_0,t_1)}. \tag{3.4}$$

In order to show the proof, we need the following preparations. First, we give the following two results under stronger conditions.

Lemma 3.5. *Let $T > 0$, $b \in \widetilde{\mathbb{L}}^{p_1}_{q_1}(T)$ for some $p_1, q_1 > 1$ with $d/p_1 + 2/q_1 < 1$. Assume* **($\mathbf{A}^\sigma$)** *holds. For any $p \in (1, p_1]$ and $q \in (1, q_1]$ with $d/p + 2/q < 2$, there is a constant $C > 0$ such that for any weak solution X_t of SDE (3.3), any $f \in \widetilde{\mathbb{L}}^p_q(t_0, t_1)$ and all $0 \leqslant t_0 < t_1 \leqslant T$, (3.4) holds.*

Proof. Without loss of generality, we assume $f \geqslant 0$ and $p, q \neq \infty$. Indeedly, once $p = \infty$ or $q = \infty$, we can choose a $p' < p$ of $q' < q$ such that $d/p' + 2/q' < 2$ and $f \in \widetilde{\mathbb{L}}^{p'}_{q'}(t_0, t_1)$. Let λ_0 be the constant in Lemma 2.8. For any $t_1 \in (0, T]$, by Lemma 2.8, there is a unique solution $u \in \widetilde{\mathbb{H}}^{2,p}_q(t_1)$ with $\partial_t u \in \widetilde{\mathbb{L}}^p_q(t_1)$ to the following backward equation:

$$\partial_t u + a_{ij}\partial_i\partial_j u + b \cdot \nabla u - \lambda_0 u = f, \quad u(t_1) = 0.$$

Let $u_n := u * \Gamma_n$ be the mollifying approximation of u in $\mathbb{R}^{d+1}$. Set

$$T_n f := \partial_t u_n + a_{ij}\partial_i\partial_j u_n + b \cdot \nabla u_n - \lambda_0 u_n.$$

For any $R > 0$, define a stopping time

$$\tau_R := \inf\{t > 0 \mid |X_t| > R\}. \tag{3.5}$$

It follows (2.54) that

$$\lim_{R\to\infty} \tau_R = \infty, \quad \text{a.e.}.$$

Since $u_n \in C_b^\infty(\mathbb{R}^{d+1})$. In view of Itô's formula, we have

$$\mathbb{E}\left(u_n(t_1 \wedge \tau_R)\Big|\mathscr{F}_{t_0}\right) - \mathbb{E}u_n(t_0 \wedge \tau_R) = \mathbb{E}\left(\int_{t_0\wedge\tau_R}^{t_1\wedge\tau_R} (T_n f + \lambda_0 u_n)(s, X_s)\mathrm{d}s\Big|\mathscr{F}_{t_0}\right).$$

Then, taking $(\alpha, p', q') = (0, \infty, \infty)$ in (2.18), one sees that

$$\mathbb{E}\left(\int_{t_0\wedge\tau_R}^{t_1\wedge\tau_R} T_n f(s, X_s)\mathrm{d}s\Big|\mathscr{F}_{t_0}\right) \leqslant (\lambda_0 + 2)\|u_n\|_{\mathbb{L}^\infty(t_0,t_1)} \lesssim \|T_n f\|_{\widetilde{\mathbb{L}}^p_q(t_0,t_1)}. \tag{3.6}$$

In particular, letting $R \to \infty$ and $0 = t_0 \leqslant t_1 = T$, we have

$$\mathbb{E}\left(\int_0^T T_n f(s, X_s)\mathrm{d}s\right) \lesssim \|T_n f\|_{\widetilde{\mathbb{L}}^p_q(T)}. \tag{3.7}$$

Now we claim for any $R > 0$,

$$\lim_{n\to\infty} \|(T_n f - f)\chi_{2R}\|_{\mathbb{L}^p_q(T)} = 0. \tag{3.8}$$

In fact, based on (2.13) and (2.19), by Hölder's inequality, we have

$$\lim_{n\to\infty}\|(T_nf-f)\chi_R\|_{\mathbb{L}^p_q(T)}\lesssim\lim_{n\to\infty}\Big(\|(\partial_tu-\partial_tu_n)\chi_R\|_{\mathbb{L}^p_q}+\|(\nabla^2u-\nabla^2u_n)\chi_R\|_{\mathbb{L}^p_q(T)}$$
$$+\|(u-u_n)\chi_R\|_{\mathbb{L}^p_q}+\|b\cdot(\nabla u-\nabla u_n)\chi_R\|_{\mathbb{L}^p_q(T)}\Big)$$
$$\lesssim\lim_{n\to\infty}\|b\chi_{2R}\|_{\mathbb{L}^{p_1}_{q_1}(T)}\|(\nabla u-\nabla u_n)\chi_R\|_{\mathbb{L}^{p'}_{q'}(T)}=0,$$

where $1/q_1+1/q'=1/q$ and $1/p_1+1/p'=1/p$. Hence,

$$A:=\{T_nf\chi_R\mid n\in\mathbb{N},f\in\widetilde{\mathbb{L}}^p_q(T)\}$$

is a dense subset of $L^q((0,T);L^p(B_R))$, which by (3.7) implies that for almost every $s\in[0,T]$, X_s admits a distributional density $p(s,y)\in L^{\bar q}((0,T);L^{\bar p}(B_R))$ where $1/\bar q+1/q=1$ and $1/\bar p+1/p=1$. As a result, by Hölder's equation, we have for any $R>0$,

$$\lim_{n\to\infty}\mathbb{E}\left(\int_{t_0\wedge\tau_R}^{t_1\wedge\tau_R}|T_nf-f|(s,X_s)\mathrm{d}s\right)\leqslant\lim_{n\to\infty}\int_{t_0}^{t_1}\int_{\mathbb{R}^d}\chi_{2R}(y)|T_nf-f|(s,y)p(s,y)\mathrm{d}s$$
$$\lesssim\lim_{n\to\infty}|\!|\!|(T_nf-f)\chi_{2R}|\!|\!|_{\widetilde{\mathbb{L}}^p_q(t_0,t_1)}=0.$$

Therefore, letting $n\to\infty$ first and $R\to\infty$ in (3.6), we have

$$\mathbb{E}\left(\int_{t_0}^{t_1}f(s,X_s)\mathrm{d}s\Big|\mathscr{F}_{t_0}\right)\lesssim|\!|\!|f|\!|\!|_{\widetilde{\mathbb{L}}^p_q(t_0,t_1)}$$

and complete the proof. □

Based on the method in [110, Lemma 5.5], when $d/p+2/q<1$, we can drop the conditions $p\leqslant p_1$ and $q\leqslant q_1$.

Lemma 3.6. *Let $T>0$, $b\in\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)$ for some $p_1,q_1>1$ with $d/p_1+2/q_1<1$. Assume* $(\mathbf{A}^\sigma)$ *holds. For any $p,q\in(1,\infty]$ with $d/p+2/q<1$, there is a constant $C>0$ such that for any weak solution X_t of SDE (3.3), any $f\in\widetilde{\mathbb{L}}^p_q(t_0,t_1)$ and all $0\leqslant t_0<t_1\leqslant T$, (3.4) holds.*

Proof. For any $t_1\in(0,T]$, by Lemma 2.8, there is a unique solution $u\in\widetilde{\mathbb{H}}^{2,p}_q(t_1)$ with $\partial_tu\in\widetilde{\mathbb{L}}^p_q(t_1)$ to the following backward equation without drift:

$$\partial_tu+a_{ij}\partial_i\partial_ju-\lambda_0u=f,\quad u(t_1)=0.$$

Let $u_n:=u*\Gamma_n$ and set

$$f_n:=\partial_tu_n+a_{ij}\partial_i\partial_ju_n-\lambda_0u_n.$$

Based (3.8), we have

$$\lim_{n\to\infty}\|(f_n-f)\chi_R\|_{\mathbb{L}^p_q}=0.$$

By Itô's formula, one sees that

$$\begin{aligned}\mathbb{E}\left(u_n(t_1\wedge\tau_R)\Big|\mathscr{F}_{t_0}\right)-\mathbb{E}u_n(t_0\wedge\tau_R)=&\mathbb{E}\left(\int_{t_0\wedge\tau_R}^{t_1\wedge\tau_R}(f_n+\lambda_0u_n)(s,X_s)\mathrm{d}s\Big|\mathscr{F}_{t_0}\right)\\&+\mathbb{E}\left(\int_{t_0\wedge\tau_R}^{t_1\wedge\tau_R}(b\cdot\nabla u_n)(s,X_s)\mathrm{d}s\Big|\mathscr{F}_{t_0}\right),\end{aligned}$$

which by taking $(\alpha,p',q')=(0,\infty,\infty)$ and $(1,\infty,\infty)$ in (2.18) implies that

$$\begin{aligned}\mathbb{E}\left(\int_{t_0\wedge\tau_R}^{t_1\wedge\tau_R}f_n(s,X_s)\mathrm{d}s\Big|\mathscr{F}_{t_0}\right)\leqslant&(\lambda_0+2)\|u_n\|_{\mathbb{L}^\infty(t_0,t_1)}\\&+\|\nabla u_n\|_{\mathbb{L}^\infty(t_0,t_1)}\mathbb{E}\left(\int_{t_0\wedge\tau_R}^{t_1\wedge\tau_R}|b(s,X_s)|\mathrm{d}s\Big|\mathscr{F}_{t_0}\right)\\\lesssim&|\!|\!|f_n|\!|\!|_{\widetilde{\mathbb{L}}^p_q(t_0,t_1)},\end{aligned}$$

where the last step is from Lemma 3.5. By the same argument in the proof of Lemma 3.5, we obtain (3.4) for $d/p+2/q<1$ and complete the proof. □

With the help of Lemma 3.6, we have the the following Zvonkin's transformation. Consider the following backward PDE:

$$\partial_tu+a_{ij}\partial_i\partial_ju+b\cdot\nabla u-\lambda u+b=0,\quad u(T)=0,\tag{3.9}$$

where $b\in\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)$ with $d/p_1+2/q<1$ and $(\mathbf{A}^\sigma)$ holds. By reversing the time variable and by Lemma 2.8, there is a unique solution $u\in\widetilde{\mathbb{H}}^{2,p_1}_{q_1}(T)$ such that for any $\alpha\in[0,2)$, $p'\in[p_1,\infty]$ and $q'\in[q_1,\infty]$ with

$$\beta:=2-\alpha-\frac{d}{p'}+\frac{2}{q'}-\frac{d}{p_1}-\frac{2}{q_1}>0,\tag{3.10}$$

there are constants $C>0$ and $\lambda_0\geqslant 1$ such that for any $\lambda\geqslant\lambda_0$,

$$\lambda^{\frac{\beta}{2}}|\!|\!|u|\!|\!|_{\widetilde{\mathbb{H}}^{\alpha,p'}_{q'}(T)}+|\!|\!|\partial_tu|\!|\!|_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)}+|\!|\!|u|\!|\!|_{\widetilde{\mathbb{H}}^{2,p_1}_{q_1}(T)}\leqslant C|\!|\!|b|\!|\!|_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)},\tag{3.11}$$

which implies by Lemma 2.2 that for any $R>0$ and $|h|<1$,

$$|\chi_R(x)(\nabla u(t,x)-\nabla u(s,x+h))|\lesssim(|t-s|+|x-y|)^\delta,\quad\text{with some }\delta>0.\tag{3.12}$$

In particular, since $d/p_1 + 2/q_1 < 1$, when $(\alpha, p', q') = (1, \infty, \infty)$, we can choose $\lambda \geqslant \lambda_0$ large enough so that

$$\|u\|_{\mathbb{L}^\infty_T} + \|\nabla u\|_{\mathbb{L}^\infty_T} \leqslant \frac{1}{2}.$$

Moreover, by Lemma 2.2, ∇u is Hölder continuous on $[0,T] \times \mathbb{R}^d$. Define

$$\Phi(t,x) := x + u(t,x). \tag{3.13}$$

Then, one sees that $x \to \Phi(t,x)$ is a C^1-diffeomorphism for any $t \in [0,T]$ and

$$\|\nabla \Phi\|_{\mathbb{L}^\infty_T} + \|\nabla \Phi^{-1}\|_{\mathbb{L}^\infty_T} \leqslant 4. \tag{3.14}$$

Lemma 3.7 (Zvonkin's transformation)**.** *Assume* $(\mathbf{A}^\sigma)$ *holds and* $b \in \widetilde{\mathbb{L}}^{p_1}_{q_1}(T)$ *with* $d/p_1 + 2/q_1 < 1$. *Let* X_t *be a weak solution to SDE* (3.3). *Then,* $Y_t := \Phi(t, X_t)$ *solves the following SDE:*

$$Y_t = \Phi(0,x) + \int_0^t \tilde{b}(s, Y_s)\mathrm{d}s + \int_0^t \tilde{\sigma}(s, Y_s)\mathrm{d}W_s, \tag{3.15}$$

where

$$\tilde{b}(s,y) := \lambda u(s, \Phi^{-1}(s,y)), \quad \tilde{\sigma}_{ij} := \big(\partial_k \Phi_i \sigma_{kj}\big)(s, \Phi^{-1}(s,y)).$$

Moreover, there is a $\gamma > 0$, *such that*

$$\tilde{b}, \nabla \tilde{b} \in \mathbb{L}^\infty_T, \quad \tilde{\sigma} \in L^\infty([0,T]; C^\gamma) \tag{3.16}$$

and for some $\tilde{\kappa}_0 \geqslant 1$,

$$\tilde{\kappa}_0^{-1}|\xi|^2 \leqslant |\tilde{\sigma}(s,y)\xi|^2 \leqslant \tilde{\kappa}_0 |\xi|^2, \quad \forall (s,x,\xi) \in \mathbb{R}_+ \times \mathbb{R}^{2d}. \tag{3.17}$$

Vice versa, if $(Y_t)_{t\geqslant 0}$ *solvs SDE* (3.15), *then* $X_t := \Phi^{-1}(t, Y_t)$ *solves SDE* (3.3).

Proof. Let $\Phi_n := \Phi * \Gamma_n = x + u * \Gamma_n$ be the mollifying approximation of Φ in $\mathbb{R}^{d+1}$. By Itô's formula, one sees that

$$\begin{aligned} \Phi_n(t, X_t) = \Phi_n(0,x) &+ \int_0^t (\partial_s \Phi_n + a_{ij}\partial_i\partial_j \Phi_n + b \cdot \nabla \Phi_n)(s, X_s)\mathrm{d}s \\ &+ \int_0^t (\sigma_{ij}\partial_i \Phi_n)(s, X_s)\mathrm{d}W^j_s. \end{aligned} \tag{3.18}$$

For any $R > 0$, we define a stopping time

$$\tau_R := \inf\{t > 0 \mid |X_t| > R\}.$$

Let χ_R be defined by (2.1). By Itô's isometric formula and (3.12), we have

$$\mathbb{E}\left|\int_0^{t\wedge\tau_R}(\sigma_{ij}\partial_i(\Phi_n-\Phi))(s,X_s)\mathrm{d}W^j_s\right|^2\leqslant\|\sigma\|^2_\infty\mathbb{E}\left(\int_0^t\chi^2_R(X_s)|\partial_i(u_n-u)|^2(s,X_s)\mathrm{d}s\right)$$
$$\lesssim\|\chi_R(\nabla u_n-\nabla u)\|^2_{\mathbb{L}^\infty_T}\lesssim n^{2\delta},$$

which converges to 0 as $n\to\infty$. By (3.4), we also have

$$\mathbb{E}\left(\int_0^{t\wedge\tau_R}|b\cdot\nabla(\Phi_n-\Phi)|(s,X_s)\mathrm{d}s\right)\leqslant\mathbb{E}\left(\int_0^t\chi_R(X_s)|b\cdot\nabla(u_n-u)|(s,X_s)\mathrm{d}s\right)$$
$$\lesssim\|\chi_R|b\cdot\nabla(u_n-u)|\|_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)}\leqslant\|b\|_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)}\|\chi_R|\nabla(u_n-u)|\|_{\mathbb{L}^\infty_T}\lesssim n^\delta$$

goes to 0 as $n\to\infty$. Similarly, we have

$$\lim_{n\to\infty}\mathbb{E}\left(\int_0^{t\wedge\tau_R}|(\partial_s+a_{ij}\partial_i\partial_j)\nabla(\Phi_n-\Phi)|(s,X_s)\mathrm{d}s\right)$$
$$\overset{(3.4)}{\lesssim}\lim_{n\to\infty}\|\chi_R|(\partial_s+a_{ij}\partial_i\partial_j)(u_n-u)|\|_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)}$$
$$\lesssim\|\chi_R|\partial_s(u_n-u)|\|_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)}+\|\chi_R|\nabla^2(u_n-u)|\|_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)}=0,$$

since (2.13). Taking $n\to\infty$ for (3.24), by Lemma 2.2 one sees that on $t\in[0,\tau_R]$,

$$\Phi(t,X_t)=\Phi(0,x)+\int_0^t(\partial_s\Phi+a_{ij}\partial_i\partial_j\Phi+b\cdot\nabla\Phi)(s,X_s)\mathrm{d}s+\int_0^t(\sigma_{ij}\partial_i\Phi)(s,X_s)\mathrm{d}W^j_s$$
$$=\Phi(0,x)+\lambda\int_0^t u(s,X_s)\mathrm{d}s+\int_0^t(\sigma_{ij}\partial_i\Phi)(s,X_s)\mathrm{d}W^j_s$$

By letting $R\to\infty$, we have (3.15). (3.16) and (3.17) are from (3.11) and (3.14).

Now, we assume $(Y_t)_{t\in[0,T]}$ is a solution to SDE (3.15) and show $(X_t)_{t\geqslant0}:=(\Phi^{-1}(t,Y_t))_{t\in[0,T]}$ solves SDE (3.3). We note that

$$\partial_t\Phi+a_{ij}\partial_i\partial_j\Phi+b\cdot\nabla\Phi-\lambda u=0,\tag{3.19}$$

and

$$0=\partial_t(\Phi^{-1}_k\circ\Phi)=\partial_t\Phi^{-1}_k\circ\Phi+\big(\partial_\ell\Phi^{-1}_k\circ\Phi\big)\partial_t\Phi_\ell;\tag{3.20}$$
$$\delta_{ik}=\partial_i(\Phi^{-1}_k\circ\Phi)=\big(\partial_\ell\Phi^{-1}_k\circ\Phi\big)\partial_i\Phi_\ell;\tag{3.21}$$
$$0=\partial_i\partial_j(\Phi^{-1}_k\circ\Phi)=\partial_i\Phi_{\ell_1}\partial_j\Phi_{\ell_2}\big(\partial_{\ell_1}\partial_{\ell_2}\Phi^{-1}_k\circ\Phi\big)+\big(\partial_\ell\Phi^{-1}_k\circ\Phi\big)\partial_i\partial_j\Phi_\ell,\tag{3.22}$$

where $\delta_{ij}=0$ for $i\neq j$ and $\delta_{ii}=1$. By (3.19), (3.20) and (3.22), one sees that

$$\partial_t\Phi^{-1}_k\circ\Phi+(\tilde a_{ij}\partial_i\partial_j\Phi^{-1}_k)\circ\Phi=\partial_t\Phi^{-1}_k\circ\Phi+a_{ij}\partial_i\Phi_{\ell_1}\partial_j\Phi_{\ell_2}\big(\partial_{\ell_1}\partial_{\ell_2}\Phi^{-1}_k\circ\Phi\big)$$
$$=-\big(\partial_\ell\Phi^{-1}_k\circ\Phi\big)\big(\partial_t\Phi_\ell+a_{ij}\partial_i\partial_j\Phi_\ell\big)$$
$$=\big(\partial_\ell\Phi^{-1}_k\circ\Phi\big)\big(b\cdot\nabla\Phi_\ell-\lambda u_\ell\big),$$

where $\tilde{a}_{ij} := \tilde{\sigma}_{ik}\tilde{\sigma}_{jk}$. It follows from (3.21) that

$$\big(\partial_\ell \Phi_k^{-1} \circ \Phi\big) b_i \partial_i \Phi_\ell = b_k,$$

which implies that

$$\partial_t \Phi^{-1} + \tilde{a}_{ij}\partial_i\partial_j \Phi^{-1} + \tilde{b} \cdot \nabla \Phi^{-1} - b \circ \Phi^{-1} = 0. \tag{3.23}$$

By the definition of Φ, we have

$$x = \Phi \circ \Phi^{-1} = \Phi^{-1} + u \circ \Phi^{-1}, \text{ i.e. } \Phi^{-1} = x - u \circ \Phi^{-1}.$$

Then, $\Phi_n^{-1} := \Phi^{-1} * \Gamma_n = x - (u \circ \Phi^{-1}) * \Gamma_n$ is well-defined. Moreover, in view of (3.22), we have

$$\|\nabla^2 \Phi^{-1}\|_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)} \leqslant \|(\nabla\Phi)^{-1}\|^2_{\mathbb{L}^\infty_T} \|\nabla \Phi^{-1}\|_{\mathbb{L}^\infty_T} \|\nabla^2 \Phi\|_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)} < \infty.$$

We note $\tilde{b}$ is bounded and $\tilde{\sigma}$ satisfies $(\mathbf{A}^\sigma)$ since (3.16) and (3.17). Thus, the Krlov estimate (3.4) holds for $(Y_t)_{t\in[0,T]}$. By the same calculation above, we have

$$\begin{aligned}\Phi^{-1}(t, Y_t) &= x + \int_0^t (\partial_s \Phi^{-1} + \tilde{a}_{ij}\partial_i\partial_j\Phi^{-1} + \tilde{b}\cdot\nabla\Phi^{-1})(s, Y_s)\mathrm{d}s + \int_0^t (\tilde{\sigma}_{ij}\partial_i \Phi^{-1})(s, Y_s)\mathrm{d}W_s^j \\ &= x + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t (\tilde{\sigma}_{ij}\partial_i\Phi^{-1})(s, \Phi(s, X_s))\mathrm{d}W_s^j.\end{aligned}$$

In light of (3.21) again, we have

$$(\tilde{\sigma}_{ij}\partial_i \Phi_k^{-1}) \circ \Phi = \partial_\ell \Phi_i \sigma_{\ell j} \partial_i \Phi_k^{-1} \circ \Phi = \sigma_{kj}$$

and complete the proof. □

Now, we can give

Proof of Lemma 3.4. First of all, it is well known that SDE (3.15) admits a unique weak solution $(Y_t)_{t\in[0,T]}$ under the conditions (3.16) and (3.17). Hence, by Lemma 3.7 there is a unique weak solution $(X_t)_{t\in[0,T]} = (\Phi^{-1}(t, Y_t))_{t\in[0,T]}$ to SDE (3.3). Moreover, since $\tilde{b} \in \mathbb{L}^\infty_T$, by Lemma 3.5, for any $f \in \widetilde{\mathbb{L}}^p_q(t_0, t_1)$ with $p, q \in (1, \infty]$ and $d/p + 2/q < 2$, (3.4) holds for Y_t. Hence, by a change of variable and (3.14), we have

$$\begin{aligned}\mathbb{E}\left(\int_{t_0}^{t_1} |f(s, X_s)|\mathrm{d}s \Big| \mathscr{F}_{t_0}\right) &= \mathbb{E}\left(\int_{t_0}^{t_1} |f(s, \Phi(s, Y_s))|\mathrm{d}s \Big| \mathscr{F}_{t_0}\right) \\ &\lesssim \|f \circ \Phi\|_{\widetilde{\mathbb{L}}^p_q(t_0,t_1)} \lesssim \|f\|_{\widetilde{\mathbb{L}}^p_q(t_0,t_1)}\end{aligned}$$

and complete the proof.

□

Applying the Krylov estimate in Lemma 3.4, we have the following two results.

Corollary 3.8 (Khasminskii's estimate). *Under the same conditions in Lemma 3.4, for any* $f \in \widetilde{\mathbb{L}}_q^p(T)$,

$$\mathbb{E}\left(\exp\left[\int_0^T |f(s,X_s)|\mathrm{d}s\right]\right) < \infty.$$

Proof. For any $m \in \mathbb{N}$, we note that

$$\left(\int_{t_0}^{t_1} g(s)\mathrm{d}s\right)^m = m!\int_{\Delta^m} g(s_1)g(s_2)\cdots g(s_m)\mathrm{d}s_1\cdots\mathrm{d}s_m,$$

where

$$\Delta^m := \left\{(s_1,...,s_m) \in \mathbb{R}_+^m \mid t_0 \leqslant s_1 \leqslant s_2 \leqslant \cdots \leqslant s_m \leqslant t_1\right\}.$$

Then, by (3.4), we have

$$\begin{aligned}
\mathbb{E}^{\mathscr{F}_{t_0}}\left(\int_{t_0}^{t_1} f(s,X_s)\mathrm{d}s\right)^m &= m!\mathbb{E}^{\mathscr{F}_{t_0}}\left(\int_{\Delta^m} f(s_1,X_{s_1})\cdots f(s_m,X_{s_m})\mathrm{d}s_1\cdots\mathrm{d}s_m\right)\\
&= \mathbb{E}^{\mathscr{F}_{t_0}}\Big(\int_{\Delta^{m-1}} f(s_1,X_{s_1})\cdots\\
&\qquad \times\left(\mathbb{E}^{\mathscr{F}_{s_{m-1}}}\int_0^t f(s_m,X_{s_m})\mathrm{d}s_m\right)\mathrm{d}s_1\cdots\mathrm{d}s_{m-1}\Big)\\
&\leqslant \mathbb{E}^{\mathscr{F}_{t_0}}\Big(\int_{\Delta^{m-1}} f(s_1,X_{s_1})\cdots(t_1-t_0)^\gamma\|f\|_{\widetilde{\mathbb{L}}_q^p(T)}\mathrm{d}s_1\cdots\mathrm{d}s_{m-1}\Big)\\
&\leqslant \cdots \leqslant m!\left(C(t_1-t_0)^\gamma\|f\|_{\widetilde{\mathbb{L}}_q^p(T)}\right)^m.
\end{aligned}$$

Hence, by taking $\delta = (2C\|f\|_{\widetilde{\mathbb{L}}_q^p(T)})^{-\gamma}$, we have

$$\begin{aligned}
\mathbb{E}^{\mathscr{F}_t}\left(\exp\left[\int_t^{t+\delta} f(s,X_s)\mathrm{d}s\right]\right) &= \sum_{m\in\mathbb{N}}\frac{1}{m!}\mathbb{E}^{\mathscr{F}_{t_0}}\left(\int_{t_0}^{t_1} f(s,X_s)\mathrm{d}s\right)^m\\
&\leqslant \sum_{m\in\mathbb{N}}(1/2)^m = 2,
\end{aligned}$$

which implies that for $M := [T/\delta]$

$$\begin{aligned}
\mathbb{E}\left(\exp\left[\int_0^T f(s,X_s)\mathrm{d}s\right]\right) &= \mathbb{E}\left(\exp\left[\int_0^{M\delta} f(s,X_s)\mathrm{d}s\right]\mathbb{E}^{\mathscr{F}_{M\delta}}\exp\left[\int_{M\delta}^T f(s,X_s)\mathrm{d}s\right]\right)\\
&\leqslant 2\mathbb{E}\left(\exp\left[\int_0^{M\delta} f(s,X_s)\mathrm{d}s\right]\right) \leqslant \cdots \leqslant 2^M.
\end{aligned}$$

This completes the proof. ∎

Corollary 3.9 (Generalized Itô's formula). *Let $T > 0$, $b \in \widetilde{\mathbb{L}}^{p_1}_{q_1}(T)$ for some $d/p_1 + 2/q_1 < 1$ and X_t solve SDE (3.3). Assume $(\mathbf{A}^\sigma)$ holds. For any $p, q \in (1, \infty]$ with $d/p + 2/q < 1$. For any $u \in \widetilde{\mathbb{H}}^{2,p}_q(T)$ with $\partial_t u \in \widetilde{\mathbb{L}}^p_q(T)$, we have*

$$u(t, X_t) = u(0, x) + \int_0^t (\partial_s u + a_{ij}\partial_i\partial_j u + b \cdot \nabla u)(s, X_s)\mathrm{d}s + \int_0^t (\sigma_{ij}\partial_i u)(s, X_s)\mathrm{d}W^j_s.$$

Proof. Let $u_n := u * \rho_n$ be the mollifying approximation of u in $\mathbb{R}^{d+1}$. By Itô's formula, one sees that

$$u_n(t, X_t) = u_n(0, x) + \int_0^t (\partial_s u_n + a_{ij}\partial_i\partial_j u_n + b \cdot \nabla u_n)(s, X_s)\mathrm{d}s \tag{3.24}$$
$$+ \int_0^t (\sigma_{ij}\partial_i u_n)(s, X_s)\mathrm{d}W^j_s. \tag{3.25}$$

For any $R > 0$, we define a stopping time

$$\tau_R := \inf\{t > 0 \mid |X_t| > R\}.$$

By the same argument as the one in proof of Lemma 3.7, taking $n \to \infty$ for (3.24), by Lemma 2.2 one sees that on $t \in [0, \tau_R]$,

$$u(t, X_t) = u(0, x) + \int_0^t (\partial_s u + a_{ij}\partial_i\partial_j u + b \cdot \nabla u)(s, X_s)\mathrm{d}s + \int_0^t (\sigma_{ij}\partial_i u_n)(s, X_s)\mathrm{d}W^j_s.$$

By letting $R \to \infty$, we have the desired formula and complete the proof. □

Remark 3.10. *Here we drop the restriction $p, q > 2$ in [109, Lemma 4.1-(iii)].*

It is the position to give

Proof of Theorem 3.3. Recall we have obtained the unique weak solution and (2.54) in Lemma 3.4. Let u and Φ defined by (3.9) and (3.13). By Yamada-Watanabe's theorem and Lemma 3.7, we only need to prove the pathwise uniqueness for SDE (3.15). For $i = 1, 2$, let $(Y^{(i)}_t)_{t\in[0,T]}$ be two solution to SDE (3.3) driven by the same Brownian motion $(W_t)_{t\geqslant 0}$ with the same staring point y, that is,

$$Y^{(i)}_t = y + \int_0^t \tilde{b}(s, Y^{(i)}_s)\mathrm{d}s + \int_0^t \tilde{\sigma}(s, Y^{(i)}_s)\mathrm{d}W_s,$$

where $\tilde{b}$ and $\tilde{\sigma}$ are defined in Lemma 3.7. We set $Z_t := |Y^{(1)}_t - Y^{(2)}_t|^2$. By Itô's formula, we have

$$Z_t = \int_0^t Z_s \mathrm{d}A_s + M_t, \tag{3.26}$$

where M_t is a continuous local martingale given by

$$M_t := \int_0^t \Big[\tilde\sigma(s,Y_s^{(1)}) - \tilde\sigma(s,Y_s^{(2)})\Big]^*(Y_s^{(1)} - Y_s^{(2)})\mathrm{d}W_s$$

and A_t is defined by

$$A_t := \int_0^t \frac{2\langle Y_s^{(1)} - Y_s^{(2)}, \tilde b(s,Y_s^{(1)}) - \tilde b(s,Y_s^{(2)})\rangle + \|\tilde\sigma(s,Y_s^{(1)}) - \tilde\sigma(s,Y_s^{(2)})\|_{HS}^2}{|Y_s^{(1)} - Y_s^{(2)}|^2}\mathrm{d}s.$$

Based on (2.15), we have

$$\begin{aligned}|A_t| \lesssim & t\|\nabla\tilde b\|_{\mathbb{L}_T^\infty} + \int_0^t \left(\mathcal{M}|\nabla\tilde\sigma|^2(s,Y_s^{(1)}) + \mathcal{M}|\nabla\tilde\sigma|^2(s,Y_s^{(2)}) + \|\tilde\sigma(s)\|_\infty^2\right)\mathrm{d}s\\ \lesssim & 1 + \int_0^t \left(\mathcal{M}|\nabla\sigma|^2(s,Y_s^{(1)}) + \mathcal{M}|\nabla\sigma|^2(s,Y_s^{(2)})\right)\mathrm{d}s\\ & + \int_0^t \left(\mathcal{M}|\nabla^2 u|^2(s,Y_s^{(1)}) + \mathcal{M}|\nabla^2 u|^2(s,Y_s^{(2)})\right)\mathrm{d}s,\end{aligned}$$

where we have used $|\nabla\tilde\sigma| \lesssim |\nabla\sigma| + |\nabla^2 u|$.

On the other hand, it follows from (2.16) that

$$\|\mathcal{M}|\nabla\sigma|^2\|_{\widetilde{\mathbb{L}}_{q_2/2}^{p_2/2}(T)} + \|\mathcal{M}|\nabla^2 u|^2\|_{\widetilde{\mathbb{L}}_{q_1/2}^{p_1/2}(T)} \lesssim \|\nabla\sigma\|_{\widetilde{\mathbb{L}}_{q_2}^{p_2}(T)}^2 + \|\nabla^2 u\|_{\widetilde{\mathbb{L}}_{q_1}^{p_1}(T)}^2 < \infty.$$

Hence, by Corollary 3.8, we have

$$\mathbb{E}\exp\left(\kappa A_T\right) < \infty, \quad \forall \kappa > 0,$$

which by (3.26) and stochastic Gronwall's inequality Lemma A.5 that

$$\mathbb{E}\Big(\sup_{t\in[0,T]} Z_t\Big) = 0.$$

This completes the proof based on Lemma 3.7. □

3.1.2 Proof of the main theorem

In this section we will show Theorem 3.1. To this end, we first establish a priori Krylov estimates like (3.4) for any solution of SDEs with mixed drifts and for any index $(q,\boldsymbol{p}) \in \mathscr{I}_1$, where $\mathscr{I}_1$ is defined in (2.29). Using this a priori estimates, one can perform the classical Zvonkin transformation like Lemma 3.6, and then establish the weak well-posedness under conditions $(\mathbf{A}^\sigma)$ and (2.49). Moreover, we also obtain the two-sided density estimates. As a byproduct, one improves the Krylov estimate to any index $(q,\boldsymbol{p}) \in \mathscr{I}_2$, which is also crucial for the strong well-posedness and the propagation of chaos.

By Theorem 2.17, we can establish the following a priori Krylov estimate.

Lemma 3.11. *Suppose that* $(\mathbf{A}^\sigma)$ *and* (2.49) *hold for some* $(q_i,\boldsymbol{p}_i)\in\mathscr{I}_1$, $i=1,..,d$. *Then for any* $(q,\boldsymbol{p})\in\mathscr{I}_1$, $\boldsymbol{\pi}\in S_d$ *and* $T>0$, *there is a constant* $C=C(\Theta,T)>0$ *such that for all* $f\in\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})$ *and any weak solution* X *of SDE* (3.3),

$$\mathbb{E}\left(\int_0^T f(s,X_s)\mathrm{d}s\right)\leqslant C\|f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{3.27}$$

Proof. By the same reason in the proof to Lemma 3.5, without loss of generality, we assume $f\geqslant 0$ and $p_i,q\neq\infty$, $i=1,...,d$. By reversing the time variable and by Theorem 2.17, there is a unique solution u to the following backward PDE

$$\partial_t u+\mathrm{tr}(a\cdot\nabla^2u)-\lambda u=f,\quad u(T)=0,$$

such that for any $\beta\in[0,2-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})$,

$$(1\vee\lambda)^{\frac12(2-\beta-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})}\|u\|_{\mathbb{L}^\infty_T(\mathcal{C}^\beta)}+\|\nabla^2u\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}\lesssim\|f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{3.28}$$

We define $u_n:=u*\Gamma_n$ and $f_n:=\partial_t u_n+\mathrm{tr}(a\cdot\nabla^2u_n)-\lambda u_n$. τ_R is defined in (3.5). For any $m\in\mathbb{N}$, we also define another stopping time here

$$\sigma_n:=\inf\{t>0:\int_0^t|b(s,X_s)|\mathrm{d}s\geqslant n\}\wedge T.$$

Based on Itô's formula, one sees that

$$\mathbb{E}u_n(\sigma_m\wedge\tau_R,X_{\sigma_m\wedge\tau_R})-u_n(0,x)=\mathbb{E}\left(\int_0^{\sigma_m\wedge\tau_R}(f_n+\lambda u_n+b\cdot\nabla u_n)(s,X_s)\mathrm{d}s\right).$$

Then, by the same argument in the proof of Lemma 3.5, taking $n\to\infty$ and $R\to\infty$, we have

$$\mathbb{E}\left(\int_0^{\sigma_m}f(s,X_s)\mathrm{d}s\right)\leqslant(2+\lambda)\|u\|_{\mathbb{L}^\infty_T}+\|\nabla u\|_{\mathbb{L}^\infty_T}\mathbb{E}\left(\int_0^{\sigma_m}|b(s,X_s)|\mathrm{d}s\right),$$

which by choosing $\beta=0,1$ in (3.28) implies that

$$\begin{aligned}\mathbb{E}\left(\int_0^{\sigma_m}f(s,X_s)\mathrm{d}s\right)\leqslant C\|f\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}\Big\{&(1\vee\lambda)^{\frac12(|\frac{1}{\boldsymbol{p}}|+\frac{2}{q})}\\&+(1\vee\lambda)^{\frac12(|\frac{1}{\boldsymbol{p}}|+\frac{2}{q}-1)}\mathbb{E}\left(\int_0^{\sigma_m}|b(s,X_s)|\mathrm{d}s\right)\Big\}.\end{aligned} \tag{3.29}$$

Taking $(q,\boldsymbol{p})=(q_i,\boldsymbol{p}_i)$ and $f=b_i$, $i=1,2,..,d$, we have

$$\begin{aligned}\mathbb{E}\left(\int_0^{\sigma_m}|b(s,X_s)|\mathrm{d}s\right)\leqslant&C\kappa_1\sum_{i=1}^d(1\vee\lambda)^{\frac12(|\frac{1}{\boldsymbol{p}_i}|+\frac{2}{q_i})}\\&+C\kappa_1\sum_{i=1}^d(1\vee\lambda)^{\frac12(|\frac{1}{\boldsymbol{p}_i}|+\frac{2}{q_i}-1)}\mathbb{E}\left(\int_0^{\sigma_m}|b(s,X_s)|\mathrm{d}s\right),\end{aligned}$$

where κ_1 is defined in (2.49). Since $(q_i, \boldsymbol{p}_i) \in \mathscr{I}_1$, $i = 1, .., d$, by taking λ large enough so that $C\kappa_1 \sum_{i=1}^d (1 \vee \lambda)^{\frac{1}{2}(|\frac{1}{\boldsymbol{p}_i}| + \frac{2}{q_i} - 1)} = \frac{1}{2}$,

$$\mathbb{E}\left(\int_0^{\sigma_m} |b(s, X_s)| \mathrm{d}s\right) \leqslant 2C\kappa_1 \sum_{i=1}^d (1 \vee \lambda)^{\frac{1}{2}(|\frac{1}{\boldsymbol{p}_i}| + \frac{2}{q_i})}.$$

Subsitutiong it into (3.29) and letting $m \to \infty$, we complete the proof. □

In the above lemma, the requirement of $(q, \boldsymbol{p}) \in \mathscr{I}_1$ is too strong for applications. We need to improve it to $(q, \boldsymbol{p}) \in \mathscr{I}_2$ like Lemma 3.4. Inspired by the proof of Lemma 3.4, we use Theorem 2.19 and the above a priori Krylov estimate to construct the Zvonkin transformation. For each $i = 1, \cdots, d$, consider the following backward PDE:

$$\partial_t u_i + \tfrac{1}{2}\mathrm{tr}((\sigma\sigma^*) \cdot \nabla^2 u_i) + b \cdot \nabla u_i - \lambda u_i + b_i = 0, \quad u_i(T) = 0. \tag{3.30}$$

By reversing the time variable and by Theorem 2.19, there is a unique solution u_i satisfying the following estimates: for any $\beta \in (0, \vartheta)$, where ϑ is defined in (2.50), there are $C_0, C_1 \geqslant 1$ such that for all $\lambda \geqslant C_0 \kappa_1^{2/\vartheta}$,

$$\lambda^{\frac{1}{2}(\vartheta - \beta)} \|u_i\|_{\mathbb{L}^\infty_T(\mathcal{C}^{1+\beta})} + \sum_{j=0}^d |\!|\!| \nabla^2 u_{ij} |\!|\!|_{\widetilde{\mathbb{L}}^{q_{ij}}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_{ij}}_{\boldsymbol{\pi}_{ij}})} \leqslant C_1 \kappa_1, \tag{3.31}$$

and for all $0 \leqslant t_0 < t_1 \leqslant T$,

$$\|u_i(t_1) - u_i(t_0)\|_\infty \leqslant C(\lambda)|t_1 - t_0|^{1/2}, \tag{3.32}$$

where

$$u_i = u_{i0} + u_{i1} + u_{i2} + \cdots + u_{id}, \tag{3.33}$$

and

$$q_{i0} = q_i,\ \boldsymbol{p}_{i0} = \boldsymbol{p}_i,\ \boldsymbol{\pi}_{i0} = \boldsymbol{\pi}_i,\ q_{ij} = q_j,\ \boldsymbol{p}_{ij} = \boldsymbol{p}_j,\ \boldsymbol{\pi}_{ij} = \boldsymbol{\pi}_j,\ j = 1, \cdots, d.$$

Below we set

$$\mathbf{u} = (u_1, \cdots, u_d).$$

By (3.31), for any $\beta \in (0, \vartheta)$, we can choose λ large enough so that

$$\|\nabla \mathbf{u}\|_{\mathbb{L}^\infty_T} \leqslant \|\mathbf{u}\|_{\mathbb{L}^\infty_T(\mathcal{C}^{1+\beta})} \leqslant \tfrac{1}{2}.$$

Once λ is chosen, it shall be fixed below without further notice. Now if we define

$$\Phi(t, x) := x + \mathbf{u}(t, x),$$

then for each t,

$$x \mapsto \Phi(t,x) \text{ is a } C^1\text{-diffeomorphism},$$

and

$$\|\nabla\Phi\|_{\mathbb{L}^\infty_T} + \|\nabla\Phi^{-1}\|_{\mathbb{L}^\infty_T} \leqslant 2, \tag{3.34}$$

and by (3.32), for all $0 \leqslant t_0 < t_1 \leqslant T$,

$$\|\Phi(t_1) - \Phi(t_0)\|_\infty \leqslant C(\lambda)(t_1 - t_0)^{1/2}. \tag{3.35}$$

We have the following result (see [110, Theorem 3.10]).

Lemma 3.12 (Zvonkin's transformation). *Under assumption* $(\mathbf{A}^\sigma)$ *and* (2.49), $Y_t := \Phi(t, X_t)$ *solves the following SDE*

$$Y_t = Y_0 + \int_0^t \tilde{b}(s, Y_s)\mathrm{d}s + \int_0^t \tilde{\sigma}(s, Y_s)\mathrm{d}W_s, \tag{3.36}$$

where $Y_0 := \Phi(0, X_0)$ *and*

$$\tilde{b}(s,y) := \lambda\mathbf{u}(s, \Phi^{-1}(s,y)), \quad \tilde{\sigma}(s,y) := (\sigma^*\nabla\Phi)(s, \Phi^{-1}(s,y)).$$

Moreover, for any $\beta \in (0, \vartheta \wedge \gamma_0)$, *where* ϑ *is defined by* (2.50),

$$\tilde{b}, \ \nabla\tilde{b} \in \mathbb{L}^\infty_T, \ \tilde{\sigma} \in \mathbb{L}^\infty_T(\mathcal{C}^\beta), \tag{3.37}$$

and for some $\tilde{\kappa}_0 \geqslant 1$,

$$\tilde{\kappa}_0^{-1}|\eta|^2 \leqslant |\tilde{\sigma}(s,y)\eta|^2 \leqslant \tilde{\kappa}_0|\eta|^2, \quad \eta \in \mathbb{R}^d. \tag{3.38}$$

Vice versa, if Y_t *solves SDE* (3.36), *then* $X_t := \Phi^{-1}(t, Y_t)$ *solves SDE* (3.3).

Proof. Similar as the proof of Lemma 3.7. For each $n \in \mathbb{N}$, define

$$\mathbf{u}^n(t,x) := (\mathbf{u}(\cdot) * \mathit{\Gamma}_{1/n})(t,x), \ \Phi^n(t,x) := x + \mathbf{u}^n(t,x).$$

By Itô's formula, we have

$$\Phi^n(t, X_t) = \Phi^n(0, X_0) + \int_0^t \mathscr{L}\Phi^n(s, X_s)\mathrm{d}s + \int_0^t (\sigma^*\nabla\Phi^n)(s, X_s)\mathrm{d}W_s, \tag{3.39}$$

where

$$\mathscr{L} := \partial_s + \mathrm{tr}(a \cdot \nabla^2) + b \cdot \nabla, \quad a := (\sigma\sigma^*)/2.$$

For $R > 0$, we define the stopping time

$$\tau_R := \inf\{t > 0 : |X_t| \geqslant R\}.$$

We note that by (3.30) and (3.31),

$$\|\partial_t u_{ij}\|_{\widetilde{\mathbb{L}}_T^{q_{ij}}(\mathbb{L}_{\boldsymbol{\pi}_{ij}}^{\boldsymbol{p}_{ij}})} \lesssim \|\nabla^2 u_{ij}\|_{\widetilde{\mathbb{L}}_T^{q_{ij}}(\mathbb{L}_{\boldsymbol{\pi}_{ij}}^{\boldsymbol{p}_{ij}})} + (1+\|\nabla u_i\|_{\mathbb{L}_T^\infty})\|b_j\|_{\widetilde{\mathbb{L}}_T^{q_{ij}}(\mathbb{L}_{\boldsymbol{\pi}_{ij}}^{\boldsymbol{p}_{ij}})} + \|u_{ij}\|_{\widetilde{\mathbb{L}}_T^{q_{ij}}(\mathbb{L}_{\boldsymbol{\pi}_{ij}}^{\boldsymbol{p}_{ij}})} < \infty.$$

For each i,j, since $(q_{ij},\boldsymbol{p}_{ij})\in\mathscr{I}_1$, by the Krylov estimates (3.27) and (3.31), we have

$$\begin{aligned}&\mathbb{E}\left(\int_0^{t\wedge\tau_m}|(\partial_t - a\cdot\nabla^2)(\Phi_{ij}^n-\nabla^2\Phi_{ij})|(s,X_s)\mathrm{d}s\right)\\ &\leqslant C_m\|\sigma\|_{\mathbb{L}_T^\infty}^2\left(\|\nabla^2(u_{ij}^n-u_{ij})\mathbb{1}_{B_m}\|_{\mathbb{L}_T^{q_{ij}}(\mathbb{L}_{\boldsymbol{\pi}_{ij}}^{\boldsymbol{p}_{ij}})}+\|\partial_t(u_{ij}^n-u_{ij})\mathbb{1}_{B_m}\|_{\mathbb{L}_T^{q_{ij}}(\mathbb{L}_{\boldsymbol{\pi}_{ij}}^{\boldsymbol{p}_{ij}})}\right)\to 0,\end{aligned}$$

$$\begin{aligned}&\mathbb{E}\left|\int_0^{t\wedge\tau_m}|\sigma^*\nabla(\Phi^n-\Phi)|(s,X_s)\mathrm{d}W_s\right|^2\\ &\leqslant C\|\sigma\|_{\mathbb{L}_T^\infty}^2\left(\|\nabla(\mathbf{u}^n-\mathbf{u})\|_{\mathbb{L}_T^\infty}\leqslant Cn^{-(\beta\wedge1/2)}\to 0,\right.\end{aligned}$$

provided by (3.32) and

$$\mathbb{E}\left|\int_0^{t\wedge\tau_m}|b\cdot\nabla(\Phi_{ij}^n-\Phi_{ij})|(s,X_s)\mathrm{d}s\right|\leqslant C\|b\|_{\widetilde{\mathbb{L}}_T^{q_{ij}}(\mathbb{L}_{\boldsymbol{\pi}_{ij}}^{\boldsymbol{p}_{ij}})}\|\nabla(\mathbf{u}^n-\mathbf{u})\|_{\mathbb{L}_T^\infty}\to 0.$$

Since $\tau_R\to\infty$ as $R\to\infty$, replacing t by $t\wedge\tau_R$ in (3.39), taking $n\to\infty$ and $R\to\infty$, we have (3.36). Moreover, (3.37) and (3.38) directly follow by their definitions and (3.31). On the other hand, if Y_t solves SDE (3.36), then by similar calculations as the ones in the proof of Lemma 3.7, $X_t := \Phi^{-1}(t,Y_t)$ solves SDE (3.3). This completes the proof. □

Remark 3.13. An immediate consequence of Zvonkin's transformation together with (3.34) and (3.35) is that for any $p\geqslant 1$ and $T>0$, there is a constant $C=C(p,T,\Theta)>0$ such that

$$\mathbb{E}|X_t-X_s|^{2p}\leqslant C|t-s|^p,\quad t,s\in[0,T]. \tag{3.40}$$

Now, we can give the Krylov estimate like Lemma 3.4. However, we cannot apply the proof of Lemma 3.4 directly. Since by Example 2.3, $|\!|\!| f\circ\Phi|\!|\!|_{\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}}}\lesssim|\!|\!| f|\!|\!|_{\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}}}$ may not hold for the localized mixed $\mathbb{L}_{\boldsymbol{\pi}}^{\boldsymbol{p}}$-space. To overcome this obstacle, we use the heat kernel estimate and show the following main result of this section.

Theorem 3.14. *Suppose that* $(\mathbf{A}^\sigma)$ *and* (2.49) *hold. For any* $\mu_0\in\mathcal{P}(\mathbb{R}^d)$, *there is a unique weak solution to SDE* (3.3) *and initial distribution* μ_0. *Moreover, we have :*

(i) For each $t>0$, X_t *admits a density* $\rho_t^X(y)$ *with the following two-sided estimate: for any* $T>0$, *there are* $\delta_1,C_1\geqslant 1$ *such that for all* $t\in(0,T]$ *and* $y\in\mathbb{R}^d$,

$$\frac{C_1^{-1}}{t^{d/2}}\int_{\mathbb{R}^d}\mathrm{e}^{-\frac{\delta_1|x-y|^2}{t}}\mu_0(\mathrm{d}x)\leqslant\rho_t^X(y)\leqslant\frac{C_1}{t^{d/2}}\int_{\mathbb{R}^d}\mathrm{e}^{-\frac{|x-y|^2}{\delta_1 t}}\mu_0(\mathrm{d}x). \tag{3.41}$$

(ii) Let ϑ be defined as in (2.50). For any $\beta \in (0, \vartheta \wedge \theta)$ and $T > 0$, there are $\delta_2, C_2 \geqslant 1$ such that for all $t \in (0, T]$ and $y, y' \in \mathbb{R}^d$,

$$\frac{|\rho_t^X(y) - \rho_t^X(y')|}{|y - y'|^\beta} \leqslant C_2 t^{-\frac{d+\beta}{2}} \left[\int_{\mathbb{R}^d} \mathrm{e}^{-\frac{|x-y|^2}{\delta_2 t}} \mu_0(\mathrm{d}x) + \int_{\mathbb{R}^d} \mathrm{e}^{-\frac{|x-y'|^2}{\delta_2 t}} \mu_0(\mathrm{d}x) \right]. \tag{3.42}$$

(iii) For any $(q, \boldsymbol{p}) \in \mathscr{I}_2$ and $T > 0$, there is a constant $C_0 > 0$ such that for any $f \in \widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})$,

$$\mathbb{E}\left(\int_0^T f(s, X_s) \mathrm{d}s \right) \leqslant C_0 \|f\|_{\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})}. \tag{3.43}$$

Proof. By (3.37) and (3.38), it is well known that SDE (3.36) admits a unique weak solution (cf. [97]). The existence and uniqueness of weak solutions for the original SDE follow from Lemma 3.12. Next we shall prove (3.41), (3.42) and (3.43).

(i) Let $\widetilde{\mathscr{L}}$ be the generator of SDE (3.36), i.e.,

$$\widetilde{\mathscr{L}} := \mathrm{tr}((\widetilde{\sigma}\widetilde{\sigma}^*) \cdot \nabla^2)/2 + \widetilde{b} \cdot \nabla.$$

By (3.37), (3.38) and Theorems 1.1, 1.3 and 2.3 of [24], there is a fundamental solution $p(s, x, t, y)$ associated with $\widetilde{\mathscr{L}}$, which satisfies the following estimates: for all $0 \leqslant s < t \leqslant T$ and $x, y \in \mathbb{R}^d$,

$$\frac{C_0^{-1}}{(t-s)^{d/2}} \mathrm{e}^{-\frac{\delta_0 |x-y|^2}{t-s}} \leqslant p(s, x, t, y) \leqslant \frac{C_0}{(t-s)^{d/2}} \mathrm{e}^{-\frac{|x-y|^2}{\delta_0 (t-s)}},$$

and for any $\beta \in (0, \vartheta \wedge \theta)$, and for all $0 \leqslant s < t \leqslant T$ and $x, y, y' \in \mathbb{R}^d$,

$$|p(s, x, t, y) - p(s, x, t, y')| \leqslant C_0 |y - y'|^\beta (t-s)^{-\frac{d+\beta}{2}} \left[\mathrm{e}^{-\frac{|x-y|^2}{\delta_0 (t-s)}} + \mathrm{e}^{-\frac{|x-y'|^2}{\delta_0 (t-s)}} \right],$$

where $\delta_0, C_0 \geqslant 1$ only depend on Θ and the bounds of $\tilde{b}$ and $\tilde{\sigma}$. In particular, $p(0, x, t, y)$ is just the density of the solution of SDE (3.3) starting from x at time zero. Note that the density $\rho_t^Y(y)$ of Y_t starting from the initial distribution $\widetilde{\mu}_0 = \mu_0 \circ \Phi(0, \cdot)^{-1}$ is given by

$$\rho_t^Y(y) = \int_{\mathbb{R}^d} p(0, x, t, y) \widetilde{\mu}_0(\mathrm{d}x).$$

This can be shown by considering a smooth approximation and taking weak limits (see [78, Section 5.1] for more details). We thus have that for any $t \in (0, T]$ and all $y, y' \in \mathbb{R}^d$,

$$\frac{C_0^{-1}}{t^{d/2}} \int_{\mathbb{R}^d} \mathrm{e}^{-\frac{\delta_0 |x-y|^2}{t}} \widetilde{\mu}_0(\mathrm{d}x) \leqslant \rho_t^Y(y) \leqslant \frac{C_0}{t^{d/2}} \int_{\mathbb{R}^d} \mathrm{e}^{-\frac{|x-y|^2}{\delta_0 t}} \widetilde{\mu}_0(\mathrm{d}x)$$

and

$$|\rho_t^Y(y)-\rho_t^Y(y')|\leqslant \frac{C_0|y-y'|^\beta}{t^{(d+\beta)/2}}\int_{\mathbb{R}^d}\left[\mathrm{e}^{-\frac{|x-y|^2}{\delta_0 t}}+\mathrm{e}^{-\frac{|x-y'|^2}{\delta_0 t}}\right]\widetilde{\mu}_0(\mathrm{d}x). \tag{3.44}$$

On the other hand, by change of variables, we have

$$\rho_t^X(y)=\rho_t^Y(\Phi(t,y))\det(\nabla\Phi(t,y)), \tag{3.45}$$

and for some $\tilde{C}_0\geqslant 1$,

$$\frac{\tilde{C}_0^{-1}}{t^{d/2}}\int_{\mathbb{R}^d}\mathrm{e}^{-\frac{\delta_0|\Phi(0,x)-\Phi(t,y)|^2}{t}}\mu_0(\mathrm{d}x)\leqslant\rho_t^X(y)\leqslant\frac{\tilde{C}_0^{-1}}{t^{d/2}}\int_{\mathbb{R}^d}\mathrm{e}^{-\frac{|\Phi(0,x)-\Phi(t,y)|^2}{\delta_0 t}}\mu_0(\mathrm{d}x).$$

which together with the following two estimates yields (3.41),

$$|\Phi(0,x)-\Phi(t,y)|^2\geqslant\tfrac12|\Phi(t,x)-\Phi(t,y)|^2-|\Phi(0,x)-\Phi(t,x)|^2\overset{(3.34)(3.35)}{\geqslant}\tfrac18|x-y|^2-Ct,$$

and

$$|\Phi(0,x)-\Phi(t,y)|^2\leqslant 2|\Phi(t,x)-\Phi(t,y)|^2+2|\Phi(0,x)-\Phi(t,x)|^2\overset{(3.34)(3.35)}{\leqslant}8|x-y|^2-Ct.$$

(ii) By (3.45) and (3.44), we have

$$\begin{aligned}|\rho_t^X(y)-\rho_t^X(y')|&\leqslant|\rho_t^Y(\Phi(t,y))-\rho_t^Y(\Phi(t,y'))|\det(\nabla\Phi(t,y))\\&\quad+\rho_t^Y(\Phi(t,y'))|\det(\nabla\Phi(t,y))-\det(\nabla\Phi(t,y'))|\\&\lesssim\frac{|y-y'|^\beta}{t^{(d+\beta)/2}}\int_{\mathbb{R}^d}\left[\mathrm{e}^{-\frac{|x-\Phi(t,y)|^2}{\delta_0 t}}+\mathrm{e}^{-\frac{|x-\Phi(t,y')|^2}{\delta_0 t}}\right]\widetilde{\mu}_0(\mathrm{d}x)\\&\quad+\frac{1}{t^{d/2}}\int_{\mathbb{R}^d}\mathrm{e}^{-\frac{|x-\Phi(t,y)|^2}{\delta_0 t}}\widetilde{\mu}_0(\mathrm{d}x)|\nabla\Phi(t,y)-\nabla\Phi(t,y')|,\end{aligned}$$

which in turn implies (3.42) by (3.31).

(iii) For nonnegative $f\in\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})$ with $(q,\boldsymbol{p})\in\mathscr{I}_2$, by (3.41) and (2.47), we get

$$\begin{aligned}\mathbb{E}\left(\int_0^T f(s,X_s)\mathrm{d}s\right)&=\int_0^T\int_{\mathbb{R}^d}f(s,y)\rho_s^X(y)\mathrm{d}y\mathrm{d}s\\&\leqslant\int_0^T\int_{\mathbb{R}^d}f(s,y)\left(\frac{C_2}{s^{d/2}}\int_{\mathbb{R}^d}\mathrm{e}^{-\frac{|x-y|^2}{2\delta_1 s}}\mu_0(\mathrm{d}x)\right)\mathrm{d}y\mathrm{d}s\\&=C_2(2\pi\delta_1)^{d/2}\int_{\mathbb{R}^d}\left(\int_0^T\mathbb{E}f(s,x-W_{\delta_1 s})\mathrm{d}s\right)\mu_0(\mathrm{d}x)\leqslant C_3\|f\|_{\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})}.\end{aligned}$$

The proof is complete. □

As a corollary, we have the following important exponential integrability of singular functionals.

Corollary 3.15. *(Khasminskii's estimate) Let X be the unique solution of SDE (3.3) in Theorem 3.14. For any $T,\lambda>0$, $(q,\boldsymbol{p})\in\mathscr{I}_2$ and $\beta\in(0,2-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})$, there is a constant $C_1>0$ depending only on $T,\lambda,d,\beta,\kappa_0,\kappa_1,q_i,\boldsymbol{p}_i,q,\boldsymbol{p}$ such that for all $f\in\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})$,*

$$\mathbb{E}\exp\left\{\lambda\int_0^T f(s,X_s)\mathrm{d}s\right\}\leqslant \mathrm{e}^{C_1\|\!|f|\!\|^{2/\beta}_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}}. \tag{3.46}$$

Moreover, if b is bounded measurable, then for some $C_2=C_2(T,\lambda,d,\beta,\kappa_0,q,\boldsymbol{p})>0$,

$$\mathbb{E}\exp\left\{\lambda\int_0^T f(s,X_s)\mathrm{d}s\right\}\leqslant \mathrm{e}^{C_2\left(\|b\|^2_{\mathbb{L}^\infty_T}+\|\!|f|\!\|^{2/\beta}_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}\right)}. \tag{3.47}$$

Proof. Let $\beta\in(0,2-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})$. For (3.46), based on the proof of Corollary 3.8 (also see [110, Lemma 3.5]), it suffices to show that for any $0\leqslant t_0<t_1\leqslant T$,

$$\mathbb{E}\left(\int_{t_0}^{t_1} f(s,X_s)\mathrm{d}s\Big|\mathscr{F}_{t_0}\right)\leqslant C_0(t_1-t_0)^{\frac{\beta}{2}}\|\!|f|\!\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}. \tag{3.48}$$

Let $\frac{1}{q'}=\frac{1}{q}+\frac{\beta}{2}$. Since $\beta\in(0,2-|\frac{1}{\boldsymbol{p}}|-\frac{2}{q})$, we have $(q',\boldsymbol{p})\in\mathscr{I}_2$. By (3.43) and Hölder's inequality,

$$\mathbb{E}\left(\int_0^{t_1-t_0} f(s,X_s)\mathrm{d}s\right)\leqslant C_0\|\!|f|\!\|_{\widetilde{\mathbb{L}}^{q'}_{t_1-t_0}(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}\leqslant C_0(t_1-t_0)^{\frac{\beta}{2}}\|\!|f|\!\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})}.$$

By the Markov property of X_t, we get (3.48). (3.47) follows by Girsanov's theorem. □

Now, we can give

Proof of Theorem 3.1. By Yamada-Watanabe's theorem, it suffices to show the pathwise uniqueness. Based on the proof of Theorem 3.3, this follows by Zvonkin's transformation (see Lemma 3.12), Lemma 2.5 and (3.46). □

3.2 Time regularity of solutions to SDEs with localized L^p drifts

In this section, letting $T > 0$, we assume that

$$B \in \widetilde{\mathbb{L}}^{p_0}(T) \text{ for some } p_0 > d \text{ and } \sigma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d \text{ satisfy } (\mathbf{A}^\sigma). \tag{3.49}$$

and consider the following SDE on a probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathbb{P})$:

$$X_{s,t}^x = x + \int_s^t B(r, X_{s,r}^x)\mathrm{d}r + \int_s^t \sigma(X_{s,r}^x)\mathrm{d}W_r, \tag{3.50}$$

where W_t is a standard d-dimensional Brownian motion. Furthermore, consider the PDE on $[0,T] \times \mathbb{R}^d$

$$\partial_t u = a_{ij}\partial_i\partial_j u - \lambda u + B \cdot \nabla u + f, \quad u(0) = \varphi, \tag{3.51}$$

where $\lambda \geqslant 0$, $f \in \widetilde{\mathbb{L}}^{p_0}(T)$, $\varphi \in C_b^\infty$ and $a_{ij} := \frac{1}{2}\sum_{k=1}^d \sigma_{ik}\sigma_{jk}$. Under condition (3.49), by Lemma 3.4 there is a unique weak solution $X_{s,\cdot}^x$ to (3.50) for any $(s,x) \in \mathbb{R}_+ \times \mathbb{R}^d$. The purpose of this section is to obtain some moment estimates for the following functionals of $X_{0,t}^x$

$$\int_0^T f(s, X_{0,\pi_h(s)}^x)\mathrm{d}s \quad \text{and} \quad \int_0^T \left[f(s, X_{0,s}^x, \mu_s^x) - f(s, X_{0,\pi_h(s)}^x, \mu_{\pi_h(s)}^x)\right]\mathrm{d}s,$$

where μ_t^x is the distribution of $X_{0,t}^x$ and $f \in \mathbb{L}_q^p(T)$ for some $2/q+d/p < 2$. The first integral is estimated by Krylov's type estimates. Compared to the case of smooth coefficients in [92], f in the second integral has no regularity. To overcome this obstacle, we use the observation mentioned in the introduction to replace $f(X_{0,t}^x) - f(X_{0,\pi_h(t)}^x)$ by $P_{0,t}^X f - P_{0,\pi_h(s)}^X f$, where P^X is the transition semi-group of X. Hence, we only need to obtain some time regularity results for the semigroup.

In Subsection 3.2.1, we consider the time-homogeneous case with $B \equiv 0$. By Girsanov's theorem, we extend the results in Subsection 3.2.1 to $X_{s,t}^x$ in Subsection 3.2.2. Moreover, we obtain additional time regularity estimates for P^X by Duhamel's formula which can not be gotten from Girsanov's theorem. In the light of Duhamel's formula again, we also have two time regularity estimates for ∇u in Subsection 3.2.3, where u is the solution to (3.51).

For simplicity, throughout this section we set

$$\Xi := (d, T, p_0, \|B\|_{\widetilde{\mathbb{L}}^{p_0}(T)}, \kappa_1, \theta, c_0).$$

3.2.1 The zero-drift case

First of all, we recall the following generalized Itô formula from [109, Lemma 4.1-iii)] (see also [67, Theorem 3.7] for the original one).

Lemma 3.16 (Generalized Itô formula)**.** *Let $p, q \in [2, \infty)$ with $2/q + d/p < 1$. For any $T > 0$ and any $u \in \widetilde{\mathbb{H}}^{2,p}_q(T)$ with $\partial_t u \in \widetilde{\mathbb{L}}^p_q(T)$, we have for any $t \in [s, T]$ and $x \in \mathbb{R}^d$,*

$$\begin{aligned} u(t, X^x_{s,t}) = u(s, x) &+ \int_s^t (\partial_r u + a_{ij}\partial_i\partial_j u + B \cdot \nabla u)(r, X_{s,r}(x))\mathrm{d}r \\ &+ \int_s^t \nabla u(r, X_{s,r}(x))\mathrm{d}W_r. \end{aligned} \tag{3.52}$$

In this subsection, we consider the following case where $B \equiv 0$:

$$Z^x_t = x + \int_0^t \sigma(Z^x_s)\mathrm{d}W_s. \tag{3.53}$$

Define $P^\sigma_t f(x) := \mathbb{E} f(Z^x_t)$. By Proposition 2.10, there is a unique solution to the following second order parabolic PDE on $[0, T] \times \mathbb{R}^d$:

$$\partial_t u = a_{ij}\partial_i\partial_j u, \quad u_0 = \varphi. \tag{3.54}$$

Lemma 3.17. *Assume* $(\mathbf{A}^\sigma)$ *holds. Let $0 \leqslant s \leqslant t$, $\varphi \in C^\infty_b(\mathbb{R}^d)$, u and Z^x_t be the solution to (3.54) and (3.53) respectively. Then we have $\mathbb{P}$-a.s.*

$$\mathbb{E}\left[\varphi(Z^x_t)|\mathscr{F}_s\right] = u(t - s, Z^x_s). \tag{3.55}$$

In particular,

$$\mathbb{E}\left[\varphi(Z^x_t)|\mathscr{F}_s\right] = P^\sigma_{t-s}\varphi(Z^x_s) \quad \mathbb{P} - \text{a.s.} \tag{3.56}$$

Moreover, for any $t \geqslant 0$ and $f \in C^2_b$,

$$P^\sigma_t f - f = \int_0^t P^\sigma_r (a_{ij}\partial_i\partial_j f)\mathrm{d}r. \tag{3.57}$$

Proof. For all $t > 0$, applying the generalized Itô formula (3.52) to $s \mapsto u(t - s, Z^x_s)$, we have

$$\begin{aligned} u(0, Z^x_t) = u(t - s, Z^x_s) &+ \int_s^t \left(-\partial_r u(t - r, Z^x_r) + a_{ij}\partial_i\partial_j u(t - r, Z^x_r)\right)\mathrm{d}r \\ &+ \int_s^t \nabla u(t - r, Z^x_r)\mathrm{d}W_r. \end{aligned}$$

Noting that $u(t-s, Z_s^x)$ is $\mathscr{F}_s$-measurable, and taking conditional expectation with respect to $\mathscr{F}_s$ on both sides, we have

$$\mathbb{E}\left[\varphi(Z_t^x)|\mathscr{F}_s\right] = u(t-s, Z_s^x) \quad \mathbb{P}-\text{a.s.},$$

which for $s = 0$ implies that

$$P_t^\sigma \varphi(x) = \mathbb{E}\varphi(Z_t^x) = u(t, x).$$

Then (3.56) is straightforward from (3.55). For (3.57), since $f \in C_b^2$, we use the classical Itô formula and have

$$f(Z_t^x) = f(x) + \int_0^t a_{ij}\partial_i\partial_j f(Z_r^x)\mathrm{d}r + \int_0^t \nabla f(Z_r^x)\mathrm{d}W_r.$$

Then, we have (3.57) by taking expectation and complete the proof. □

Based on Lemma 3.17 and the uniqueness of (3.54), we have the following Chapman-Kolmogorov equations

$$P_s^\sigma P_t^\sigma = P_{s+t}^\sigma. \tag{3.58}$$

Recall

$$g_t(x) := (4\pi t)^{-d/2} e^{-|x|^2/(4t)}.$$

Then the following lemma is from [24, Theorem 2.3].

Lemma 3.18. *Assume* $(\mathbf{A}^\sigma)$ *holds. Then there is a unique function* $p_\cdot^\sigma(\cdot,\cdot) : \mathbb{R}_+ \times \mathbb{R}^{2d} \to \mathbb{R}$ *such that for any* $j = 0, 1, 2$

$$|\nabla_x^j p_t^\sigma(x, y)| \leqslant c_1 t^{-\frac{j}{2}} g_{c_2 t}(x - y) \tag{3.59}$$

and

$$P_t^\sigma f(x) = \int_{\mathbb{R}^d} f(y) p_t^\sigma(x, y)\mathrm{d}y \tag{3.60}$$

for any $f \in C(\mathbb{R}^d)$*, where* c_1 *and* c_2 *are positive constants depending on* Ξ.

For any $h \in (0, 1)$, recall that $\pi_h(t) := t$ for $t \in [0, h)$ and

$$\pi_h(t) := kh, \quad t \in [kh, (k+1)h), \; k \geqslant 1.$$

Remark 3.19. *The reason why we define* $\pi_h(t) = t$ *for* $t \in [0, h)$ *is that the function space here is* L^p*. If the initial data don't have an* L^q *density,* $\mathbb{E}f(Z_{\pi_h(t)}) = \mathbb{E}f(Z_0)$*,* $f \in L^p$*, will blow up for all* $t < h$.

Now we give the following Krylov estimate and Khasminskii estimate.

Lemma 3.20. *Assume* $(\mathbf{A}^\sigma)$ *holds. For any* $T>0$, $k=0,1,2$, $p\in[1,\infty]$ *and* $q\in[p,\infty]$, *there is a constant* $C=C(\Xi,p,q)$ *such that for all* $0\leqslant s<t\leqslant T$, $x\in\mathbb{R}^d$ *and nonnegative functions* $f\in\widetilde{L}^p$

$$\|\nabla^k P_t^\sigma f\|_q\leqslant Ct^{-k/2-d/(2p)+d/(2q)}\|f\|_p \tag{3.61}$$

and for $2/q+d/p<2$, $h>0$ *and nonnegative functions* $f\in\widetilde{\mathbb{L}}_q^p(T)$

$$\mathbb{E}\int_s^t f(r,Z_r^x)\mathrm{d}r+\mathbb{E}\int_s^t f(r,Z_{\pi_h(r)}^x)\mathrm{d}r\leqslant C(t-s)^{1-\frac{1}{q}-\frac{d}{2p}}\|f\|_{\widetilde{\mathbb{L}}_q^p(T)}. \tag{3.62}$$

Moreover, for any $f\in\widetilde{\mathbb{L}}_q^p(T)$ *with* $d/p+2/q<2$,

$$\sup_{x\in\mathbb{R}^d}\mathbb{E}\exp\left(\int_0^T f(t,Z_t^x)\mathrm{d}t\right)<\infty. \tag{3.63}$$

Proof. Without loss of generality we assume that $c_2=1$ in (3.59). Combining Lemma 3.18 and Young's convolution inequality, one sees that

$$\begin{aligned}
\|\nabla^k P_t^\sigma f\|_q&\lesssim\Big\|\int_{\mathbb{R}^d}f(y)\nabla_x^k p_t^\sigma(\cdot,y)\mathrm{d}y\Big\|_q\\
&\lesssim t^{-k/2}\|g_t*f\|_q\lesssim t^{-k/2}\sup_w\|\mathbf{1}_{|\cdot-w|\leqslant1}\int_{\mathbb{R}^d}f(\cdot-y)g_t(y)\mathrm{d}y\|_q\\
&\lesssim t^{-k/2}\sup_w\|\mathbf{1}_{|\cdot-w|\leqslant1}\frac{1}{|B_1|}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbf{1}_{|y-z|\leqslant1}f(\cdot-y)\mathbf{1}_{|y-z|\leqslant1}g_t(y)\mathrm{d}y\mathrm{d}z\|_q\\
&\lesssim t^{-k/2}\sup_w\|\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbf{1}_{|\cdot-y-w+z|\leqslant2}f(\cdot-y)\mathbf{1}_{|y-z|\leqslant1}g_t(y)\mathrm{d}y\mathrm{d}z\|_q\\
&\lesssim t^{-k/2}\sup_w\int_{\mathbb{R}^d}\|\int_{\mathbb{R}^d}\mathbf{1}_{|\cdot-y-w+z|\leqslant2}f(\cdot-y)\mathbf{1}_{|y-z|\leqslant1}g_t(y)\mathrm{d}y\|_q\mathrm{d}z\\
&\lesssim t^{-k/2}\int_{\mathbb{R}^d}\sup_w\|\mathbf{1}_{|\cdot-w+z|\leqslant2}f(\cdot)\|_p\|\mathbf{1}_{|\cdot-z|\leqslant1}g_t(\cdot)\|_r\mathrm{d}z\\
&\lesssim t^{-k/2}\int_{\mathbb{R}^d}\left(\int_{|y-z|\leqslant1}|g_t(y)|^r\mathrm{d}y\right)^{1/r}\mathrm{d}z\|f\|_p,
\end{aligned}$$

where $1+1/q=1/r+1/p$ and $B_1:\{x\in\mathbb{R}^d:|x|\leqslant1\}$. Next, one sees that

$$\int_{\mathbb{R}^d}\left(\int_{|y-z|\leqslant1}(g_t(y))^r\mathrm{d}y\right)^{1/r}\mathrm{d}z\lesssim\|g_t\|_r+\int_{|z|>2}\left(\int_{|y-z|\leqslant1}(g_t(y))^r\mathrm{d}y\right)^{1/r}\mathrm{d}z.$$

We note that

$$|z|>2, |y-z|\leqslant 1 \Rightarrow |y|\geqslant |z|-|y-z|\geqslant \frac{|z|}{2},$$

which implies that $g_t(y)\lesssim g_t(z/2)$, and $\|g_t\|_r\lesssim t^{-d/2+d/(2r)}=t^{-d/(2p)+d/(2q)}$, we have

$$\begin{aligned}\int_{\mathbb{R}^d}\Big(\int_{|y-z|\leqslant 1}(g_t(y))^r\mathrm{d}y\Big)^{1/r}\mathrm{d}z &\lesssim \|g_t\|_r+\int_{|z|>2}\Big(\int_{|y-z|\leqslant 1}(g_t(z/2))^r\mathrm{d}y\Big)^{1/r}\mathrm{d}z\\ &\lesssim t^{-d/(2p)+d/(2q)}+\int_{|z|>2}g_t(z/2)\mathrm{d}z\lesssim t^{-d/(2p)+d/(2q)}+1\end{aligned}$$

and obtain (3.61).

Now we show (3.62). Set $p':=\frac{p}{p-1}$ and $q':=\frac{q}{q-1}$. Without loss of generality, we take $s=0$. By (3.61), for any $h>0$,

$$\mathbb{E}\int_0^t f(s,Z^x_{\pi_h(s)})\mathrm{d}s\lesssim\int_{\mathbb{R}^d}\Big(\int_0^t\big(\int_{|y-z|\leqslant 1}|g_{\pi_h(s)}(y)|^{p'}\mathrm{d}y\big)^{q'/p'}\mathrm{d}s\Big)^{1/q'}\mathrm{d}z|\!|\!|f|\!|\!|_{\widetilde{\mathbb{L}}^p_q(T)}.$$

Then, we have

$$\begin{aligned}\mathscr{I}:=&\int_{\mathbb{R}^d}\Big(\int_0^t\big(\int_{|y-z|\leqslant 1}|g_{\pi_h(s)}(y)|^{p'}\mathrm{d}y\big)^{q'/p'}\mathrm{d}s\Big)^{1/q'}\mathrm{d}z\\ \lesssim&\big(\int_0^t\|g_{\pi_h(s)}\|_{p'}^{q'}\mathrm{d}s\big)^{1/q'}+\int_{|z|>2}\Big(\int_0^t|g_{\pi_h(s)}(z/2)|^{q'}\mathrm{d}s\Big)^{1/q'}\mathrm{d}z\\ \lesssim&\big(\int_0^t(\pi_h(s))^{-dq'/(2p)}\mathrm{d}s\big)^{1/q'}+t\int_{|z|>2}|z|^{-d}\exp(-|z|^2/(16T))\mathrm{d}z\\ \lesssim&\, t^{1-1/q-d/(2p)}+t,\end{aligned}$$

since $dq'/(2p)<1$ and $\rho_s(z)\leqslant C|z|^{-d}\exp(-\frac{|z|^2}{4T})$ for all $s<T$. Similarly, we obtain

$$\mathbb{E}\int_0^t f(s,Z^x_s)\mathrm{d}s\lesssim t^{1-\frac{1}{q}-\frac{d}{2p}}+t.$$

Finally, noting that by (3.56)

$$\mathbb{E}\Big[\int_s^t f(s,Z^x_r)\mathrm{d}r\Big|\mathscr{F}_s\Big]=\mathbb{E}\int_s^t f(s,Z^y_{r-s})\mathrm{d}r\Big|_{y=Z^x_s},$$

(3.63) is direct from (3.62) (see [116, Corollary 3.5] for example) and we complete the proof.

□

Remark 3.21. *(i)We note that for any fixed $h > 0$ and $x \in \mathbb{R}^d$*

$$\mathbb{E}\exp\Big(\int_0^T f(t, Z^x_{\pi_h(t)})\mathrm{d}t\Big) < \infty$$

is not true. For example, letting $Z^x_t = W_t$, when $d \geqslant 2$ and $f(t,x) = |x|^{-1/2} \in \widetilde{L}^{d+1}(\mathbb{R}^d)$, we have

$$\mathbb{E}\exp\left(\int_h^{2h} f(t, W_{\pi_h(t)})\mathrm{d}t\right) = \mathbb{E}\exp(h|W_h|^{-1/2})$$
$$= \int_{\mathbb{R}^d} e^{\frac{h}{\sqrt{|x|}}} g_h(x)\mathrm{d}x \geqslant \frac{h^{2d}}{(2d)!}\int_{\mathbb{R}^d} \frac{1}{|x|^d} g_h(x)\mathrm{d}x = \infty.$$

(ii) The Krylov estimate (3.4) is obtained from backward PDE, but here (3.62) is from the heat kernel estimates.

Moreover, we have the following results from the Schauder estimate Lemma A.6.

Lemma 3.22. *Assume $(\mathbf{A}^\sigma)$ holds. Then, for any $T > 0$, there is a constant $C = C(\Xi)$ such that for any $\varphi \in \mathbf{C}^\infty_b$ and $t \in (0,T]$,*

$$\|P^\sigma_t \varphi\|_{\mathcal{C}^{2+\theta}} \leqslant Ct^{-\frac{2+\theta}{2}}\|\varphi\|_\infty, \tag{3.64}$$

where θ is the parameter in $(\mathbf{A}^\sigma)$ and for any $\alpha \in (0,1)$ there is a constant $C = C(\Xi,\alpha)$ such that for any $\varphi \in \mathbf{C}^\infty_b$ and $t \in (0,T]$,

$$\|\nabla P^\sigma_t \varphi\|_\infty \leqslant Ct^{-\frac{1-\alpha}{2}}\|\varphi\|_{\mathcal{C}^\alpha}, \tag{3.65}$$

.

Proof. Letting $u(t,x) := P^\sigma_t\varphi$, by (3.55) we have

$$\partial_t u = a_{ij}\partial_i\partial_j u, \quad u_0 = \varphi.$$

We set $v(t) := u(t) - P_t\varphi$, where $P_t f = g_t * f$. Then, v satisfier the following PDE:

$$\partial_t v = a_{ij}\partial_i\partial_j v + f, \quad v_0 = 0,$$

where

$$f(t) := -\partial_t P_t\varphi + a_{ij}\partial_i\partial_j P_t\varphi = -\Delta P_t\varphi + a_{ij}\partial_i\partial_j P_t\varphi$$

satisfying

$$\|f(t)\|_{\mathcal{C}^\theta} \leqslant (1 + \|a\|_\theta)\|P_t\nabla^2\varphi\|_{\mathcal{C}^\theta} \lesssim t^{-\frac{\theta}{2}}\|\nabla^2\varphi\|_\infty$$

because of $\partial_t P_t = \Delta P_t$ and (A.3.3). Then, based on Theorem A.6, we have

$$\|v(t)\|_{\mathcal{C}^{2+\theta}} \lesssim t^{-\frac{\theta}{2}}\|\nabla^2\varphi\|_\infty,$$

which implies that

$$\|P_t^\sigma\varphi\|_{\mathcal{C}^{2+\theta}} \lesssim t^{-\frac{\theta}{2}}\|\nabla^2\varphi\|_\infty + \|P_t\varphi\|_{\mathcal{C}^{2+\theta}} \lesssim t^{-\frac{\theta}{2}}\|\nabla^2\varphi\|_\infty.$$

Taking $\varphi = P_t^\sigma\varphi$, by (3.58) and (3.61) we have

$$\|P_{2t}^\sigma\varphi\|_{\mathcal{C}^{2+\theta}} \lesssim t^{-\frac{\theta}{2}}\|\nabla^2 P_t^\sigma\varphi\|_\infty \lesssim t^{-\frac{2+\theta}{2}}\|\varphi\|_\infty,$$

which is (3.64).

For (3.64), we note that $\nabla P_t^\sigma 1 = 0$ and have

$$|\nabla P_t^\sigma\varphi(x)| = \left|\int_{\mathbb{R}^d}\nabla_x p_t^\sigma(x,y)(\varphi(y)-\varphi(x))\mathrm{d}y\right| \leqslant \|\varphi\|_{\mathcal{C}^\alpha}\int_{\mathbb{R}^d}|\nabla_x p_t^\sigma(x,y)||x-y|^\alpha\mathrm{d}y.$$

Since (3.59), one sees that

$$|\nabla P_t^\sigma\varphi(x)| \lesssim t^{-\frac{1}{2}}\|\varphi\|_{\mathcal{C}^\alpha}\int_{\mathbb{R}^d}|g_{c_2t}(y)||y|^\alpha\mathrm{d}y \lesssim t^{-\frac{1-\alpha}{2}}\|\varphi\|_{\mathcal{C}^\alpha},$$

because of the scaling of $g_t(y) = t^{-d/2}g_1(t^{-1/2}y)$. This completes the proof. □

Lemma 3.23. *Assume* $(\mathbf{A}^\sigma)$ *holds. Let* $T>0$, $k=0,1$, *and* $1\leqslant p\leqslant q\leqslant\infty$. *Then there is a constant* $C=C(\Xi,p,q)$ *such that for all* $0<s\leqslant t\leqslant T$,

$$\|\nabla^k(P_t^\sigma\varphi - P_s^\sigma\varphi)\|_q \leqslant C\Big([(t-s)^{\frac{2-k}{2}}s^{\frac{k-2}{2}}]\wedge 1\Big)s^{-\frac{k}{2}-\frac{d}{2p}+\frac{d}{2q}}\|\varphi\|_p. \tag{3.66}$$

In particular,

$$\|\nabla(P_t^\sigma\varphi - P_s^\sigma\varphi)\|_\infty \leqslant C\Big([(t-s)^{\frac{1+\theta}{2}}s^{\frac{-1-\theta}{2}}]\wedge 1\Big)s^{-\frac{1}{2}}\|\varphi\|_\infty. \tag{3.67}$$

Proof. Based on (3.58) and (3.57), one sees that

$$P_t^\sigma\varphi - P_s^\sigma\varphi = P_{t-s}^\sigma(P_s^\sigma\varphi) - P_s^\sigma\varphi = \int_0^{t-s}P_r^\sigma(a_{ij}\partial_i\partial_j P_s^\sigma\varphi)\mathrm{d}r. \tag{3.68}$$

By (3.61), we have

$$\begin{aligned}\|\nabla^k(P_t^\sigma\varphi - P_s^\sigma\varphi)\|_q &\lesssim \int_0^{t-s} r^{-\frac{k}{2}}\|\nabla^2 P_s^\sigma\varphi\|_q\mathrm{d}r\\ &\lesssim \int_0^{t-s} r^{-\frac{k}{2}}s^{-1+\frac{d}{2q}-\frac{d}{2p}}\|\varphi\|_p\mathrm{d}r\\ &\lesssim \Big[(t-s)^{\frac{2-k}{2}}s^{\frac{k-2}{2}}\Big]s^{-\frac{k}{2}-\frac{d}{2p}+\frac{d}{2q}}\|\varphi\|_p.\end{aligned}$$

Moreover, noting that by (3.61)

$$\begin{aligned}|||\nabla^k(P^\sigma_t\varphi - P^\sigma_s\varphi)|||_q &\leqslant |||\nabla^k P^\sigma_t\varphi|||_q + |||\nabla^k P^\sigma_s\varphi|||_q \\ &\lesssim s^{-\frac{k}{2}-\frac{d}{2p}+\frac{d}{2q}}|||\varphi|||_p\end{aligned}$$

for $s \leqslant t$, we have (3.66).

For (3.67), we only need to show

$$\|\nabla(P^\sigma_t\varphi - P^\sigma_s\varphi)\|_\infty \leqslant C(t-s)^{\frac{1+\theta}{2}} s^{-1-\frac{\theta}{2}}\|\varphi\|_\infty.$$

Based on (3.68), by (3.65) and (3.64), we have

$$\begin{aligned}\|\nabla(P^\sigma_t\varphi - P^\sigma_s\varphi)\|_\infty &\lesssim \int_0^{t-s} r^{-\frac{1-\theta}{2}}\|a\|_{\mathcal{C}^\theta}\|\nabla^2 P^\sigma_s\varphi\|_{\mathcal{C}^\theta}\mathrm{d}r \\ &\lesssim \int_0^{t-s} r^{-\frac{1-\theta}{2}} s^{-1-\frac{\theta}{2}}\|\varphi\|_\infty \mathrm{d}r \\ &\lesssim (t-s)^{\frac{1+\theta}{2}} s^{-1-\frac{\theta}{2}}\|\varphi\|_\infty\end{aligned}$$

and complete the proof. □

When $p = \infty$ and $\sigma \equiv \mathbb{I}$, the following lemma has been proved in [28, Lemma 2.1] for Brownian motion. For classical L^p spaces, Lê and Ling obtained these results by the stochastic sewing lemma in [73]. For the localized L^p space, we provide a different proof here, which is based on (3.66) and (3.61).

Lemma 3.24. *Assume* $(\mathbf{A}^\sigma)$ *holds. Then for any* $T > 0$ *and* $p \in (d \vee 2, \infty)$*, there is a constant* $C = C(\Xi, p)$ *such that for any stopping time* $\tau \leqslant T$*,* $h \in (0,1)$*,* $x \in \mathbb{R}^d$ *and* $f \in \widetilde{\mathbb{L}}^p(T)$*,*

$$\mathbb{E}\left|\int_0^\tau (f(t, Z^x_t) - f(t, Z^x_{\pi_h(t)}))\mathrm{d}t\right|^2 \leqslant Ch\log h^{-1}|||f|||^2_{\widetilde{\mathbb{L}}^p(T)}. \tag{3.69}$$

Proof. Set $F_h(r) := f(r, Z^x_r) - f(r, Z^x_{\pi_h(r)})$. We divide the proof into two steps. In Step 1, we prove that

$$\mathbb{E}\left|\int_s^t F_h(r)\mathrm{d}r\right|^2 \leqslant C\big(hs^{-\frac{d}{2p}}(t-s)^{1-\frac{d}{2p}} + h\log h^{-1}(t-s)^{1-\frac{d}{p}}\big)|||f|||^2_{\widetilde{\mathbb{L}}^p(T)} \tag{3.70}$$

and obtain (3.69) for $\tau = T$ by Lemma A.1; In Step 2, we show (3.69) for any stopping time τ with $\tau \leqslant T$.

(Step 1) First, we note that by Hölder's inequality and (3.62),

$$\mathbb{E}\left(\int_0^{2h} F_h(t)\mathrm{d}t\right)^2 \leqslant 2h\mathbb{E}\int_0^{2h} |F_h(t)|^2\,\mathrm{d}t \lesssim h|\!|\!| f|\!|\!|_p.$$

Hence, without loss of generality, we may assume $s > 2h$. The symmetry implies

$$\begin{aligned}\mathbb{E}\Big|\int_s^t F_h(r)\mathrm{d}r\Big|^2 =&2\int_s^t\int_{r_1}^t \mathbb{E}\left(F_h(r_1)F_h(r_2)\right)\mathrm{d}r_2\mathrm{d}r_1\\ =&2\left(\int_s^t\int_{r_1}^{r_1+h}+\int_s^t\int_{r_1+h}^t\right)\mathbb{E}\left(F_h(r_1)F_h(r_2)\right)\mathrm{d}r_2\mathrm{d}r_1 =: 2\mathscr{I}_1+2\mathscr{I}_2.\end{aligned}$$

By Hölder's inequality and (3.61), one sees that

$$\begin{aligned}\mathbb{E}\left(F_h(r_1)F_h(r_2)\right) \leqslant& \left(\mathbb{E}\left|F_h(r_1)\right|^2\right)^{1/2}\left(\mathbb{E}\left|F_h(r_2)\right|^2\right)^{1/2}\\ &\lesssim (r_1^{-\frac{d}{2p}}+\pi_h(r_1)^{-\frac{d}{2p}})(r_2^{-\frac{d}{2p}}+\pi_h(r_2)^{-\frac{d}{2p}})|\!|\!| f|\!|\!|_p^2,\end{aligned}$$

which implies that

$$\begin{aligned}\mathscr{I}_1 &\lesssim |\!|\!| f|\!|\!|_p^2\int_s^t\int_{r_1}^{r_1+h}(r_1^{-\frac{d}{2p}}+\pi_h(r_1)^{-\frac{d}{2p}})(r_2^{-\frac{d}{2p}}+\pi_h(r_2)^{-\frac{d}{2p}})\mathrm{d}r_2\mathrm{d}r_1\\ &\lesssim |\!|\!| f|\!|\!|_p^2(s-h)^{-\frac{d}{2p}}h\int_s^t(r_1^{-\frac{d}{2p}}+\pi_h(r_1)^{-\frac{d}{2p}})\mathrm{d}r_1 \lesssim hs^{-\frac{d}{2p}}(t-s)^{1-\frac{d}{2p}}|\!|\!| f|\!|\!|_p^2.\end{aligned}$$

For $\mathscr{I}_2$, we use conditional expectation and the Markov property (3.56). For simplicity, let

$$\mathbb{E}^{\mathscr{G}}[\cdot] := \mathbb{E}[\cdot|\mathscr{G}].$$

Then, noting that $r_1 \leqslant r_2 - h < r_2$, by the Markov property (3.56), we have

$$\begin{aligned}\mathscr{I}_2 &= \int_s^t\int_{r_1+h}^t \mathbb{E}\Big(F_h(r_1)\mathbb{E}^{\mathscr{F}_{r_1}}[f(Z^x_{r_2})-f(Z^x_{\pi_h(r_2)})]\Big)\mathrm{d}r_2\mathrm{d}r_1\\ &= \int_s^t\int_{r_1+h}^t \mathbb{E}\Big(F_h(r_1)\Big[P^\sigma_{r_2-r_1}f(Z^x_{r_1})-P^\sigma_{\pi_h(r_2)-r_1}f(Z^x_{r_1})\Big]\Big)\mathrm{d}r_2\mathrm{d}r_1.\end{aligned}$$

Setting $G := P^\sigma_{r_2-r_1}f - P^\sigma_{\pi_h(r_2)-r_1}f$, in view of Hölder's inequality and (3.61), one sees that

$$\begin{aligned}\mathscr{I}_2 &\leqslant \int_s^t\int_{r_1+h}^t\Big(\mathbb{E}|F_h(r_1)|^2\Big)^{1/2}\Big(\mathbb{E}|G(Z^x_{r_1})|^2\Big)^{1/2}\mathrm{d}r_2\mathrm{d}r_1\\ &\lesssim |\!|\!| |f|^2|\!|\!|_{p/2}^{1/2}\int_s^t\int_{r_1+h}^t(r_1^{-\frac{d}{2p}}+\pi_h(r_1)^{-\frac{d}{2p}})r_1^{-\frac{d}{2p}}|\!|\!| |G|^2|\!|\!|_{p/2}^{1/2}\mathrm{d}r_2\mathrm{d}r_1\\ &\lesssim |\!|\!| f|\!|\!|_p\int_s^t\int_{r_1+h}^t(r_1-h)^{-\frac{d}{p}}|\!|\!| G|\!|\!|_p\mathrm{d}r_2\mathrm{d}r_1.\end{aligned}$$

We note that by (3.66),

$$\begin{aligned}\|G\|_p &= \|P^\sigma_{r_2-r_1}f - P^\sigma_{\pi_h(r_2)-r_1}f\|_p \lesssim \Big([(r_2-\pi_h(r_2))(\pi_h(r_2)-r_1)^{-1}]\wedge 1\Big)\|f\|_p\\ &\lesssim \Big[(h(r_2-r_1-h)^{-1})\wedge 1\Big]\|f\|_p,\end{aligned}$$

By a change of variables we have

$$\begin{aligned}\mathscr{I}_2 &\lesssim \|f\|_p^2\int_s^t\int_{r_1+h}^t (r_1-h)^{-\frac{d}{p}}\Big[(h(r_2-r_1-h)^{-1})\wedge 1\Big]\mathrm{d}r_2\mathrm{d}r_1\\ &\lesssim \|f\|_p^2\int_{s-h}^{t-h}\int_0^t (r_1)^{-\frac{d}{p}}\Big[(h(r_2)^{-1})\wedge 1\Big]\mathrm{d}r_2\mathrm{d}r_1\\ &\lesssim h(t-s)^{1-\frac{d}{p}}\|f\|_p^2\int_0^{t/h}\Big[(r_2)^{-1}\wedge 1\Big]\mathrm{d}r_2 \lesssim h\log h^{-1}(t-s)^{1-\frac{d}{p}}\|f\|_p^2,\end{aligned}$$

and we obtain (3.70). Then, by Lemma A.1, we have (3.69) with $\tau = T$. In particular, for any $a, b \in [0,T]$, defining $f_{a,b}(t,x) := \mathbf{1}_{t\in[a,b]}f(t,x)$, by Step 2, one sees that

$$\begin{aligned}\sup_{a,b\in[0,T]}\mathbb{E}\left(\int_a^b (f(t,Z^x_t)-f(t,Z^x_{\pi_h(t)}))\mathrm{d}t\right)^2 &= \mathbb{E}\Big(\int_0^T (f_{a,b}(t,Z^x_t)-f_{a,b}(t,Z^x_{\pi_h(t)}))\mathrm{d}t\Big)^2\\ &\leqslant Ch\log h^{-1}\|f_{a,b}\|^2_{\widetilde{\mathbb{L}}^p(T)} \leqslant Ch\log h^{-1}\|f\|_{\widetilde{\mathbb{L}}^p(T)}.\end{aligned} \tag{3.71}$$

(Step 2) Without loss of generality, we assume that $h < T/2$ and that τ only takes finite values $a_1, a_2, ..., a_n \in [0,T]$. Otherwise, for any stopping time, we choose $\tau_n, n \in \mathbb{N}$, which only take finite values to approximate τ and (3.69) follows from (3.62) and the dominated convergence theorem. First, we have

$$\mathbb{E}\left|\int_\tau^T F_h(t)\mathrm{d}t\right|^2 \leqslant 2\mathbb{E}\left|\int_\tau^{(\tau+2h)\wedge T} F_h(t)\mathrm{d}t\right|^2 + 2\mathbb{E}\left|\int_{(\tau+2h)\wedge T}^T F_h(t)\mathrm{d}t\right|^2.$$

We note that by Hölder's inequality and (3.62) for $p/2 > d/2$,

$$\mathbb{E}\Big|\int_\tau^{(\tau+2h)\wedge T} F_h(t)\mathrm{d}t\Big|^2 \leqslant \mathbb{E}\Big|\int_0^T \mathbf{1}_{t\in[\tau,\tau+2h]}\mathrm{d}t\int_0^T |F_h(t)|^2\mathrm{d}t\Big| \lesssim h\||f|^2\|_{\widetilde{\mathbb{L}}^{p/2}(T)} \lesssim h\|f\|^2_{\widetilde{\mathbb{L}}^p(T)}.$$

Now we estimate the second term. In fact,

$$\mathbb{E}\left(\int_{(\tau+2h)\wedge T}^T F_h(t)\mathrm{d}t\right)^2 = \sum_{i=1}^n \mathbb{E}\left[\mathbf{1}_{\tau=a_i}\left(\int_{(a_i+2h)\wedge T}^T F_h(t)\mathrm{d}t\right)^2\right].$$

Noting that $\mathbf{1}_{\tau=a_i} \in \mathscr{F}_{a_i} \subset \mathscr{F}_{([a_i/h]+1)h}$, without loss of generality, assuming $a_i + 2h < T$, one sees that

$$\mathbb{E}\left[\mathbf{1}_{\tau=a_i}\left(\int_{a_i+2h}^T F_h(t)\mathrm{d}t\right)^2\right] = \mathbb{E}\Big(\mathbf{1}_{\tau=a_i}\mathbb{E}^{\mathscr{F}_{a_i}}\left(\int_{a_i+2h}^T F_h(t)\mathrm{d}t\right)^2\Big) := \mathbb{E}(\mathbf{1}_{\tau=a_i}\mathscr{A}_i),$$

where $\mathscr{A}_i = \mathbb{E}^{\mathscr{F}_{a_i}}\mathbb{E}^{\mathscr{F}_{([a_i/h]+1)h}}(\int_{a_i+2h}^T F_h(t)\mathrm{d}t)^2 =: \mathbb{E}^{\mathscr{F}_{a_i}}\mathscr{B}_i$. Moreover, by the Markov property (3.56), we have

$$\begin{aligned}
\mathscr{B}_i &= \mathbb{E}\Big(\int_{a_i+2h}^T (f(t, Z^y_{t-([a_i/h]+1)h}) - f(t, Z^y_{\pi_h(t)-([a_i/h]+1)h}))\mathrm{d}t\Big)^2\Big|_{y=Z^x_{([a_i/h]+1)h}} \\
&= \mathbb{E}\Big(\int_{a_i+2h-([a_i/h]+1)h}^{T-([a_i/h]+1)h} (f(t+([a_i/h]+1)h, Z^y_t) \\
&\qquad\qquad - f(t+([a_i/h]+1)h, Z^y_{\pi_h(t)}))\mathrm{d}t\Big)^2\Big|_{y=Z^x_{([a_i/h]+1)h}} \\
&\lesssim h\log h^{-1}\|f\|^2_{\widetilde{\mathbb{L}}^p(T)}
\end{aligned}$$

by (3.71). Therefore, we have

$$\begin{aligned}
\mathbb{E}\Big|\int_\tau^T F_h(t)\mathrm{d}t\Big|^2 &\lesssim h\|f\|^2_{\widetilde{\mathbb{L}}^p(T)} + \sum_{i=1}^n \mathbb{E}(\mathbf{1}_{\tau=a_i}\mathscr{A}_i) \\
&\lesssim h\|f\|^2_{\widetilde{\mathbb{L}}^p(T)} + h\log h^{-1}\|f\|^2_{\widetilde{\mathbb{L}}^p(T)}\sum_{i=1}^n \mathbb{E}\mathbf{1}_{\tau=a_i} \lesssim h\log h^{-1}\|f\|^2_{\widetilde{\mathbb{L}}^p(T)}.
\end{aligned}$$

Noting that $\int_0^\tau = \int_0^T - \int_\tau^T$, we have

$$\mathbb{E}\left|\int_0^\tau F_h(t)\mathrm{d}t\right|^2 \lesssim h\log h^{-1}\|f\|^2_{\widetilde{\mathbb{L}}^p(T)}$$

and complete the proof. □

Corollary 3.25. *Assume* $(\mathbf{H}^\sigma)$ *holds. For any* $T > 0$, $p \in (d \vee 2, \infty)$ *and* $\delta > 0$, *there is a constant* $C = C(\Xi, p)$ *such that for any* $x \in \mathbb{R}^d$ *and* $f \in \widetilde{\mathbb{L}}^p(T)$,

$$\mathbb{E}\left(\sup_{t\in[0,T]}\left|\int_0^t (f(s, Z^x_s) - f(s, Z^x_{\pi_h(s)}))\mathrm{d}s\right|^2\right) \leqslant Ch^{1-\delta}\|f\|^2_{\widetilde{\mathbb{L}}^p(T)}. \tag{3.72}$$

Proof. Let

$$\eta_t := \left|\int_0^t (f(s, Z^x_s) - f(s, Z^x_{\pi_h(s)}))\mathrm{d}s\right|^2$$

and $\eta_t^* := \sup_{s\in[0,t]} \eta_s$. First of all, it follows from Hölder's inequality and (3.62) that for any $\gamma \in (1, p/(d\vee 2))$,

$$\begin{aligned}\mathbb{E}(\eta_T^*)^\gamma &\lesssim \mathbb{E}\int_0^T \big(|f|^{2\gamma}(s, Z_s^x) + |f|^{2\gamma}(s, Z_{\pi_h(s)}^x)\big)\,\mathrm{d}s \\ &\lesssim \| |f|^{2\gamma}\|_{\widetilde{\mathbb{L}}^{p/(2\gamma)}(T)} \lesssim \|f\|_{\widetilde{\mathbb{L}}^p(T)}^{2\gamma}.\end{aligned} \tag{3.73}$$

For any $\lambda > 0$, let

$$\tau_\lambda := \inf\{t \geqslant 0,\ \eta_t > \lambda\}.$$

We note that $\eta_{\tau_\lambda} = \lambda$, since η is a continuous process. Then,

$$\lambda\mathbb{P}(\eta_T^* > \lambda) \leqslant \lambda\mathbb{P}(\tau_\lambda \leqslant T) \leqslant \mathbb{E}\big(\eta_{\tau_\lambda}\mathbf{1}_{\{\tau_\lambda\leqslant T\}}\big) \leqslant \mathbb{E}\eta_{\tau_\lambda\wedge T}.$$

In view of (3.69), we have

$$\lambda\mathbb{P}(\eta_T^* > \lambda) \lesssim h\log h^{-1}\|f\|_{\widetilde{\mathbb{L}}^p(T)}^2.$$

Set $\Xi_h := h\log h^{-1}\|f\|_{\widetilde{\mathbb{L}}^p(T)}^2$. Then, for any $\delta \in (0,1)$, by a change of variables,

$$\begin{aligned}\mathbb{E}\big(\eta_T^*\big)^{1-\delta} &= (1-\delta)\int_0^\infty \lambda^{-\delta}\mathbb{P}(\eta_T^* > \lambda)\mathrm{d}\lambda \\ &\lesssim \int_0^\infty \lambda^{-\delta}\big(1\wedge(\Xi_h\lambda^{-1})\big)\mathrm{d}\lambda \\ &\lesssim \Xi_h^{1-\delta}\int_0^\infty \lambda^{-\delta}(1\wedge\lambda^{-1})\mathrm{d}\lambda \lesssim (h\log h^{-1})^{(1-\delta)}\|f\|_{\widetilde{\mathbb{L}}^p(T)}^{2(1-\delta)}.\end{aligned} \tag{3.74}$$

Combining (3.73) and (3.74), in view of Hölder's inequality, for any $\delta > 0$ small enough, we have

$$\begin{aligned}\mathbb{E}\eta_T^* &= \mathbb{E}\left[\big(\eta_T^*\big)^{1-\delta-\sqrt{\delta}}\big(\eta_T^*\big)^{\delta+\sqrt{\delta}}\right] \\ &\lesssim \left[\mathbb{E}\big(\eta_T^*\big)^{1-\delta}\right]^{\frac{1-\delta-\sqrt{\delta}}{1-\delta}}\left[\mathbb{E}\big(\eta_T^*\big)^{(\sqrt{\delta}+1)(1-\delta)}\right]^{\frac{\sqrt{\delta}}{1-\delta}} \\ &\lesssim h^{1-\delta-\sqrt{\delta}}(\log h^{-1})^{1-\delta-\sqrt{\delta}}\|f\|_{\widetilde{\mathbb{L}}^p(T)}^2\end{aligned}$$

and complete the proof.

□

3.2.2 Time discretization for SDEs with localized L^p drift

Now, let us extend estimate (3.72) from Z^x_{t-s} to the solution $X^x_{s,t}$ of SDE (3.50) both in the sense of paths and distributions (see (3.79) and (3.91) below). Recall that

$$X^x_{s,t} = x + \int_s^t B(r, X^x_{s,r})\mathrm{d}r + \int_s^t \sigma(X^x_{s,r})\mathrm{d}W_r. \tag{3.75}$$

Let $\mu^x_{s,t}$ denote the distribution of $X^x_{s,t}$. For simplicity, we also set

$$X^x_t := X^x_{0,t} \quad \text{and} \quad \mu^x_t := \mu^x_{0,t}.$$

The following estimates follow from Girsanov's transform and estimates for Z^x_t (see [109, Lemma 4.1] for (ii) and (iii)).

Lemma 3.26. *Assume (3.49) with $p_0 > d \vee 2$.*
(i) For any $T > 0$ and $p \in (1, \infty)$, there is a constant $C = C(\Xi, p)$ such that for any $x \in \mathbb{R}^d$, $0 \leqslant s < t \leqslant T$ and nonnegative $f \in \widetilde{L}^p$,

$$\mathbb{E} f(X^x_t) \leqslant C t^{-d/(2p)} \|\!| f \|\!|_p. \tag{3.76}$$

(ii)For any $T > 0$ and $2/q + d/p < 2$, there is a constant $C = C(\Xi, p, q)$ such that for any $x \in \mathbb{R}^d$, $0 \leqslant s < t \leqslant T$ and nonnegative $f \in \widetilde{\mathbb{L}}^p_q(T)$,

$$\mathbb{E}\int_s^t f(r, X^x_r)\mathrm{d}r + \mathbb{E}\int_s^t f(r, X^x_{\pi_h(r)})\mathrm{d}r \leqslant C(t-s)^{1-\frac{1}{q}-\frac{d}{2p}} \|\!| f \|\!|_{\widetilde{\mathbb{L}}^p_q(T)}. \tag{3.77}$$

(iii)For any $T > 0$, $d/p + 2/q < 2$ and $f \in \widetilde{\mathbb{L}}^p_q(T)$,

$$\sup_{x\in\mathbb{R}^d} \mathbb{E}\exp\left(\int_0^T f(t, X^x_t)\mathrm{d}t\right) < \infty. \tag{3.78}$$

(iv)For any $T > 0$, $\delta > 0$ and $p > d \vee 2$, there is a constant $C = C(\Xi, p)$ such that for any $x \in \mathbb{R}^d$, $h \in (0,1)$ and $f \in \widetilde{\mathbb{L}}^p(T)$,

$$\mathbb{E}\left(\sup_{t\in[0,T]}\left|\int_0^t (f(s, X^x_s) - f(s, X^x_{\pi_h(s)}))\mathrm{d}s\right|^2\right) \leqslant C h^{1-\delta} \|\!| f \|\!|^2_{\widetilde{\mathbb{L}}^p(T)}. \tag{3.79}$$

Proof. Let $\widetilde{Z}^x$ be a solution on a probability space $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, (\widetilde{\mathscr{F}}_t)_{t\geqslant 0}, \widetilde{\mathbb{P}})$ to the following SDE

$$\widetilde{Z}^x_t = x + \int_0^t \sigma(\widetilde{Z}^x_r)\mathrm{d}\widetilde{W}_r,$$

where $\widetilde{W}_t$ is a standard d-dimensional Brownian motion. Since $p_0 > 2 \vee d$, $B^2 \in \widetilde{\mathbb{L}}^{p_0/2}(T)$. By (3.63), one sees that for any $\gamma > 0$

$$\sup_x \widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}} \exp\left(\gamma \int_0^T |\sigma^{-1} B(t, \widetilde{Z}_t^x)|^2 \mathrm{d}t\right) < \infty.$$

Hence, by Novikov's criterion,

$$\mathbb{Z}_T := \exp\left(-\int_0^T \sigma^{-1} B(t, \widetilde{Z}_t^x) \mathrm{d}\widetilde{W}_t - \frac{1}{2}\int_0^T |\sigma^{-1} B(t, \widetilde{Z}_t^x)|^2 \mathrm{d}t\right)$$

is integrable and for any $q > 0$

$$\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}} |\mathbb{Z}_T|^q \leqslant C(\Xi, q). \tag{3.80}$$

Define $\mathrm{d}\mathbb{Q} := \mathbb{Z}_T \mathrm{d}\widetilde{\mathbb{P}}$. Then, by Girsanov's theorem,

$$\bar{W}_t := \widetilde{W}_t - \int_0^t \sigma^{-1} B(s, \widetilde{Z}_s^x) \mathrm{d}s \text{ is a } \mathbb{Q}\text{-martingale.}$$

In other words,

$$\widetilde{Z}_t^x = x + \int_0^t B(s, \widetilde{Z}_s) \mathrm{d}s + \int_0^t \sigma(\widetilde{Z}_s) \mathrm{d}\bar{W}_s, \quad \mathbb{Q} - a.e.$$

Therefore, by the uniqueness of (3.75), we have

$$\mathbb{Q} \circ (\widetilde{Z}_\cdot^x)^{-1} = \mathbb{P} \circ (X_\cdot^x)^{-1}. \tag{3.81}$$

Now, we show (i)-(iv) one by one.

(i): In view of (3.81), one sees that

$$\mathbb{E} f(X_t^x) = \widetilde{\mathbb{E}}_{\mathbb{Q}} f(\widetilde{Z}_t^x) = \widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\Big[\mathbb{Z}_T f(\widetilde{Z}_t)\Big].$$

By Hölder's inequality, (3.80) and (3.61), we have for any $r \in (1, p)$ and $1/r' + 1/r = 1$,

$$\mathbb{E} f(X_t^x) \leqslant \left(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}} |\mathbb{Z}_T|^{r'}\right)^{1/r'} \left(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}} |f(\widetilde{Z}_t)|^r\right)^{1/r} \lesssim \left(t^{-dr/(2p)} \| |f|^r \|_{p/r}\right)^{1/r} \lesssim t^{-d/(2p)} \|f\|_p,$$

which is (3.76).

(ii): Similarly to (i), by Hölder's inequality and (3.62), we have

$$\begin{aligned}
\mathbb{E} \int_s^t f(u, X_u^x) \mathrm{d}u &\leqslant \left(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}} |\mathbb{Z}_T|^{r'}\right)^{1/r'} \left(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}} \left[\int_s^t f(u, \widetilde{Z}_u^x) \mathrm{d}u\right]^r\right)^{1/r} \\
&\lesssim \left(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}} \left((t-s)^{r-1} \int_s^t |f(u, \widetilde{Z}_u^x)|^r \mathrm{d}u\right)\right)^{1/r} \\
&\lesssim (t-s)^{1-1/q-d/(2p)} \left(\| |f|^r \|_{\widetilde{\mathbb{L}}^{p/r}_{q/r}}\right)^{1/r} \lesssim (t-s)^{1-1/q-d/(2p)} \|f\|_{\widetilde{\mathbb{L}}^p_q}.
\end{aligned}$$

The term $\mathbb{E}\int_s^t f(u, X^x_{\pi_h(u)})\mathrm{d}u$ can be estimated the same way.

(iii): For (3.78), it again follows from Hölder's inequality that

$$\mathbb{E}\exp\left(\int_0^T f(t,X_t^x)\mathrm{d}t\right) = \widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\left[\mathbb{Z}_T\exp\Big(\int_0^T f(t,\widetilde{Z}_t^x)\mathrm{d}t\Big)\right]$$
$$\leqslant \big(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}|\mathbb{Z}_T|^{r'}\big)^{1/r'}\left(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\exp\Big(r\int_0^T f(t,\widetilde{Z}_t^x)\mathrm{d}t\Big)\right)^{1/r} < \infty$$

by (3.80) and (3.63).

(iv): Let

$$A^f(h,X) := \sup_{t\in[0,T]}\Big|\int_0^t (f(s,X_s^x) - f(s,X^x_{\pi_h(s)}))\mathrm{d}u\Big|^2$$

and

$$A^f(h,\widetilde{Z}) := \sup_{t\in[0,T]}\Big|\int_0^t (f(s,\widetilde{Z}_s^x) - f(s,\widetilde{Z}^x_{\pi_h(s)}))\mathrm{d}u\Big|^2.$$

For any $\delta \in (0,1)$, we note that

$$\begin{aligned}\mathbb{E}A^f(h,X) &= \widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\big(\mathbb{Z}_T A^f(h,\widetilde{Z})\big)\\ &= \widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\big(\mathbb{Z}_T|A^f(h,\widetilde{Z})|^\delta|A^f(h,\widetilde{Z})|^{1-\delta}\big).\end{aligned}$$

Based on Hölder's inequality, (3.80), (3.77) and (3.72), for $1/r' + 1/r = 1$ with some $r \in (1, p/2)$, we have

$$\begin{aligned}\mathbb{E}A^f(h,X) &\leqslant \left[\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\big(|\mathbb{Z}_T|^{1/\delta}A^f(h,\widetilde{Z})\big)\right]^\delta\left[\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}A^f(h,\widetilde{Z})\right]^{1-\delta}\\ &\leqslant \big(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}|\mathbb{Z}_T|^{r'/\delta}\big)^{\delta/r'}\big(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}|A^f(h,\widetilde{Z})|^r\big)^{\delta/r}\big(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}A^f(h,\widetilde{Z})\big)^{1-\delta}\\ &\lesssim \||f|^{2r}\|^{\delta/r}_{\widetilde{\mathbb{L}}^{p/(2r)}(T)} h^{(1-\delta_0)}\|f\|^{2(1-\delta)}_{\widetilde{\mathbb{L}}^p(T)}\\ &\lesssim h^{1-\delta_0}\|f\|^2_{\widetilde{\mathbb{L}}^p(T)},\end{aligned}$$

where $\delta_0 = \delta_0(\delta) \to 0$ as $\delta \to 0$, and complete the proof.

□

Next, we want to prove an estimate for $\sup_x \|\mu_s^x - \mu_t^x\|_{var}$. To this end, we will use the relation between the PDE and the SDE. For any $T > 0$, consider the following backward PDE:

$$\partial_t u^T + B\cdot\nabla u^T + a_{ij}\partial_i\partial_j u^T = 0, \quad u^T(T) = \varphi, \tag{3.82}$$

where $\varphi \in C_b^\infty$. By Proposition 2.10, there exists a unique solution u^T to (3.82) in the sense of Definition 2.6. Set

$$P^X_{s,t} f(x) := \mathbb{E} f(X^x_{s,t}), \quad P^X_t := P^X_{0,t}.$$

By (3.76), the domain of $P^X_{s,t}$ includes $\widetilde{L}^p$ for any $p \in (1, \infty]$. Then we have the following probabilistic representation.

Proposition 3.27. *Let $T > 0$, $\varphi \in C_b^\infty(\mathbb{R}^d)$, u^T and $X^x_{s,t}$ be the solutions to (3.82) and (3.75) respectively. Then,*

$$u^T(s, x) = \mathbb{E}\varphi(X^x_{s,T}) = P^X_{s,T}\varphi(x). \tag{3.83}$$

Proof. It is straightforward to obtain (3.83) by applying the generalized Itô formula (3.52) to the function $t \mapsto u^T(t, X^x_{s,t})$ and taking expectation. □

Apart from the probabilistic representation, by the generalized Itô formula, we have the following Duhamel formula.

Lemma 3.28 (Duhamel formula)**.** *For any $\varphi \in C_b^\infty(\mathbb{R}^d)$,*

$$P^X_{s,t}\varphi(x) = P^\sigma_{t-s}\varphi(x) + \int_s^t P^X_{s,r}\big(B(r) \cdot \nabla P^\sigma_{t-r}\varphi\big)(x)\mathrm{d}r. \tag{3.84}$$

Proof. For any $t \in [0, T]$, let $v^t = v^t(t, x)$ be the solution to the following backward PDE:

$$\partial_r v^t + a_{ij}\partial_i\partial_j v^t = 0, \quad v^t(t) = \varphi.$$

Based on (3.83), one sees that $v^t(r) = P^\sigma_{t-r}\varphi$. By the generalized Itô formula (3.52), we have

$$\mathbb{E}v^t(t, X^x_{s,t}) = v^t(s, x) + \mathbb{E}\int_s^t (\partial_r v^t + a_{ij}\partial_i\partial_j v^t + B \cdot \nabla v^t)(r, X^x_{s,r})\mathrm{d}r,$$

which implies that

$$\begin{aligned} P^X_{s,t}\varphi(x) &= P^\sigma_{t-s}\varphi + \mathbb{E}\int_s^t (B(r) \cdot \nabla P^\sigma_{t-r}\varphi)(X^x_{s,r})\mathrm{d}r \\ &= P^\sigma_{t-s}\varphi + \int_s^t P^X_{s,r}(B(r) \cdot \nabla P^\sigma_{t-r}\varphi)(x)\mathrm{d}r, \end{aligned}$$

and we complete the proof. □

Now we can prove the estimate for $\|\mu^x_t - \mu^x_s\|_{var}$. We note that similar results have been proved in [117, Lemma 3.8 (1)(ii)] based on heat kernel estimates. It should be mentioned that the order of time regularity in [117] depends on the Hölder index β of σ and thus this is not applicable to our case.

Lemma 3.29. *Assume (3.49). For any $T > 0$ and $q \in [p_0, \infty]$, there is a constant $C = C(\Xi, q)$ such that for any $0 < s \leqslant t \leqslant T$ and $\varphi \in C_b^\infty$,*

$$\|P_t^X \varphi - P_s^X \varphi\|_\infty \leqslant C\Big[[(t-s)^{\frac{1}{2}-\frac{d}{2q}} s^{-\frac{1}{2}+\frac{d}{2q}}] \wedge 1\Big] s^{-\frac{d}{2q}} \|\varphi\|_q. \tag{3.85}$$

In particular, when $q = \infty$,

$$\|\mu_t^x - \mu_s^x\|_{var} \leqslant C\Big[[(t-s)^{1/2} s^{-1/2}] \wedge 1\Big]. \tag{3.86}$$

Proof. For simplicity, let

$$\alpha := \frac{1}{2} - \frac{d}{2q}.$$

From (3.84), one sees that

$$\begin{aligned} P_t^X\varphi - P_s^X\varphi =& \big(P_t^\sigma \varphi - P_s^\sigma \varphi\big) + \int_s^t P_r^X \big(B(r) \cdot \nabla P_{t-r}^\sigma \varphi\big) \mathrm{d}r \\ &+ \int_0^s P_r^X \Big[B(r) \cdot \nabla \big(P_{t-r}^\sigma - P_{s-r}^\sigma\big)\varphi\Big] \mathrm{d}r \\ :=& \mathscr{I}_1 + \mathscr{I}_2 + \mathscr{I}_3. \end{aligned}$$

Based on (3.66), we have

$$\begin{aligned} \|\mathscr{I}_1\|_\infty &\lesssim \Big[[(t-s)s^{-1}] \wedge 1\Big] s^{-\frac{d}{2q}} \|\varphi\|_q \\ &\lesssim \Big[(t-s)^\alpha s^{-\alpha}\Big] s^{-\frac{d}{2q}} \|\varphi\|_q. \end{aligned}$$

By (3.76) and (3.61), we have

$$\begin{aligned} \|\mathscr{I}_2\|_\infty &\lesssim \int_s^t r^{-d/(2p_0)} \|B(r) \cdot \nabla P_{t-r}^\sigma \varphi\|_{p_0} \mathrm{d}r \\ &\lesssim \int_s^t r^{-d/(2p_0)} \|\nabla P_{t-r}^\sigma \varphi\|_\infty \mathrm{d}r \\ &\lesssim s^{-d/(2q)} \int_s^t r^{-d/(2p_0)+d/(2q)} (t-r)^{-1/2-d/(2q)} \mathrm{d}r \|\varphi\|_q \\ &:= s^{-d/(2q)} K(t,s) \|\varphi\|_q, \end{aligned}$$

where

$$\begin{aligned} K(t,s) &\leqslant \left(s^{-\frac{d}{2p_0}+\frac{d}{2q}} \int_s^t (t-r)^{\alpha-1} \mathrm{d}r\right) \wedge \left((t-s)^{\frac{1}{2}-\frac{d}{2p_0}} \int_0^1 r^{-\frac{d}{2p_0}+\frac{d}{2q}} (1-r)^{-\frac{1}{2}-\frac{d}{2q}} \mathrm{d}r\right) \\ &\lesssim (s^{-\alpha}(t-s)^\alpha) \wedge 1 \end{aligned}$$

since $q \geqslant p_0 > d$.

It remains to estimate $\mathscr{I}_3$. By (3.76), we have

$$\begin{aligned}\|\mathscr{I}_3\|_\infty &\lesssim \int_0^s r^{-d/(2p_0)} \|B(r)\cdot\nabla\big(P^\sigma_{t-r}-P^\sigma_{s-r}\big)\varphi\|_{p_0}\mathrm{d}r \\ &\lesssim \int_0^s r^{-d/(2p_0)}\|\nabla\big(P^\sigma_{t-r}-P^\sigma_{s-r}\big)\varphi\|_\infty \mathrm{d}r. \end{aligned} \tag{3.87}$$

When $q<\infty$, it follows (3.66) that

$$\|\mathscr{I}_3\|_\infty \lesssim \int_0^s r^{-d/(2p_0)}\Big[[(t-s)^{\frac12}(s-r)^{-\frac12}]\wedge 1\Big](s-r)^{-1/2-d/(2q)}\mathrm{d}r\|\varphi\|_q. \tag{3.88}$$

We note

$$\begin{aligned}\mathscr{J} &:= \int_0^s r^{-d/(2p_0)}\Big[[(t-s)^{\frac12}(s-r)^{-\frac12}]\wedge 1\Big](s-r)^{-1/2-d/(2q)}\mathrm{d}r \\ &\leqslant \int_0^s r^{-d/(2p_0)}(s-r)^{-1/2-d/(2q)}\mathrm{d}r \lesssim s^{1/2-d/(2p_0)-d/(2q)} \lesssim s^{-d/(2q)}, \end{aligned} \tag{3.89}$$

since $p_0>d$. In addition, when $r\in(0,\frac{s}{2}]$, one sees that

$$\Big[(t-s)^{\frac12}(s-r)^{-\frac12}\Big]\wedge 1 \leqslant (t-s)^{\frac12-\frac{d}{2q}}(s-r)^{-\frac12+\frac{d}{2q}} \leqslant (t-s)^{\frac12-\frac{d}{2q}}\Big(\frac{s}{2}\Big)^{-\frac12+\frac{d}{2q}}.$$

Hence,

$$\begin{aligned}\mathscr{J} \lesssim& s^{-1/2-d/(2q)}\int_0^{s/2} r^{-d/(2p_0)}\Big[[(t-s)^{\frac12}(s-r)^{-\frac12}]\wedge 1\Big]\mathrm{d}r \\ &+ s^{-d/(2p_0)}\int_{s/2}^s \Big[[(t-s)^{\frac12}(s-r)^{-\frac12}]\wedge 1\Big](s-r)^{-1/2-d/(2q)}\mathrm{d}r \\ \lesssim& s^{-1}(t-s)^{1/2-d/(2q)}\int_0^{s/2} r^{-d/(2p_0)}\mathrm{d}r \\ &+ s^{-d/(2p_0)}\int_{s/2}^s \Big[[(t-s)^{\frac12}(s-r)^{-\frac12}]\wedge 1\Big](s-r)^{-1/2-d/(2q)}\mathrm{d}r. \end{aligned}$$

By two changes of variables, we have

$$\begin{aligned}\mathscr{J} \lesssim& s^{-d/(2p_0)}(t-s)^{1/2-d/(2q)} + s^{-d/(2p_0)}\int_0^{s/2}\Big[(t-s)^{\frac12}r^{-\frac12}\wedge 1\Big]r^{-1/2-d/(2q)}\mathrm{d}r \\ \lesssim& s^{-d/(2p_0)}(t-s)^{1/2-d/(2q)} \\ &+ s^{-d/(2p_0)}(t-s)^{1-1/2-d/(2q)}\int_0^\infty \Big[r^{-\frac12}\wedge 1\Big]r^{-1/2-d/(2q)}\mathrm{d}r \\ \lesssim& s^{-d/(2p_0)}(t-s)^{1-1/2-d/(2q)} \lesssim s^{-\frac12}(t-s)^{\frac12-\frac{d}{2q}}, \end{aligned} \tag{3.90}$$

since $q<\infty$ and $p_0>d$, which combined with (3.89) implies that

$$\mathscr{J}\lesssim\left(s^{-\alpha}(t-s)^{\alpha}\right)\wedge s^{-\frac{d}{2q}}=\Big([(t-s)^{\alpha}s^{-\alpha}]\wedge 1\Big)s^{-\frac{d}{2q}}.$$

When $q=\infty$, by (3.87) and (3.67), we have

$$\|\mathscr{I}_3\|_\infty\lesssim\int_0^s r^{-d/(2p_0)}\Big[[(t-s)^{\frac{1+\beta}{2}}(s-r)^{-\frac{1+\beta}{2}}]\wedge 1\Big](s-r)^{-\frac{1}{2}}\mathrm{d}r\|\varphi\|_\infty.$$

By the same calculation as (3.89) and (3.90), we have

$$\mathscr{J}\lesssim\left(s^{-\frac{1}{2}}(t-s)^{\frac{1}{2}}\int_0^\infty\Big[r^{-\frac{1+\beta}{2}}\wedge 1\Big]r^{-1/2-d/(2q)}\mathrm{d}r\right)\wedge 1\lesssim\left(s^{-\frac{1}{2}}(t-s)^{\frac{1}{2}}\right)\wedge 1.$$

Thus, we obtain (3.85). In particular, by Lusin's theorem and (3.85), we have

$$\|\mu_t^x-\mu_s^x\|_{var}=\sup_{\varphi\in C_b^\infty}\frac{|P_t^X\varphi(x)-P_s^X\varphi(x)|}{\|\varphi\|_\infty}\lesssim(t-s)^{\frac{1}{2}}s^{-\frac{1}{2}}.$$

Moreover, it is easy to see that

$$\|\mu_t^x-\mu_s^x\|_{var}\leqslant 2,$$

which completes the proof. □

The following lemma is the distribution dependent version of (3.79).

Lemma 3.30. *For any $T>0$, $p\in(d\vee 2,\infty)$, assume that $f:\mathbb{R}_+\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)\to\mathbb{R}$ such that*

$$\kappa_f:=\sup_{t\in[0,T]}\sup_{\mu,\nu}\left(\|f(t,\cdot,\mu)\|_p+\frac{\|f(t,\cdot,\mu)-f(t,\cdot,\nu)\|_p}{\|\mu-\nu\|_{var}}\right)<\infty.$$

Then, for any $\delta>0$, there is a constant $C=C(\Xi,p,\delta)$ such that for any $x\in\mathbb{R}^d$ and $h\in(0,1)$

$$\mathbb{E}\Big(\sup_{t\in[0,T]}\Big|\int_0^t(f(s,X_s^x,\mu_s^x)-f(s,X_{\pi_h(s)}^x,\mu_{\pi_h(s)}^x))\mathrm{d}s\Big|^2\Big)\leqslant C(\kappa_f)^2h^{1-\delta}.\tag{3.91}$$

Proof. For simplicity, we drop the superscript x from X^x and μ^x. First of all, we note that

$$\begin{aligned}&\mathbb{E}\Big(\sup_{t\in[0,T]}\Big|\int_0^t(f(s,X_s,\mu_s)-f(s,X_{\pi_h(s)},\mu_{\pi_h(s)}))\mathrm{d}s\Big|^2\Big)\\ \lesssim\,&\mathbb{E}\Big(\sup_{t\in[0,T]}\Big|\int_0^t(f(s,X_s,\mu_s)-f(s,X_{\pi_h(s)},\mu_s))\mathrm{d}s\Big|^2\Big)\\ &+\mathbb{E}\Big(\sup_{t\in[0,T]}\Big|\int_0^t(f(s,X_{\pi_h(s)},\mu_s)-f(s,X_{\pi_h(s)},\mu_{\pi_h(s)}))\mathrm{d}s\Big|^2\Big)\\ :=&\mathscr{I}_1^h+\mathscr{I}_2^h.\end{aligned}$$

By (3.79), for any $\delta > 0$, we have

$$\mathscr{I}_1^h \lesssim h^{1-\delta} \sup_{t\in[0,T]} \|f(t,\cdot,\mu_t)\|_p^2 \lesssim h^{1-\delta}(\kappa_f)^2.$$

For $\mathscr{I}_2^h$, we use the same method as in Step 2 of the proof of Lemma 3.24. For any $0 \leqslant s < t \leqslant T$, set

$$\mathscr{J}_{s,t}^h := \int_{s+h}^{t+h} |f(r, X_{\pi_h(r)}, \mu_r) - f(r, X_{\pi_h(r)}, \mu_{\pi_h(r)})|\mathrm{d}r$$

and $\|\cdot\| := \big(\mathbb{E}|\cdot|^2\big)^{1/2}$. Then

$$\mathscr{I}_2^h \leqslant \mathbb{E}\Big|\int_h^{T+h} |f(s, X_{\pi_h(s)}, \mu_s) - f(s, X_{\pi_h(s)}, \mu_{\pi_h(s)})|\mathrm{d}s\Big|^2 = \|\mathscr{J}_{0,T}^h\|^2,$$

since $\pi_h(r) = r$, if $r < h$. Based on Hölder's inequality and (3.77), one sees that if $2/q + d/p < 1$, then

$$\begin{aligned}
\mathbb{E}|\mathscr{J}_{s,t}^h|^2 &\lesssim (t-s)\mathbb{E}\int_{s+h}^{t+h} \Big|f(r, X_{\pi_h(r)}, \mu_r) - f(r, X_{\pi_h(r)}, \mu_{\pi_h(r)})\Big|^2 \mathrm{d}r \\
&\lesssim (t-s)\Big(\int_{s+h}^{t+h} \|f(r,\cdot,\mu_r) - f(r,\cdot,\mu_{\pi_h(r)})\|_p^q \mathrm{d}r\Big)^{2/q} \\
&\lesssim (\kappa_f)^2(t-s)\Big(\int_{s+h}^{t+h} \|\mu_r - \mu_{\pi_h(r)}\|_{var}^q \mathrm{d}r\Big)^{2/q}.
\end{aligned}$$

Then, by (3.86) and the fact that $q > 2$, for any $\delta > 0$ we have

$$\begin{aligned}
\|\mathscr{J}_{s,t}^h\| &\lesssim \kappa_f(t-s)^{1/2}\Big(\int_{s+h}^{t+h} h^{(\frac{1}{2}-\delta)q}(\pi_h(r))^{-q/2}\mathrm{d}r\Big)^{1/q} \\
&\lesssim \kappa_f(t-s)^{1/2}h^{1/2-\delta}\Big(\int_s^t r^{-q/2}\mathrm{d}r\Big)^{1/q} \lesssim \kappa_f h^{1/2-\delta}(t-s)^{1/2+1/q}s^{-1/2}.
\end{aligned}$$

Taking $t_n := 2^{-n}T$, we have

$$\begin{aligned}
\|\mathscr{J}_{0,T}^h\| \leqslant \sum_{n=0}^{\infty} \|\mathscr{J}_{t_{n+1},t_n}^h\| &\lesssim \kappa_f h^{1/2-\delta} \sum_{n=0}^{\infty}(t_n - t_{n+1})^{1/2+1/q}(t_{n+1})^{-1/2} \\
&\lesssim \kappa_f h^{1/2-\delta}T^{1/q}\sum_{n=0}^{\infty} 2^{-\frac{n+1}{q}} \lesssim \kappa_f h^{1/2-\delta},
\end{aligned}$$

which implies $\mathscr{I}_2^h \lesssim (\kappa_f)^2 h^{1-2\delta}$ and we complete the proof.

□

3.2.3 Time regularity for solutions to parabolic PDEs

In this section, we establish the time regularity for the solution of PDE (3.51). First, we give the following probabilistic representation of the solution to PDE if $B \equiv 0$.

Lemma 3.31. *Let $B \equiv 0$ and u be a solution to PDE (3.51). Then,*

$$u(t) = \int_0^t e^{-\lambda(t-s)} P^\sigma_{t-s} f(s)\mathrm{d}s + e^{-\lambda t} P^\sigma_t \varphi. \tag{3.92}$$

Proof. Applying the generalized Itô formula (3.52) to

$$s \to e^{-\lambda s} u(t-s, Z^x_s),$$

we get

$$\begin{aligned} e^{-\lambda t} u(0, Z^x_t) - u(t,x) = &\int_0^t e^{-\lambda s}(-\partial_s u + a_{ij}\partial_i\partial_j u - \lambda u)(t-s, Z^x_s)\mathrm{d}s \\ &+ \int_0^t e^{-\lambda s} \nabla u(t-s, Z^x_s)\mathrm{d}W_s. \end{aligned}$$

Taking expectation of both sides, we obtain that

$$e^{-\lambda t} P^\sigma_t \varphi - u(t) = -\int_0^t e^{-\lambda s} P^\sigma_s f(t-s)\mathrm{d}s,$$

which is (3.92) by a change of variable and this completes the proof. □

Using the above lemma, we have the following time Hölder regularity of ∇u.

Lemma 3.32. *Assume $\varphi \equiv 0$. Under condition (3.49) with some $p_0 \in (d, \infty)$, for any $\lambda \geqslant 0$, there is a constant $C = C(\Xi, p, \lambda)$ such that for all $t, s \in [0, T]$ and $f \in \widetilde{\mathbb{L}}^{p_0}(T)$ the solution u to (3.51) in the sense of Definition 2.6 satisfies*

$$\|\nabla u(t) - \nabla u(s)\|_\infty \leqslant C|t-s|^{\frac{1}{2} - \frac{d}{2p_0}} |\!|\!| f |\!|\!|_{\widetilde{\mathbb{L}}^{p_0}(T)}. \tag{3.93}$$

Remark 3.33. By Remark 2.9 and (3.93), we further have that there is a version of the solution such that $u \in C([0,T]; \mathbf{C}^1)$.

Proof. First, since $B, f \in \widetilde{\mathbb{L}}^{p_0}(T) \subset \widetilde{\mathbb{L}}^{p_0}_q(T)$, $\forall q$, we indeed have a unique solution u. Set

$$g(s) := B \cdot \nabla u(s) + f(s).$$

In view of (2.20),

$$|\!|\!| g |\!|\!|_{\widetilde{\mathbb{L}}^{p_0}(T)} \leqslant |\!|\!| b |\!|\!|_{\widetilde{\mathbb{L}}^{p_0}(T)} \|\nabla u\|_{\mathbb{L}^\infty_T} + |\!|\!| f |\!|\!|_{\widetilde{\mathbb{L}}^{p_0}(T)} \lesssim |\!|\!| f |\!|\!|_{\widetilde{\mathbb{L}}^{p_0}(T)}.$$

Then, for any $0 \leqslant s < t \leqslant T$, by (3.92), one sees that

$$\begin{aligned}\|\nabla u(t)-\nabla u(s)\|_\infty = &\left\|\int_s^t e^{-\lambda(t-r)}\nabla P^\sigma_{t-r}g(r)\mathrm{d}r\right\|_\infty\\ &+\left\|\int_0^s \left(e^{-\lambda(t-r)}-e^{-\lambda(s-r)}\right)\nabla P^\sigma_{t-r}g(r)\mathrm{d}r\right\|_\infty\\ &+\left\|\int_0^s e^{-\lambda(s-r)}\left(\nabla P^\sigma_{t-r}-\nabla P^\sigma_{s-r}\right)g(r)\mathrm{d}r\right\|_\infty\\ =:& \mathscr{I}_1+\mathscr{I}_2+\mathscr{I}_3.\end{aligned}$$

By (3.61), one sees that

$$\begin{aligned}\mathscr{I}_1 &\lesssim \int_s^t (t-r)^{-1/2-d/(2p_0)}\|g(r)\|_{p_0}\mathrm{d}r\\ &\lesssim \int_s^t (t-r)^{-1/2-d/(2p_0)}\mathrm{d}r\|g\|_{\widetilde{\mathbb{L}}^{p_0}(T)}\\ &\lesssim |t-s|^{\frac{1}{2}-\frac{d}{2p_0}}\|f\|_{\widetilde{\mathbb{L}}^{p_0}(T)}.\end{aligned}$$

For $\mathscr{I}_2$, noting that $|e^{-x}-e^{-y}| \leqslant |x-y|$ for any $x, y > 0$, it follows from (3.61) that

$$\begin{aligned}\mathscr{I}_2 &\lesssim |t-s|\int_0^s (t-r)^{-1/2-d/(2p_0)}\|g(r)\|_{p_0}\mathrm{d}r\\ &\lesssim |t-s|\|g\|_{\widetilde{\mathbb{L}}^{p_0}(T)} \lesssim |t-s|\|f\|_{\widetilde{\mathbb{L}}^{p_0}(T)}.\end{aligned}$$

For $\mathscr{I}_3$, by (3.66) with $q=\infty$, we have

$$\mathscr{I}_3 \lesssim \int_0^s ([(t-s)^{\frac{1}{2}}(s-r)^{-\frac{1}{2}}]\wedge 1)(s-r)^{-1/2-d/(2p_0)}\mathrm{d}r\|g\|_{\widetilde{\mathbb{L}}^{p_0}(T)}.$$

By a change of variable, we have

$$\begin{aligned}\mathscr{I}_3 &\lesssim |t-s|^{\frac{1}{2}-\frac{d}{2p_0}}\int_0^\infty ([(s-r)^{-\frac{1}{2}}]\wedge 1)(s-r)^{-1/2-d/(2p_0)}\mathrm{d}r\|f\|_{\widetilde{\mathbb{L}}^{p_0}(T)}\\ &\lesssim |t-s|^{1/2-d/(2p_0)}\|f\|_{\widetilde{\mathbb{L}}^{p_0}(T)}\end{aligned}$$

and complete the proof.

□

Moreover, we also have an estimate of time regularity for the solution to the following Cauchy problem.

Lemma 3.34. *Assume (3.49). Let* $\varphi \in C_b^\infty$ *and let* u *be the unique solution to the following Cauchy problem on* $[0,T]$ *in the sense of Definition 2.6*

$$\partial_t u = a_{ij}\partial_i\partial_j u + B\cdot\nabla u, \quad u_0 = \varphi. \tag{3.94}$$

Then there is a constant $C = C(\Xi)$ *such that for all* $0 \leqslant s < t \leqslant T$*,*

$$\|\nabla u(t) - \nabla u(s)\|_\infty \leqslant C(t-s)^{\frac{1}{2}-\frac{d}{2p_0}} s^{-1+\frac{d}{2p_0}}\|\varphi\|_\infty. \tag{3.95}$$

Proof. First, by (3.92), we have

$$u(t) = \int_0^t P^\sigma_{t-s}(B\cdot\nabla u)(s)\mathrm{d}s + P^\sigma_t\varphi, \tag{3.96}$$

which implies that

$$\begin{aligned}\|\nabla u(t)\|_\infty &\lesssim \int_0^t (t-s)^{-\frac{1}{2}-\frac{d}{2p_0}}\|B\cdot\nabla u(s)\|_{p_0}\mathrm{d}s + t^{-\frac{1}{2}}\|\varphi\|_\infty \\ &\lesssim \int_0^t (t-s)^{-\frac{1}{2}-\frac{d}{2p_0}}\|\nabla u(s)\|_\infty\mathrm{d}s + t^{-\frac{1}{2}}\|\varphi\|_\infty\end{aligned}$$

because of (3.61). Hence, by Gronwall's inequality, we have

$$\|\nabla u(t)\|_\infty \lesssim t^{-\frac{1}{2}}\|\varphi\|_\infty. \tag{3.97}$$

By (3.96), (3.61) and (3.66), one sees that

$$\begin{aligned}\|\nabla u(t) - \nabla u(s)\|_\infty \leqslant& \int_s^t \|\nabla P^\sigma_{t-r}(B\cdot\nabla u)(r)\|_\infty\mathrm{d}r + \int_0^s \|\nabla(P^\sigma_{t-r} - P^\sigma_{s-r})B\cdot\nabla u(r)\|_\infty\mathrm{d}r \\ &+ \|\nabla(P^\sigma_t - P^\sigma_s)\varphi\|_\infty. \\ \lesssim& \int_s^t (t-r)^{-\frac{1}{2}-\frac{d}{2p_0}}\|\nabla u(r)\|_\infty\mathrm{d}r \\ &+ \int_0^s \left\{\left[(t-s)^{\frac{1}{2}}(s-r)^{-\frac{1}{2}}\right]\wedge 1\right\}(s-r)^{-\frac{1}{2}-\frac{d}{2p_0}}\|\nabla u(r)\|_\infty\mathrm{d}r \\ &+ \left\{\left[(t-s)^{\frac{1}{2}}s^{-\frac{1}{2}}\right]\wedge 1\right\}s^{-\frac{1}{2}}\|\varphi\|_\infty \\ :=& \mathscr{I}_1 + \mathscr{I}_2 + \mathscr{I}_3.\end{aligned}$$

Then, noting that $x\wedge 1 \leqslant x^\theta$ for all $x \geqslant 0$ and $\theta\in[0,1]$, we have

$$\mathscr{I}_3 \lesssim \left\{\left[(t-s)^{\frac{1}{2}}s^{-\frac{1}{2}}\right]\wedge 1\right\}s^{-\frac{1}{2}}\|\varphi\|_\infty \lesssim (t-s)^{\frac{1}{2}-\frac{d}{2p_0}}s^{-1+\frac{d}{2p_0}}\|\varphi\|_\infty.$$

Based on (3.97) and a change of variable,

$$\begin{aligned}\mathscr{I}_1 &\lesssim \|\varphi\|_\infty \int_s^t (t-r)^{-\frac{1}{2}-\frac{d}{2p_0}} r^{-\frac{1}{2}} \mathrm{d}r \lesssim (t-s)^{\frac{1}{2}-\frac{d}{2p_0}} s^{-\frac{1}{2}} \|\varphi\|_\infty \\ &\lesssim (t-s)^{\frac{1}{2}-\frac{d}{2p_0}} s^{-1+\frac{d}{2p_0}} \|\varphi\|_\infty\end{aligned}$$

since $p_0 > d$. For $\mathscr{I}_2$, again by (3.97), we divide $[0,s]$ into $[0,s/2)$ and $[s/2,s]$ and have

$$\begin{aligned}\frac{\mathscr{I}_2}{\|\varphi\|_\infty} \lesssim &\Big(\int_0^{s/2} + \int_{s/2}^s\Big)\Big\{\Big[(t-s)^{\frac{1}{2}}(s-r)^{-\frac{1}{2}}\Big] \wedge 1\Big\}(s-r)^{-\frac{1}{2}-\frac{d}{2p_0}} r^{-\frac{1}{2}} \mathrm{d}r \\ \lesssim &(t-s)^{\frac{1}{2}-\frac{d}{2p_0}} s^{-1} \int_0^{s/2} r^{-\frac{1}{2}} \mathrm{d}r \\ &+ s^{-\frac{1}{2}} \int_{s/2}^s \Big\{\Big[(t-s)^{\frac{1}{2}}(s-r)^{-\frac{1}{2}}\Big] \wedge 1\Big\}(s-r)^{-\frac{1}{2}-\frac{d}{2p_0}} \mathrm{d}r \\ \lesssim &(t-s)^{\frac{1}{2}-\frac{d}{2p_0}} s^{-\frac{1}{2}} + s^{-\frac{1}{2}} \int_0^s \Big\{\Big[(t-s)^{\frac{1}{2}} r^{-\frac{1}{2}}\Big] \wedge 1\Big\} r^{-\frac{1}{2}-\frac{d}{2p_0}} \mathrm{d}r,\end{aligned}$$

where we used the fact that

$$\begin{aligned}&\Big[\Big((t-s)^{\frac{1}{2}}(s-r)^{-\frac{1}{2}}\Big) \wedge 1\Big] (s-r)^{-\frac{1}{2}-\frac{d}{2p_0}} \\ &\leqslant (t-s)^{\frac{1}{2}-\frac{d}{2p_0}} (s-r)^{-1} \leqslant 2(t-s)^{\frac{1}{2}-\frac{d}{2p_0}} s^{-1} \quad \forall r \in [0, s/2).\end{aligned}$$

From a change of variable, we have

$$\begin{aligned}\mathscr{I}_2 &\lesssim \Big[(t-s)^{\frac{1}{2}-\frac{d}{2p_0}} s^{-\frac{1}{2}} + s^{-\frac{1}{2}} (t-s)^{1-\frac{1}{2}-\frac{d}{2p_0}} \int_0^\infty (r^{-\frac{1}{2}} \wedge 1) r^{-\frac{1}{2}-\frac{d}{2p_0}} \mathrm{d}r\Big] \|\varphi\|_\infty \\ &\lesssim (t-s)^{\frac{1}{2}-\frac{d}{2p_0}} s^{-\frac{1}{2}} \|\varphi\|_\infty \lesssim (t-s)^{\frac{1}{2}-\frac{d}{2p_0}} s^{-1+\frac{d}{2p_0}} \|\varphi\|_\infty\end{aligned}$$

since $p_0 \in (d, \infty)$ and complete the proof.

□

Chapter 4

Well-posedness of McKean-Vlasov SDEs with mixed L^p coefficients

In this chapter, we consider the following dDDSDE:

$$\mathrm{d}X_t = b(t, X_t, \rho_t(X_t), \mu_{X_t})\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \tag{4.1}$$

where $\rho_t(x)$ is the density of X_t and $b(t, x, r, \mu) : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ is a measurable function.

Here is the main result in this chapter

Theorem 4.1. (Weak well-posedness) *Suppose that* $(\mathbf{A}^\sigma)$ *holds and for any* $T > 0$ *and* $i = 1, \cdots, d$*, there are indices* $(q_i, \boldsymbol{p}_i) \in \mathscr{I}^o$ *and* $\boldsymbol{\pi}_i \in S_d$ *such that*

$$\sup_{\mu \in C([0,T];\mathcal{P}(\mathbb{R}^d))} \||\sup_{r \geqslant 0} |b^i(\cdot, \cdot, r, \mu_\cdot)|\||_{\widetilde{\mathbb{L}}^{q_i}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i})} \leqslant \kappa_1, \tag{4.2}$$

and for some $h_i \in \mathbb{L}^{q_i}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i})$ *and for all* $t, x \in [0, T] \times \mathbb{R}^d$*,* $r, r' \geqslant 0$ *and* $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$*,*

$$|b^i(t, x, r, \mu) - b^i(t, x, r', \nu)| \leqslant h_i(t, x)\big(|r - r'| + \|\mu - \nu\|_{\mathrm{var}}\big). \tag{4.3}$$

Then for any probability measure $\mu_0(\mathrm{d}x) = \rho_0(x)\mathrm{d}x$ *with* $\rho_0 \in \mathbb{L}^\infty$*, there is a unique weak solution* $(X, W, \mathfrak{U})$*, or equivalently, a martingale solution to dDDSDE (4.1) with initial distribution* μ_0*.*

As a consequence, we have the following strong well-posedness.

Theorem 4.2. (Strong well-posedness) *In addition to the assumptions in Theorem 4.1, we also assume* $(\mathbf{H}^\sigma_{\mathbf{mix}})$ *holds. Then there is a unique strong solution.*

Proof. It is directly from Theorems 4.1 and 3.1. □

The proof of Theorem 4.1 is presented in the following. Section 4.1 establishes stability (1.31), which is a crucial prerequisite for the subsequent analysis. The Picard iteration is then utilized in Section 4.2 to demonstrate Theorem 4.1.

4.1 On the stability of densities of solutions to SDEs with respect to drifts

In this section we prepare a stability result about the density of classical SDEs. Our starting point is the associated Fokker-Planck equation. Fix $z \in \mathbb{R}^d$. Let

$$A^z_{s,t} := \int_s^t A(r,z)\mathrm{d}r \quad \text{with} \quad A(r,z) = (a_{ij}(r,z)) = ((\sigma\sigma^*)_{ij}(r,z))/2.$$

Let $P^z_{s,t}f(x)$ be the Gaussian heat kernel associated with $A^z_{s,t}$, i.e.,

$$P^z_{s,t}f(x) = \int_{\mathbb{R}^d} h_{A^z_{s,t}}(x-y)f(y)\mathrm{d}y,$$

where for a symmetric positive definite matrix A,

$$h_A(x) := \frac{\mathrm{e}^{-\langle A^{-1}x,x\rangle/4}}{\sqrt{(4\pi)^d \det(A)}}.$$

Lemma 4.3. *Let $\beta \in [0,1]$, $k \in \mathbb{N}_0$, $\boldsymbol{p} \in [1,\infty]^d$ and $\boldsymbol{\pi} \in S_d$. Under $(\mathbf{A}^\sigma)$, for any $T > 0$, there is a constant $C = C(T,d,\beta,k,\boldsymbol{p},\kappa_0) > 0$ such that for all $0 \leqslant s < t \leqslant T$ and $0 \leqslant f \in \widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}$,*

$$|\nabla^k P^z_{s,t}(|\cdot|^\beta f)(0)| \leqslant C(t-s)^{\frac{1}{2}(\beta-k-|\frac{1}{\boldsymbol{p}}|)}|||f|||_{\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}}.$$

Proof. First of all, by definition and $(\mathbf{A}^\sigma)$, it is easy to see that for some $\lambda > 0$,

$$|\nabla^k h_{A^z_{s,t}}(x)| \lesssim (t-s)^{-\frac{k+d}{2}}\mathrm{e}^{-\frac{|x|^2}{\lambda(t-s)}} = (t-s)^{-\frac{k}{2}}(2\pi\lambda)^{\frac{d}{2}}g_{\lambda(t-s)}(x),$$

and for some $\lambda' > \lambda$,

$$\begin{aligned}|\nabla^k P^z_{s,t}(|\cdot|^\beta f)(0)| &\lesssim (t-s)^{-\frac{k}{2}}\int_{\mathbb{R}^d} g_{\lambda(t-s)}(y)|y|^\beta f(y)\mathrm{d}y \\ &\lesssim (t-s)^{\frac{\beta}{2}-\frac{k}{2}}\int_{\mathbb{R}^d} g_{\lambda'(t-s)}(y)f(y)\mathrm{d}y.\end{aligned}$$

Let $\boldsymbol{p'} \in (1,\infty)^d$ be defined by $\frac{1}{\boldsymbol{p}} + \frac{1}{\boldsymbol{p'}} = \mathbf{1}$. Fix $r > 0$. By Hölder's inequality we have

$$\begin{aligned}\int_{\mathbb{R}^d} g_{\lambda'(t-s)}(y)f(y)\mathrm{d}y &= \frac{1}{|B^r_0|}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} g_{\lambda'(t-s)}(y)\mathbb{1}_{B^r_z}(y)f(y)\mathrm{d}y\mathrm{d}z \\ &\leqslant \frac{1}{|B^r_0|}\int_{\mathbb{R}^d}\|\mathbb{1}_{B^r_z}g_{\lambda'(t-s)}\|_{\mathbb{L}^{\boldsymbol{p'}}_{\boldsymbol{\pi}}}\|\mathbb{1}_{B^r_z}f\|_{\mathbb{L}^{\boldsymbol{p}}_{\boldsymbol{\pi}}}\mathrm{d}z \\ &\leqslant \frac{1}{|B^r_0|}\left(\int_{\mathbb{R}^d}\|\mathbb{1}_{B^r_z}g_{\lambda'(t-s)}\|_{\mathbb{L}^{\boldsymbol{p'}}_{\boldsymbol{\pi}}}\mathrm{d}z\right)|||f|||_{\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}}.\end{aligned} \tag{4.4}$$

Below, without loss of generality, we suppose $s=0$. By a change of variables, we have

$$\begin{aligned}\int_{\mathbb{R}^d}\|\mathbb{1}_{B^r_z}g_{\lambda t}\|_{\mathbb{L}^{\boldsymbol{p}'}_{\boldsymbol{\pi}}}\mathrm{d}z&=(2\pi\lambda t)^{-\frac{d}{2}}\int_{\mathbb{R}^d}\|\mathbb{1}_{B^r_z}\mathrm{e}^{-\frac{|\cdot|^2}{\lambda t}}\|_{\mathbb{L}^{\boldsymbol{p}'}_{\boldsymbol{\pi}}}\mathrm{d}z\\&\lesssim t^{-\frac{d}{2}}\prod_{i=1}^d\int_{\mathbb{R}}\left(\int_{|y_i-z_i|\leqslant r}\mathrm{e}^{-\frac{p_i'|y_i|^2}{\lambda t}}\mathrm{d}y_i\right)^{\frac{1}{p_i'}}\mathrm{d}z_i=:t^{-\frac{d}{2}}\prod_{i=1}^d\mathcal{J}_i.\end{aligned}$$

For each i, we have

$$\begin{aligned}\mathcal{J}_i&=\int_{|z_i|\leqslant 2r}\left(\int_{|y_i-z_i|\leqslant r}\mathrm{e}^{-\frac{p_i'|y_i|^2}{\lambda t}}\mathrm{d}y_i\right)^{\frac{1}{p_i'}}\mathrm{d}z_i+\int_{|z_i|>2r}\left(\int_{|y_i-z_i|\leqslant r}\mathrm{e}^{-\frac{p_i'|y_i|^2}{\lambda t}}\mathrm{d}y_i\right)^{\frac{1}{p_i'}}\mathrm{d}z_i\\&\leqslant\int_{|z_i|\leqslant 2r}\left(\int_{\mathbb{R}}\mathrm{e}^{-\frac{p_i'|y_i|^2}{\lambda t}}\mathrm{d}y_i\right)^{\frac{1}{p_i'}}\mathrm{d}z_i+\int_{|z_i|>2r}\mathrm{e}^{-\frac{p_i'(|z_i|-r)^2}{\lambda t}}\left(\int_{|y_i-z_i|\leqslant r}\mathrm{d}y_i\right)^{\frac{1}{p_i'}}\mathrm{d}z_i\\&\lesssim\left(\int_{\mathbb{R}}\mathrm{e}^{-\frac{p_i'|y_i|^2}{\lambda t}}\mathrm{d}y_i\right)^{\frac{1}{p_i'}}+\int_{\mathbb{R}}\mathrm{e}^{-\frac{p_i'|z_i|^2}{2\lambda t}}\mathrm{d}z_i\lesssim t^{\frac{1}{2p_i'}}+t^{\frac{1}{2}}\lesssim t^{\frac{1}{2p_i'}}=t^{\frac{1}{2}(1-\frac{1}{p_i})}.\end{aligned}$$

Hence,

$$\int_{\mathbb{R}^d}\|\mathbb{1}_{B^r_z}g_{\lambda t}\|_{\mathbb{L}^{\boldsymbol{p}'}_{\boldsymbol{\pi}}}\mathrm{d}z\lesssim t^{-\frac{d}{2}}\prod_{i=1}^d t^{\frac{1}{2}(1-\frac{1}{p_i})}=t^{-|\frac{1}{\boldsymbol{p}}|/2}.$$

Combining the above estimates, we obtain the desired estimate. □

The following stability result shall be used below to show the existence and uniqueness.

Lemma 4.4. *Let b_0,b_1 be two Borel measurable functions satisfying (2.49) and for $k=0,1$, $\mu_k(\mathrm{d}x):=\rho^o_k(x)\mathrm{d}x$ with $\rho^o_k\in\mathbb{L}^\infty$. Let $\mathbb{P}_k\in\mathcal{M}^{\sigma,b_k}_{\mu_k}$ be the unique martingale solution and $\rho_k(t,x)$ be the density of the coordinated process w_t under $\mathbb{P}_k$. Then for any $T>0$, there is a constant $C=C(T,\Theta)>0$ such that for all $t\in[0,T]$,*

$$\|\rho_0(t)-\rho_1(t)\|_{\mathbb{L}^\infty}\lesssim_C\|\rho^o_0-\rho^o_1\|_{\mathbb{L}^\infty}+\sum_{i=1}^d\int_0^t(t-s)^{-\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}_i}|)}\|b^i_0(s)-b^i_1(s)\|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i}}\mathrm{d}s.\quad(4.5)$$

Proof. First of all, by the heat kernel estimate (3.41), we have for all t,y,

$$\rho_k(t,y)\leqslant\frac{C_1}{t^{d/2}}\int_{\mathbb{R}^d}\mathrm{e}^{-\frac{|x-y|^2}{2\delta_1 t}}\rho^o_k(x)\mathrm{d}x\lesssim\|\rho^o_k\|_{\mathbb{L}^\infty},\ k=0,1.\quad(4.6)$$

Note that ρ_k solves the following Fokker-Planck equation in the distributional sense:

$$\partial_t\rho_k=\partial_i\partial_j(a_{ij}\rho_k)+\mathrm{div}(b_k\rho_k),\ k=0,1,$$

where $a = \sigma\sigma^*/2$ and we use the Einstein convention for summation. Below we use the freezing technique to show our result. Fix $z \in \mathbb{R}^d$. For a function f, we set

$$\tau_z f(x) := f(x+z), \quad \ell(t,x) := \rho_0(t,x) - \rho_1(t,x).$$

By the invariance of shifting the spatial variable x, we have

$$\begin{aligned}\partial_t \tau_z \ell &= \partial_i\partial_j(\tau_z a_{ij}\tau_z\ell) + \mathrm{div}(\tau_z b_0 \tau_z \ell) + \mathrm{div}(\tau_z(b_0 - b_1)\tau_z\rho_1)\\ &= a_{ij}(t,z)\partial_i\partial_j\tau_z\ell + \partial_i\partial_j((\tau_z a_{ij} - a_{ij}(t,z))\tau_z\ell)\\ &\quad + \mathrm{div}(\tau_z b_0\tau_z\ell) + \mathrm{div}(\tau_z(b_0-b_1)\tau_z\rho_1).\end{aligned}$$

By Duhamel's formula we have

$$\begin{aligned}\tau_z\ell(t,x) = P^z_{0,t}\tau_z\ell(0,x) &+ \int_0^t P^z_{s,t}(\partial_i\partial_j((\tau_z a_{ij} - a_{ij}(s,z))\tau_z\ell))(s,x)\mathrm{d}s\\ &+ \int_0^t P^z_{s,t}\mathrm{div}(\tau_z b_0\tau_z\ell)(s,x)\mathrm{d}s + \int_0^t P^z_{s,t}\mathrm{div}(\tau_z(b_0-b_1)\tau_z\rho_1)(s,x)\mathrm{d}s.\end{aligned}$$

By $(\mathbf{A}^\sigma)$ and Lemma 4.3 we have

$$\begin{aligned}|\tau_z\ell(t,0)| \lesssim |P^z_{0,t}\tau_z\ell(0,0)| &+ \int_0^t (t-s)^{\frac{\gamma_0}{2}-1}\|\tau_z\ell\|_{\mathbb{L}^\infty}\mathrm{d}s\\ &+ \sum_{i=1}^d \int_0^t (t-s)^{-\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}_i}|)} |\!|\!| \tau_z b_0^i \tau_z\ell |\!|\!|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i}}\mathrm{d}s\\ &+ \sum_{i=1}^d \int_0^t (t-s)^{-\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}_i}|)} |\!|\!| \tau_z (b_0^i - b_1^i)\tau_z\rho_1 |\!|\!|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i}}\mathrm{d}s.\end{aligned}$$

Noting that

$$|\!|\!| \tau_z b_0^i\tau_z\ell |\!|\!|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i}} \leqslant |\!|\!| \tau_z b_0^i |\!|\!|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i}} \|\tau_z\ell\|_{\mathbb{L}^\infty} \leqslant \kappa_1\|\ell\|_{\mathbb{L}^\infty},$$

and by (4.6),

$$\begin{aligned}|\!|\!| \tau_z(b_0^i - b_1^i)\tau_z\rho_1 |\!|\!|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i}} &\leqslant |\!|\!| \tau_z(b_0^i - b_1^i) |\!|\!|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i}} \|\tau_z\rho_1\|_{\mathbb{L}^\infty}\\ &= |\!|\!| b_0^i - b_1^i |\!|\!|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i}} \|\rho_1\|_{\mathbb{L}^\infty} \lesssim |\!|\!| b_0^i - b_1^i |\!|\!|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i}} \|\rho_1^o\|_{\mathbb{L}^\infty},\end{aligned}$$

we further have

$$\begin{aligned}\|\ell(t)\|_{\mathbb{L}^\infty} = \sup_z|\tau_z\ell(t,0)| \lesssim \|\ell(0)\|_{\mathbb{L}^\infty} &+ \int_0^t (t-s)^{\frac{\gamma_0}{2}-1}\|\ell(s)\|_{\mathbb{L}^\infty}\mathrm{d}s\\ &+ \sum_{i=1}^d \int_0^t (t-s)^{-\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}_i}|)}\|\ell(s)\|_{\mathbb{L}^\infty}\mathrm{d}s\\ &+ \sum_{i=1}^d \int_0^t (t-s)^{-\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}_i}|)} |\!|\!| b_0^i - b_1^i |\!|\!|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i}}\mathrm{d}s.\end{aligned}$$

By Lemma A.4, we obtain the desired estimate. □

4.2 Proof of the well-posedness

Now we are ready to prove the main result of this section.

Proof of Theorem 4.1. We divide the proof into three steps.

(Step 1). Let $\mu_t^0 \equiv \mu_0$ for any $t \geqslant 0$. We consider the following Picard iteration: for $n \in \mathbb{N}$,

$$\mathrm{d}X_t^n = b_n(t, X_t^n)\mathrm{d}t + \sigma(t, X_t^n)\mathrm{d}W_t, \quad X_0^n \overset{(d)}{=} \mu_0, \tag{4.7}$$

where

$$b_n(t,x) := b(t, x, \rho_t^{n-1}(x), \mu_t^{n-1}),$$

and

$$\mu_t^{n-1} \text{ is the marginal distribution of } X_t^{n-1}\text{, which has a density } \rho_t^{n-1}. \tag{4.8}$$

By (4.2), one sees that for each $i = 1, \cdots, d$,

$$\sup_n \|b_n^i\|_{\widetilde{\mathbb{L}}_T^{q_i}(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}_i}^{\boldsymbol{p}_i})} \leqslant \kappa_1. \tag{4.9}$$

Thus, by Theorem 3.14, for each $n \in \mathbb{N}$, there is a unique weak solution $(X^n, W^n, \mathfrak{U}^n)$ to SDE (4.7), where

$$\mathfrak{U}^n := (\Omega^n, \mathscr{F}^n, \mathbf{P}_n; (\mathscr{F}_t^n)_{t\geqslant 0}),$$

and for each $t > 0$, X_t^n admits a density ρ_t^n satisfying the following estimate: for all $(t, y) \in [0, T] \times \mathbb{R}^d$,

$$\rho_t^n(y) \leqslant \frac{C_1}{t^{d/2}} \int_{\mathbb{R}^d} \mathrm{e}^{-\frac{|x-y|^2}{\delta_1 t}} \rho_0(x)\mathrm{d}x \lesssim \|\rho_0\|_\infty. \tag{4.10}$$

Moreover, for any $T > 0$, by (3.40), there is a constant $C > 0$ such that

$$\sup_n \mathbb{E}_{\mathbf{P}_n} |X_t^n - X_s^n|^4 \leqslant C|t-s|^2, \quad s, t \in [0, T],$$

and by (3.43), for any $(q_0, \boldsymbol{p}_0) \in \mathscr{I}_2$, there is a constant $C > 0$ such that for all $f \in \widetilde{\mathbb{L}}_T^{q_0}(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}_0})$,

$$\sup_n \mathbb{E}_{\mathbf{P}_n}\left(\int_0^T f(s, X_s^n)\mathrm{d}s\right) \leqslant C\|f\|_{\widetilde{\mathbb{L}}_T^{q_0}(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}_0})}. \tag{4.11}$$

In particular, by Kolmogorov's criterion,

$$\text{the laws } \mathbb{P}_n \text{ of } X_\cdot^n \text{ in } \mathbb{C}_T \text{ are tight.} \tag{4.12}$$

(Step 2). For simplicity of notations, we write

$$\Gamma_{n,m}(t) := \|\rho_t^h - \rho_t^m\|_{\mathbb{L}^\infty} + \|\rho_t^h - \rho_t^m\|_{\mathbb{L}^1}.$$

Noting that by (4.3) and (4.8),

$$\begin{aligned} |b_n^i(s,x) - b_m^i(s,x)| &\leqslant h_i(s,x)\big(|\rho_s^{n-1}(x) - \rho_s^{m-1}(x)| + \|\mu_s^{n-1} - \mu_s^{m-1}\|_{\mathrm{var}}\big) \\ &\leqslant h_i(s,x)\Gamma_{n-1,m-1}(s), \end{aligned}$$

we have

$$\|b_n^i(s) - b_m^i(s)\|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i}} \leqslant \|h_i(s)\|_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i}} \Gamma_{n-1,m-1}(s) =: \ell_i(s)\Gamma_{n-1,m-1}(s). \tag{4.13}$$

Since $(\frac{q_i}{2}, \frac{\boldsymbol{p}_i}{2}) \in \mathscr{I}_2$, by Lemma 3.26 and (4.11), (4.13), we have

$$\begin{aligned} \mathcal{H}(\mu_t^n|\mu_t^m) &\leqslant \frac{1}{2}\mathbb{E}_{\mathbb{P}_m}\left(\int_0^t |\sigma^{-1}(s,w_s)(b_n(s,w_s) - b_m(s,w_s))|^2 \mathrm{d}s\right) \\ &\leqslant \frac{\|\sigma^{-1}\|_\infty^2}{2}\mathbb{E}_{\mathbb{P}_m}\left(\int_0^t |b_n(s,w_s) - b_m(s,w_s)|^2 \mathrm{d}s\right) \\ &\lesssim \sum_{i=1}^d \left(\int_0^t \||b_n^i(s) - b_m^i(s)|^2\|^{q_i/2}_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i/2}_{\boldsymbol{\pi}_i}} \mathrm{d}s\right)^{\frac{2}{q_i}} \\ &= \sum_{i=1}^d \left(\int_0^t \|b_n^i(s) - b_m^i(s)\|^{q_i}_{\widetilde{\mathbb{L}}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i}} \mathrm{d}s\right)^{\frac{2}{q_i}} \\ &\lesssim \sum_{i=1}^d \left(\int_0^t \ell_i^{q_i}(s)\Gamma^{q_i}_{n-1,m-1}(s)\mathrm{d}s\right)^{\frac{2}{q_i}}. \end{aligned}$$

By Pinsker's inequality (2.58), we get

$$\|\rho_t^n - \rho_t^m\|_{\mathbb{L}^1} = \|\mu_t^n - \mu_t^m\|_{\mathrm{var}} \lesssim \sum_{i=1}^d \left(\int_0^t \ell_i^{q_i}(s)\Gamma^{q_i}_{n-1,m-1}(s)\mathrm{d}s\right)^{\frac{1}{q_i}}. \tag{4.14}$$

On the other hand, by (4.5), (4.13) and Hölder's inequality, for $q_i' = \frac{q_i}{q_i-1}$, we have

$$\begin{aligned} \|\rho_t^h - \rho_t^m\|_{\mathbb{L}^\infty} &\lesssim \sum_{i=1}^d \int_0^t (t-s)^{-\frac{1}{2}(1+|\frac{1}{\boldsymbol{p}_i}|)}\ell_i(s)\Gamma_{n-1,m-1}(s)\mathrm{d}s \\ &\lesssim \sum_{i=1}^d \left(\int_0^t (t-s)^{-\frac{q_i'}{2}(1+|\frac{1}{\boldsymbol{p}_i}|)}\mathrm{d}s\right)^{\frac{1}{q_i'}} \left(\int_0^t \ell_i^{q_i}(s)\Gamma^{q_i}_{n-1,m-1}(s)\mathrm{d}s\right)^{\frac{1}{q_i}} \\ &\lesssim \sum_{i=1}^d \left(\int_0^t \ell_i^{q_i}(s)\Gamma^{q_i}_{n-1,m-1}(s)\mathrm{d}s\right)^{\frac{1}{q_i}}, \end{aligned}$$

which together with (4.14) yields

$$\Gamma_{n,m}(t) \lesssim \sum_{i=1}^{d} \left(\int_0^t \ell_i^{q_i}(s) \Gamma_{n-1,m-1}^{q_i}(s) \mathrm{d}s \right)^{\frac{1}{q_i}}.$$

Let $q = q_1 \vee \cdots \vee q_d$. By Hölder's inequality with respect to $\ell_i^{q_i}(s)\mathrm{d}s$, we get

$$\begin{aligned} \Gamma_{n,m}^{q}(t) &\lesssim \sum_{i=1}^{d} \left(\int_0^t \ell_i^{q_i}(s) \Gamma_{n-1,m-1}^{q}(s) \mathrm{d}s \right) \left(\int_0^t \ell_i^{q_i}(s) \mathrm{d}s \right)^{\frac{q}{q_i}-1} \\ &\lesssim \int_0^t \sum_{i=1}^{d} \ell_i^{q_i}(s) \Gamma_{n-1,m-1}^{q}(s) \mathrm{d}s. \end{aligned}$$

Therefore, by (4.10) and the Fatou lemma,

$$\varlimsup_{n,m\to\infty} \Gamma_{n,m}^{q}(t) \lesssim \int_0^t \sum_{i=1}^{d} \ell_i^{q_i}(s) \varlimsup_{n,m\to\infty} \Gamma_{n-1,m-1}^{q}(s) \mathrm{d}s,$$

which implies by the Gronwall inequality that for each $t \in [0,T]$,

$$\varlimsup_{n,m\to\infty} \left(\|\rho_t^h - \rho_t^m\|_{\mathbb{L}^\infty} + \|\rho_t^h - \rho_t^m\|_{\mathbb{L}^1} \right) = \varlimsup_{n,m\to\infty} \Gamma_{n,m}^{q}(t) = 0. \tag{4.15}$$

Now by (4.12), there is a subsequence n_k such that as $k \to \infty$,

$$\mathbb{P}_{n_k} \text{ weakly converges to some } \mathbb{P} \in \mathcal{P}(\mathbb{C}_T),$$

and by (4.15), $\mathbb{P} \circ w_t^{-1}(\mathrm{d}x) = \mu_t(\mathrm{d}x) = \rho_t(x)\mathrm{d}x$ and for each $t \in (0,T]$,

$$\varlimsup_{n\to\infty} \left(\|\rho_t^h - \rho_t\|_{\mathbb{L}^\infty} + \|\rho_t^h - \rho_t\|_{\mathbb{L}^1} \right) = 0. \tag{4.16}$$

(Step 3). In this step we show $\mathbb{P} \in \mathcal{M}_{\mu_0}^{\sigma,b}$. More precisely, we want to show that for fixed $f \in C_c^2(\mathbb{R}^d)$, the process M_t^f defined by (2.56) is a $\mathcal{B}_t$-martingale under $\mathbb{P}$, that is, for any $t_0 < t_1$ and every bounded $\mathcal{B}_{t_0}$-measurable continuous function η,

$$\mathbb{E}\big((M_{t_1}^f - M_{t_0}^f)\eta\big) = 0. \tag{4.17}$$

Note that for each $k \in \mathbb{N}$, by SDE (4.7) and Itô's formula,

$$\mathbb{E}_{\mathbb{P}_{n_k}}\big((M_{t_1}^k - M_{t_0}^k)\eta\big) = 0,$$

where

$$M_t^k := f(w_t) - f(w_0) - \int_0^t \Big(\mathrm{tr}(a_{n_k} \cdot \nabla^2 f) + b_{n_k} \cdot \nabla f \Big)(s, w_s) \mathrm{d}s.$$

Since $x \mapsto a_{n_k}(s,x)$ is continuous, to show (4.17), the key point is to prove the following:

$$\lim_{k\to\infty} \mathbb{E}_{\mathbb{P}_{n_k}}\left(\eta\int_{t_0}^{t_1} b_{n_k}(s,w_s)\cdot\nabla f(s,w_s)\mathrm{d}s\right) = \mathbb{E}\left(\eta\int_{t_0}^{t_1} b(s,w_s,\rho_s(w_s),\mu_s)\cdot\nabla f(s,w_s)\mathrm{d}s\right),$$

which follows from:

$$\lim_{m\to\infty}\sup_k \mathbb{E}_{\mathbb{P}_{n_k}}\left(\int_{t_0}^{t_1}|b_{n_m}(s,w_s)-b(s,w_s,\rho_s(w_s),\mu_s)|\mathrm{d}s\right)=0, \tag{4.18}$$

together with

$$\lim_{k\to\infty} \mathbb{E}_{\mathbb{P}_{n_k}}\left(\eta\int_{t_0}^{t_1} b_{n_m}(s,w_s)\cdot\nabla f(w_s)\mathrm{d}s\right) = \mathbb{E}_{\mathbb{P}}\left(\eta\int_{t_0}^{t_1} b_{n_m}(s,w_s)\cdot\nabla f(w_s)\mathrm{d}s\right) \tag{4.19}$$

for each $m\in\mathbb{N}$. The first limit (4.18) follows by the Krylov estimates (4.11), (4.3) and (4.16). For the second, let

$$b^\varepsilon_{n_m}(s,x) := b_{n_m}(s,\cdot)*\Gamma_\varepsilon(x),\quad \varepsilon\in(0,1),$$

where Γ_ε is the mollifiers in (0.1). For each $\varepsilon\in(0,1)$, since $x\mapsto b^\varepsilon_{n_m}(s,x)$ is bounded continuous, by the weak convergence of $\mathbb{P}_{n_k}$, we have

$$\lim_{k\to\infty} \mathbb{E}_{\mathbb{P}_{n_k}}\left(\eta\int_{t_0}^{t_1} b^\varepsilon_{n_m}(s,w_s)\cdot\nabla f(w_s)\mathrm{d}s\right) = \mathbb{E}_{\mathbb{P}}\left(\eta\int_{t_0}^{t_1} b^\varepsilon_{n_m}(s,w_s)\cdot\nabla f(w_s)\mathrm{d}s\right). \tag{4.20}$$

Moreover, for each $m\in\mathbb{N}$ and $R>0$, by the Krylov estimate (4.11), we also have

$$\begin{aligned}&\lim_{\varepsilon\to 0}\sup_k \mathbb{E}_{\mathbb{P}_{n_k}}\left(\int_{t_0}^{t_1}|b^\varepsilon_{n_m}-b_{n_m}|(s,w_s)|\mathbb{1}_{|w_s|\leqslant R}\mathrm{d}s\right)\\ &\lesssim \lim_{\varepsilon\to 0}\sum_{i=1}^d \|(b^\varepsilon_{n_m}-b_{n_m})^i\mathbb{1}_{B_0^R}\|_{\mathbb{L}_T^{q_i}(\mathbb{L}^{\boldsymbol{p}_i}_{\boldsymbol{\pi}_i})}=0,\end{aligned} \tag{4.21}$$

and

$$\lim_{R\to\infty}\sup_{k,\varepsilon} \mathbb{E}_{\mathbb{P}_{n_k}}\left(\int_{t_0}^{t_1}|b^\varepsilon_{n_m}-b_{n_m}|(s,w_s)|\mathbb{1}_{|w_s|\geqslant R}\mathrm{d}s\right)=0. \tag{4.22}$$

Combining (4.20), (4.21) and (4.22), we obtain (4.19). Thus we complete the proof of existence. On the other hand, by the same calculations as in (4.15), one can show that any two weak solutions have the same marginal distribution. Then by Theorem 3.14, we get the weak uniqueness. □

Remark 4.5. If b does not depend on the density variable r, then we can drop the assumption $\mu_0(\mathrm{d}x)=\rho_0(x)\mathrm{d}x$. In this case, we can only use (4.14) to show that μ^n_t is a Cauchy sequence. We note that a similar result has been shown in [105]. However, even in the non-mixed norm case, the results in [105] do not cover our case since we are using the total variational norm in (4.3). Moreover, our proofs are based on the Fokker-Planck equation, and Wang's proofs are based on the backward Kolmogorov equation.

Chapter 5

Propagation of chaos of McKean-Vlasov SDEs with singular interactions

Let $\phi : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^m$, $F : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ and $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ be Borel measurable functions. For a (sub)-probability measure μ over $\mathbb{R}^d$, we define

$$b(t, x, \mu) := F(t, x, (\phi_t \circledast \mu)(x)),$$

where $\phi_t(x, y) := \phi(t, x, y)$ and

$$(\phi_t \circledast \mu)(x) := \int_{\mathbb{R}^d} \phi_t(x, y)\mu(\mathrm{d}y).$$

Consider the following interacting system of N-particles,

$$\mathrm{d}X_t^{N,i} = b\big(t, X_t^{N,i}, \eta_{\mathbf{X}_t^N}\big)\mathrm{d}t + \sigma\big(t, X_t^{N,i}\big)\mathrm{d}W_t^i, \quad i = 1, \cdots, N, \tag{5.1}$$

where $\mathbf{X}_t^N := (X_t^{N,1}, \cdots, X_t^{N,N})$ and $\eta_{\mathbf{X}_t^N}$ stands for the empirical distribution measure,

$$\eta_{\mathbf{X}_t^N}(\mathrm{d}y) := \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{N,j}}(\mathrm{d}y),$$

and $\{W^i, i \in N\}$ is a sequence of independent standard Brownian motions on some stochastic basis $(\Omega, \mathscr{F}, \mathbb{P}, (\mathscr{F}_t)_{t\geqslant 0})$.

In this chapter we are mainly concerned with the weak and strong convergence of the solutions to (5.1) with general L^p-*singular* interaction $\phi_t(x, y)$ to the solution of the following DDSDE when $N \to \infty$:

$$\mathrm{d}X_t = b(t, X_t, \mu_{X_t})\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t^1, \tag{5.2}$$

where μ_{X_t} denotes the distribution of X_t.

Moreover, we are also interested in the moderately interacting kernel $\phi_t(x,y) = \phi_{\varepsilon_N}(x-y)$, where ϕ_{ε_N} is a family of mollifiers and $\varepsilon_N \to 0$ as $N \to \infty$. In this case, the solution to the interacting particle system

$$\mathrm{d}X_t^{N,i} = F\big(t, X_t^{N,i}, (\phi_{\varepsilon_N} \circledast \eta_{\mathbf{X}_t^N})(X_t^{N,i})\big)\mathrm{d}t + \sigma\big(t, X_t^{N,i}\big)\mathrm{d}W_t^i, \quad i = 1, \cdots, N, \tag{5.3}$$

is expected to converge to the solution of the following dDSDE (see [81, 61]):

$$\mathrm{d}X_t = F(t, X_t, \rho_{X_t}(X_t))\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \tag{5.4}$$

where ρ_{X_t} stands for the density of X_t. Here $\rho := (\rho_{X_t})_{t\geqslant 0}$ solves the following nonlinear and *local (or Nemytskii-type)* Fokker-Planck equation:

$$\partial_t \rho = \partial_i \partial_j (a_{ij}\rho) + \mathrm{div}(F(\rho)\rho). \tag{5.5}$$

It should be kept in mind that for $d = 1$ and $F(\rho) = \rho$, this is Burgers-type equation.

Now, we make the following assumptions for b.

($\mathbf{H}^b$) Suppose that $\phi_t(x,x) = 0$ and for some measurable $h : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}_+$ and $\kappa_1 > 0$,

$$|F(t,x,r)| \leqslant h(t,x) + \kappa_1 |r|, \quad |F(t,x,r) - F(t,x,r')| \leqslant \kappa_1 |r - r'|, \tag{5.6}$$

and for some $(q, \boldsymbol{p}) \in \mathscr{I}^o$ and $\boldsymbol{\pi} \in S_d$ and for any $T > 0$,

$$\|h\|_{\mathbb{L}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \left[\int_0^T \sup_{y\in\mathbb{R}^d} \Big(\|\phi_t(\cdot, y)\|^q_{\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} + \|\phi_t(y, \cdot)\|^q_{\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} \Big) \mathrm{d}t \right]^{\frac{1}{q}} \leqslant \kappa_1. \tag{5.7}$$

Example 1. We provide two examples to illustrate condition (5.7).

(i) Let $d \geqslant 2$ and $\phi_t(x,y) = c_t(x,y)/|x-y|^\alpha$, where $c_t(x,y)$ is bounded measurable and $\alpha \in (0,1)$. It is easy to see that (5.7) holds for q close to ∞ and $p \in (d, \frac{d}{\alpha})$ with $\frac{d}{p} + \frac{2}{q} < 1$.

(ii) Let $d \geqslant 1$ and $\phi_t(x,y) = c_t(x,y)/\Pi_{i=1}^d |x_i - y_i|^{\alpha_i}$, where $\alpha_i \in (0, \frac{1}{2})$ satisfies $\alpha_1 + \cdots + \alpha_d < 1$ and $c_t(x,y)$ is bounded measurable. Note that one can choose q close to ∞ and $p_i > 2$ close to $1/\alpha_i$ so that $|\frac{1}{\boldsymbol{p}}| + \frac{2}{q} < 1$ and (5.7) holds. In this case, the kernel is allowed to have singularities along each axis.

The aim of this chapter is to show the following strong convergence of the particle approximation.

Theorem 5.1. *Let $T > 0$. Under ($\mathbf{H}^\sigma_{\mathbf{mix}}$) and ($\mathbf{H}^b$), for any initial values $\mathbf{X}_0^N$ and X_0, there are unique strong solutions $\mathbf{X}_t^N$ and X_t to particle system (5.1) and DDSDE (5.2), respectively. Moreover, letting μ_0^N be the law of $\mathbf{X}_0^N$ in $\mathbb{R}^{dN}$ and μ_0 the law of X_0 in $\mathbb{R}^d$, we have the following strong convergence results:*

(i) **(Singular kernel)** *Suppose that* μ_0^N *is symmetric and* μ_0*-chaotic, and*

$$\lim_{N\to\infty} \mathbb{E}|X_0^{N,1} - X_0|^2 = 0.$$

Then for any $\gamma \in (0,1)$,

$$\lim_{N\to\infty} \mathbb{E}\left(\sup_{t\in[0,T]} |X_t^{N,1} - X_t|^{2\gamma}\right) = 0. \tag{5.8}$$

(ii) **(Bounded kernel)** *If* h *and* ϕ *in* $(\mathbf{H}^b)$ *are bounded measurable and*

$$\kappa_2 := \sup_N \mathcal{H}\big(\mu_0^N | \mu_0^{\otimes N}\big) < \infty, \tag{5.9}$$

where $\mu_0^{\otimes N} \in \mathcal{P}((\mathbb{R}^d)^N)$ *is the* N*-tensor of* μ_0 *and* $\mathcal{H}$ *stands for the relative entropy (see* (2.57) *below), then for any* $\delta > 2$ *and* $\gamma \in (0,1)$*, there are constants* $C_i = C_i(T,\gamma,\delta,\Theta) > 0$, $i = 1,2$ *independent of* ϕ *and* κ_2 *such that*

$$\mathbb{E}\left(\sup_{t\in[0,T]} |X_t^{N,1} - X_t|^{2\gamma}\right) \leqslant C_1 \mathrm{e}^{C_2\|\phi\|_\infty^\delta} \left(\mathbb{E}|X_0^{N,1} - X_0|^2 + \frac{\kappa_2+1}{N}\right)^\gamma. \tag{5.10}$$

Remark 5.2. If $\sup_N \mathbb{E}|X_0^{N,1}|^p < \infty$ for some $p > 2$, then by interpolation one in fact has

$$\lim_{N\to\infty} \mathbb{E}\left(\sup_{t\in[0,T]} |X_t^{N,1} - X_t|^{p\gamma}\right) = 0, \quad \gamma \in (0,1).$$

The Euler approximation for particle system (5.1) with bounded interaction kernel was studied in [115], which combined with (5.10) implies the full discretization approximation for DDSDE (5.2).

Example 2. Let $d = 1$. Consider the following rank-based interaction:

$$b(t,x,\mu) = F(t,x,\mu(-\infty,x]). \tag{5.11}$$

In this case, the interaction kernel is $\phi(x,y) = \mathbb{1}_{(-\infty,x]}(y) = \mathbb{1}_{x-y\geqslant 0}$, which is bounded and discontinuous. Thus, by (5.10) we have the strong convergence rate of the particle approximation. In particular, if we let $V(x) := \mu((-\infty,x])$, $\sigma(t,x) = \sqrt{2}$ and $F(t,x,r) = g(r)$, then V solves the following Burgers type equation:

$$\partial_t V = \Delta V + \left(\int_0^V g(r)\mathrm{d}r\right)'.$$

For $g(r) = r$, this is the classical Burgers equation. In this way, the above Burgers type equation has been studied in [13, 60, 69]. In the following Example 3, we have another way to simulate Burgers equation via moderate interaction particle system.

Next we turn to the moderate interaction system (5.3) and have the following result.

Theorem 5.3. *Let $T > 0$. Suppose that* $(\mathbf{H}^{\sigma}_{\mathbf{mix}})$ *holds, and*

$$|F(t,x,r)| \leqslant \kappa_1, \quad |F(t,x,r) - F(t,x,r')| \leqslant \kappa_1 |r - r'|, \tag{5.12}$$

and for $\varepsilon_N \in (0,1)$ with $\varepsilon \to 0$ as $N \to \infty$,

$$\phi_t(x,y) = \phi_{\varepsilon_N}(x-y) = \varepsilon_N^{-d}\phi((x-y)/\varepsilon_N),$$

where ϕ is a bounded probability density function in $\mathbb{R}^d$ with support in the unit ball. Then for any initial value X_0 with bounded density ρ_0, there is a unique strong solution X to density-dependent SDE (5.4) such that for each $t > 0$, X_t admits a density ρ_t with

$$\|\rho_t\|_\infty \leqslant C(T,\Theta)\|\rho_0\|_\infty, \ t \in [0,T]. \tag{5.13}$$

Moreover, under (5.9), for any $T > 0$, $\beta \in (0,\gamma_0)$, $\gamma \in (0,1)$ and $\delta > 2$, there are constants $C_i = C_i(T,\beta,\gamma,\delta,\Theta) > 0$, $i = 1,2,3$ such that for all $N \geqslant 2$,

$$\mathbb{E}\left(\sup_{t\in[0,T]} |X_t^{N,1} - X_t|^{2\gamma}\right) \leqslant C_1 \mathrm{e}^{C_2\varepsilon_N^{-\delta d}} \left(\mathbb{E}|X_0^{N,1} - X_0|^2 + \frac{\kappa_2 + 1}{N}\right)^{\gamma} + C_3 \varepsilon_N^{2\beta\gamma}. \tag{5.14}$$

Remark 5.4. Suppose that for some $C > 0$,

$$\mathbb{E}|X_0^{N,1} - X_0|^2 \leqslant C/N.$$

If one chooses $\varepsilon_N = C_4/(\ln N)^{1/(\delta d)}$ with C_4 being large enough, then by (5.14), for some $C > 0$,

$$\mathbb{E}\left(\sup_{t\in[0,T]} |X_t^{N,1} - X_t|^{2\gamma}\right) \leqslant \frac{C}{(\ln N)^{(2\beta\gamma)/(\delta d)}}.$$

Our results weaken the smoothness assumptions on F, ϕ and ρ_0 of Jourdain and Méléard's result (1.14). However, compared to Oelschläger's work [81], the moderate interaction of ε_N can not be chosen with $\varepsilon_N^{-d}/N = o(1)$. It is noted that the results in [81] are only about the weak convergence. In a future work, we shall study the strong convergence when $\varepsilon_N = N^{-\beta}$ for some $\beta > 0$.

Although we assume that F is bounded in (5.12), once we can establish the existence of bounded solutions to the Fokker-Planck equation (5.5) under linear growth assumptions of F in r, then the boundedness of F in (5.12) is no longer a restriction. We illustrate this in the following example.

Example 3. Consider the following special case:

$$\partial_t \rho = \Delta\rho + \mathrm{div}(F(\rho)\rho),$$

where $F : \mathbb{R}_+ \to \mathbb{R}^d$ satisfies $\sum_{i=1}^d |F_i'(r)| \leqslant \kappa_1$. Since the above equation can be written in the following transport form:

$$\partial_t \rho = \Delta \rho + (F(\rho) + F'(\rho)\rho) \cdot \nabla \rho,$$

it is easy to see that by the maximum principle,

$$\|\rho_t\|_\infty \leqslant \|\rho_0\|_\infty.$$

This can be established rigorously by considering the truncated F as $F_n(r) = F(r \wedge n)$, where $n > \|\rho_0\|_\infty$. In particular, the above example covers the one dimensional Burgers equation, i.e., $F(r) = r$. In this case, if one takes $\phi(x) = \mathbb{1}_{[-1,1]}(x)/2$ in (5.3), then

$$(\phi_{\varepsilon_N} \circledast \eta_{\mathbf{X}_t^N})(X_t^{N,i}) = \frac{1}{2N\varepsilon_N} \sum_{j=1}^N \mathbb{1}_{|X_t^{N,i} - X_t^{N,j}| \leqslant \varepsilon_N}.$$

We believe that this is useful for numerical experiments.

5.1 Outline

In this section, we give a brief outline of the proof to Theorem 5.1 and Theorem 5.3.

In Section 5.2, we give two different method to prove weak convergence of N-particle systems (5.1). On the one hand, in Section 5.2.1, by the classical martingale method we show that the propagation of chaos for (5.2) with singular kernels holds in the weak sense, where the key point is to use the partial Girsanov transformation used in [59, 99] to derive some uniform estimate for the exponential functional. Here the strong well-posedness of N-particle systems (5.1) can be used to treat the chaos of the initial distributions. This extends the assumption of i.i.d. initial distributions in [99, 53]. On the other hand, in Section 5.2.2, we also provide a detailed proof for Jabin and Wang's quantitative result [58] for bounded interaction kernels. This is not new and only for the readers' convenience.

In Section 5.3, we give the proof of Theorem 5.1 and show how to use Zvonkin's transformation to derive the strong convergence from weak convergence obtainded in Section 5.2, where the key point is Lemma 5.12.

Finally, in Section 5.4 we apply Zvonkin's transformation again and the stability results obtained in Lemma 4.4 to prove Theorem 5.3.

5.2 Weak convergence

Throughout this section we assume $(\mathbf{H}^\sigma_{\mathbf{mix}})$ and $(\mathbf{H}^b)$. Let

$$\mathbf{X}_t^N := (X_t^{N,1}, \cdots, X_t^{N,N}), \quad \mathbf{W}_t^N := (W^1, \cdots, W^N),$$

and for $\mathbf{x}^N = (x^1, \cdots, x^N)$, define

$$B(t, \mathbf{x}^N) := \left(F\Big(t, x^1, \frac{1}{N}\sum_{j=1}^N \phi_t(x^1, x^j)\Big), \cdots, F\Big(t, x^N, \frac{1}{N}\sum_{j=1}^N \phi_t(x^N, x^j)\Big) \right), \tag{5.15}$$

and a $(dN) \times (dN)$-matrix $\boldsymbol{\sigma}$ by

$$\boldsymbol{\sigma}(t, \mathbf{x}^N) := \mathrm{diag}_N(\sigma(t, x^1), \cdots, \sigma(t, x^N)). \tag{5.16}$$

Then the particle system (5.1) can be written as an SDE in $\mathbb{R}^{dN}$:

$$\mathrm{d}\mathbf{X}_t^N = B(t, \mathbf{X}_t^N)\mathrm{d}t + \boldsymbol{\sigma}(t, \mathbf{X}_t^N)\mathrm{d}\mathbf{W}_t^N.$$

Noting that by $(\mathbf{H}^b)$,

$$|B_i(t, \mathbf{x}^N)| \leqslant h(t, x^i) + \frac{\kappa_1}{N}\sum_{j=1}^N |\phi_t(x^i, x^j)|,$$

we have for $\vec{\boldsymbol{p}} = (\infty, \cdots, \infty, \boldsymbol{p}) \in [1, \infty]^{dN}$ and for $\boldsymbol{\pi}_i = (1, \cdots, i-1, i+1, \cdots, N, i)$,

$$\|B_i\|_{\mathbb{L}^q_T(\widetilde{\mathbb{L}}^{\vec{\boldsymbol{p}}}_{\boldsymbol{\pi}_i})} \leqslant \|h\|_{\mathbb{L}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} + \kappa_1 \left[\int_0^T \sup_{y\in\mathbb{R}^d} \|\phi_t(\cdot, y)\|^q_{\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}} \mathrm{d}t\right]^{\frac{1}{q}} < \infty.$$

Then, by Theorem 3.1, for any initial value $\mathbf{X}_0^N$, there is a unique strong solution to the above SDE. In particular, there is a measurable functional $\Phi : \mathbb{R}^{dN} \times \mathbb{C}_T^N \to \mathbb{C}_T^N$ such that

$$\mathbf{X}_t^N = \Phi(\mathbf{X}_0^N, \mathbf{W}_\cdot^N)(t), \quad t \in [0, T]. \tag{5.17}$$

5.2.1 Martingale approach

In this section we use the classical martingale approach to show the following qualitative result of weak convergence.

Theorem 5.5. *For any $N \in \mathbb{N}$, let $\xi_1^N, \cdots, \xi_N^N$ be N-random variables and $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$. Suppose that the law of $(\xi_1^N, \cdots, \xi_N^N)$ is invariant under any permutation of $\{1, \cdots, N\}$, and for any $k \leqslant N$,*

$$\mathbb{P} \circ \left(\xi_1^N, \cdots, \xi_k^N\right)^{-1} \to \mu_0^{\otimes k}, \quad N \to \infty. \tag{5.18}$$

Then for any $k \leqslant N$ and $T > 0$,

$$\mathbb{P} \circ \left(X^{N,1}_{[0,T]}, \cdots, X^{N,k}_{[0,T]}\right)^{-1} \to \mu^{\otimes k}_{[0,T]}, \quad N \to \infty, \tag{5.19}$$

where $\mu_{[0,T]}$ is the law of the unique solution of dDDSDE (5.2) with initial distribution μ_0 on $\mathbb{C}_T$.

First of all, we use the partial Girsanov transform as used in [59, 99] to show some uniform Krylov estimate for particle system (5.1). Let $\{\widetilde{W}^i_t, i \in \mathbb{N}\}$ be a sequence of independent d-dimensional standard Brownian motions. For each $x \in \mathbb{R}^d$, let $Z_t(x)$ be the unique strong solution of the following SDE starting from x:

$$\mathrm{d}Z_t = \sigma\big(t, Z_t\big)\mathrm{d}\widetilde{W}^1_t, \quad Z_0 = x.$$

For each $\boldsymbol{z} = (z^2, \cdots, z^N) \in \mathbb{R}^{(N-1)d}$, let $\boldsymbol{Z}^N_t(\boldsymbol{z}) := \boldsymbol{Z}^N_t := (Z^{N,2}_t, \cdots, Z^{N,N}_t)$ be the unique strong solution of the following SDE starting from $\boldsymbol{z}$:

$$\mathrm{d}Z^{N,k}_t = b\big(t, Z^{N,k}_t, \eta_{\boldsymbol{Z}^N_t}\big)\mathrm{d}t + \sigma\big(t, Z^{N,k}_t\big)\mathrm{d}\widetilde{W}^k_t, \quad Z^{N,k}_0 = z^k,$$

where $k = 2, \cdots, N$ and

$$\eta_{\boldsymbol{z}}(\mathrm{d}y) := \frac{1}{N}\sum_{j=2}^{N} \delta_{z^j}(\mathrm{d}y).$$

In particular, as Brownian functionals of $\widetilde{W}^1$ and $(\widetilde{W}^2, \cdots, \widetilde{W}^N)$ respectively,

$$Z_\cdot(\cdot) \text{ is independent of } \boldsymbol{Z}^N_\cdot(\cdot), \tag{5.20}$$

and by the notion of strong solution of SDEs (see (5.17)),

$$\widetilde{X}^{N,1}_t := Z_t(\xi^N_1), \quad (\widetilde{X}^{N,2}_t, \cdots, \widetilde{X}^{N,N}_t) := \boldsymbol{Z}^N_t(\xi^N_2, \cdots, \xi^N_N) =: \boldsymbol{Y}^N_t, \tag{5.21}$$

solves the following SDE:

$$\begin{cases} \mathrm{d}\widetilde{X}^{N,1}_t = \sigma\big(t, \widetilde{X}^{N,1}_t\big)\mathrm{d}\widetilde{W}^1_t, \quad \widetilde{X}^{N,1}_0 = \xi^N_1, \\ \qquad \text{and for each } k = 2, \cdots, N, \\ \mathrm{d}\widetilde{X}^{N,k}_t = b\big(t, \widetilde{X}^{N,k}_t, \eta_{\boldsymbol{Y}^N_t}\big)\mathrm{d}t + \sigma\big(t, \widetilde{X}^{N,k}_t\big)\mathrm{d}\widetilde{W}^k_t, \quad \widetilde{X}^{N,k}_0 = \xi^N_k, \end{cases} \tag{5.22}$$

where

$$\eta_{\boldsymbol{Y}^N_t} := \frac{1}{N}\sum_{j=2}^{N} \delta_{\widetilde{X}^{N,j}_t}(\mathrm{d}y).$$

Now let us define

$$\eta_{\widetilde{\mathbf{X}}^N_t}(\mathrm{d}y) := \frac{1}{N}\sum_{j=1}^{N} \delta_{\widetilde{X}^{N,j}_t}(\mathrm{d}y), \quad H^{N,1}_t := \sigma\big(t, \widetilde{X}^{N,1}_t\big)^{-1} b\big(t, \widetilde{X}^{N,1}_t, \eta_{\widetilde{\mathbf{X}}^N_t}\big),$$

and for $k = 2, \cdots, N$,

$$H^{N,k}_t := \sigma\big(t, \widetilde{X}^{N,k}_t\big)^{-1}\Big[b\big(t, \widetilde{X}^{N,k}_t, \eta_{\widetilde{\mathbf{X}}^N_t}\big) - b\big(t, \widetilde{X}^{N,k}_t, \eta_{\boldsymbol{Y}^N_t}\big)\Big].$$

By the above definition, we clearly have for each $i = 1, \cdots, N$,

$$\mathrm{d}\widetilde{X}_t^{N,i} = b\big(t, \widetilde{X}_t^{N,i}, \eta_{\widetilde{\mathbf{X}}_t^N}\big)\mathrm{d}t + \sigma\big(t, \widetilde{X}_t^{N,i}\big)\big(\mathrm{d}\widetilde{W}_t^i - H_t^{N,i}\mathrm{d}t\big). \tag{5.23}$$

The following uniform estimate is the key step for performing the Girsanov transform to derive the Krylov estimate for the particle system, whose proof strongly depends on the independence in (5.20) and the strong uniqueness used in (5.22).

Lemma 5.6. *For any* $\gamma, T > 0$,

$$\sup_N \mathbb{E} \exp\left\{\gamma \sum_{i=1}^N \int_0^T |H_t^{N,i}|^2 \mathrm{d}t\right\} < \infty. \tag{5.24}$$

Proof. For $x \in \mathbb{R}^d$ and $\boldsymbol{y} = (y^2, \cdots, y^N) \in \mathbb{R}^{(N-1)d}$, let us write $\eta_{\boldsymbol{y}} := \frac{1}{N}\sum_{j=2}^N \delta_{y^j}$ and define

$$\Gamma_1(t, x, \boldsymbol{y}) := \sigma(t, x)^{-1} b\left(t, x, \frac{\delta_x}{N} + \eta_{\boldsymbol{y}}\right),$$

and for $k = 2, \cdots, N$,

$$\Gamma_k(t, x, \boldsymbol{y}) := \sigma(t, y^k)^{-1}\left[b\left(t, y^k, \frac{\delta_x}{N} + \eta_{\boldsymbol{y}}\right) - b\left(t, y^k, \eta_{\boldsymbol{y}}\right)\right].$$

From the very definition, one sees that for each $i = 1, \cdots, N$,

$$H_s^{N,i} = \Gamma_i\big(s, \widetilde{X}_s^{N,1}, \boldsymbol{Y}_s^N\big),$$

and by (5.21) and (5.20),

$$\begin{aligned}
\mathbb{E} \exp\left\{\gamma \sum_{i=1}^N \int_0^T |H_s^{N,i}|^2 \mathrm{d}s\right\} &= \mathbb{E} \exp\left\{\gamma \int_0^T \sum_{i=1}^N |\Gamma_i\big(s, Z_s(\xi_1^N), \boldsymbol{Y}_s^N\big)|^2 \mathrm{d}s\right\} \\
&= \mathbb{E}\left(\mathbb{E} \exp\left\{\gamma \int_0^T \sum_{i=1}^N |\Gamma_i\big(s, Z_s(x), \boldsymbol{y}_s\big)|^2 \mathrm{d}s\right\}\Big|_{(x, \boldsymbol{y}_\cdot) = (\xi_1^N, \boldsymbol{Y}_\cdot^N)}\right) \\
&\leqslant \sup_{x, \boldsymbol{y}_\cdot} \mathbb{E} \exp\left\{\gamma \int_0^T \sum_{i=1}^N |\Gamma_i\big(s, Z_s(x), \boldsymbol{y}_s\big)|^2 \mathrm{d}s\right\} \\
&= \sup_{x, \boldsymbol{y}_\cdot} \mathbb{E} \exp\left\{\gamma \int_0^T f_{\boldsymbol{y}}\big(s, Z_s(x)\big)\mathrm{d}s\right\},
\end{aligned} \tag{5.25}$$

where for $\boldsymbol{y} = (\boldsymbol{y}_s)_{s \in [0,T]}$,

$$f_{\boldsymbol{y}}(s, x) := \sum_{i=1}^N |\Gamma_i\big(s, x, \boldsymbol{y}_s\big)|^2.$$

Note that by (5.6) and because $\phi_t(x,x)=0$,

$$|\Gamma_1(t,x,\boldsymbol{y})| = \left|\sigma(t,x)^{-1}F\left(t,x,\frac{1}{N}\Big(\phi_t(x,x)+\sum_{j=2}^N|\phi_t(x,y^j)|\Big)\right)\right|$$
$$\leqslant \|\sigma^{-1}\|_\infty\left(h(t,x)+\frac{\kappa_1}{N}\sum_{j=2}^N\phi_t(x,y^j)\right),$$

and

$$|\Gamma_k(t,x,\boldsymbol{y})| \leqslant \frac{\kappa_1\|\sigma^{-1}\|_\infty}{N}|\phi_t(y^k,x)|,$$

and by (5.7),

$$\left(\int_0^T \sup_{\boldsymbol{y}} \|\Gamma_1(t,\cdot,\boldsymbol{y})\|^q_{\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}}\mathrm{d}t\right)^{1/q} \leqslant \|\sigma^{-1}\|_\infty(\kappa_1+\kappa_1^2)$$

and

$$\left(\int_0^T \sup_{\boldsymbol{y}} \|\Gamma_k(t,\cdot,\boldsymbol{y})\|^q_{\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}}\mathrm{d}t\right)^{1/q} \leqslant \frac{\kappa_1^2\|\sigma^{-1}\|_\infty}{N-1}.$$

From these two estimates, by Minkowskii's inequality, we derive

$$\left(\int_0^T \sup_{\boldsymbol{y}} \|f_{\boldsymbol{y}\cdot}(s,\cdot)\|^{q/2}_{\widetilde{\mathbb{L}}^{\boldsymbol{p}/2}_{\boldsymbol{\pi}}}\mathrm{d}t\right)^{2/q} \leqslant \sum_{i=1}^N\left(\int_0^T \sup_{\boldsymbol{y}} \||\Gamma_i(t,\cdot,\boldsymbol{y})|^2\|^{q/2}_{\widetilde{\mathbb{L}}^{\boldsymbol{p}/2}_{\boldsymbol{\pi}}}\mathrm{d}t\right)^{2/q}$$
$$= \sum_{i=1}^N\left(\int_0^T \sup_{\boldsymbol{y}} \|\Gamma_i(t,\cdot,\boldsymbol{y})\|^{q}_{\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}}}\mathrm{d}t\right)^{2/q}$$
$$\leqslant \|\sigma^{-1}\|^2_\infty\left((\kappa_1+\kappa_1^2)^2+\frac{\kappa_1^4}{N}\right).$$

Thus, because $(\frac{q}{2},\frac{\boldsymbol{p}}{2})\in\mathscr{I}_2$, by (3.46) we have

$$\sup_{x,\boldsymbol{y}}\mathbb{E}\exp\left\{\gamma\int_0^T f_{\boldsymbol{y}}\big(s,Z_s(x)\big)\mathrm{d}s\right\}\leqslant C,$$

which together with (5.25) yields (5.24). □

Now if we define

$$\mathscr{E}_t^N := \exp\left\{\sum_{i=1}^N\int_0^t H_s^{N,i}\mathrm{d}\widetilde{W}_s^i-\frac{1}{2}\sum_{i=1}^N\int_0^t|H_s^{N,i}|^2\mathrm{d}s\right\},$$

then by (5.24) and Novikov's criterion, $t \mapsto Z_t$ is an exponential martingale and

$$\mathscr{E}_t^N = 1 + \sum_{i=1}^N \int_0^t H_s^{N,i} \mathscr{E}_s^N \mathrm{d}\widetilde{W}_s^i.$$

Thus, by Girsanov's theorem, $\big(\widetilde{W}_t^i - \int_0^t H_s^{N,i}\mathrm{d}s\big)_{t\in[0,T]}^{i=1,\cdots,N}$ are N-independent standard Brownian motions under the new probability measure

$$\mathbb{Q} := \mathscr{E}_T^N \mathbb{P}.$$

Moreover, by (5.23) and the weak uniqueness for SDE (5.1), we have

$$\mathbb{Q} \circ \big(\widetilde{\mathbf{X}}_{[0,T]}^N\big)^{-1} = \mathbb{P} \circ \big(\mathbf{X}_{[0,T]}^N\big)^{-1}, \tag{5.26}$$

and for any $\gamma \in \mathbb{R}$, by (5.24) it is standard to derive that

$$\sup_N \mathbb{E}\left(\sup_{t\in[0,T]} |\mathscr{E}_t^N|^\gamma\right) < \infty. \tag{5.27}$$

From these, we can derive the following crucial Krylov estimate for the particle system.

Lemma 5.7. *(i) The law of $(X_t^{N,1})_{t\in[0,T]}$, $N \in \mathbb{N}$, in $\mathbb{C}_T$ is tight.*

(ii) For any $T > 0$, $(q, \boldsymbol{p}) \in \mathscr{I}_2$ and $\boldsymbol{\pi} \in S_d$, there is a constant $C_1 = C_1(T, \Theta) > 0$ such that for any $f \in \widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})$,

$$\sup_N \mathbb{E}\left(\int_0^T f\big(t, X_t^{N,1}\big)\mathrm{d}t\right) \leqslant C_1 |\!|\!| f |\!|\!|_{\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})}, \tag{5.28}$$

and for any $\lambda > 0$ and $\beta \in (0, 2 - |\frac{1}{\boldsymbol{p}}| - \frac{2}{q})$, there is a $C_2 = C_2(T, \Theta, \lambda, \beta) > 0$ such that for any $f \in \widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})$,

$$\sup_N \mathbb{E}\exp\left\{\lambda \int_0^T f\big(t, X_t^{N,1}\big)\mathrm{d}t\right\} \leqslant \mathrm{e}^{C_2 |\!|\!| f |\!|\!|^{2/\beta}_{\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})}}. \tag{5.29}$$

(iii) Let $\boldsymbol{p}_1, \boldsymbol{p}_2 \in (1,\infty)^d$ and let $q \in (1,\infty)$ with $|\frac{1}{\boldsymbol{p}_1}| + |\frac{1}{\boldsymbol{p}_2}| + \frac{2}{q} < 2$ and $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2 \in S_d$. Then for any $T > 0$, it holds that for some $C_3 = C_3(T, \Theta) > 0$,

$$\sup_N \mathbb{E}\left(\int_0^T f\big(t, X_t^{N,1}, X_t^{N,2}\big)\mathrm{d}t\right) \leqslant C_3 |\!|\!| f |\!|\!|_{\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}_1}^{\boldsymbol{p}_1}(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}_2}^{\boldsymbol{p}_2}))}, \tag{5.30}$$

where $\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}_1}^{\boldsymbol{p}_1}(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}_2}^{\boldsymbol{p}_2}))$ is the localization of $\mathbb{L}_T^q(\mathbb{L}_{\boldsymbol{\pi}_1}^{\boldsymbol{p}_1}(\mathbb{L}_{\boldsymbol{\pi}_2}^{\boldsymbol{p}_2}))$ as in (2.10).

Proof. (i) By (5.26), Hölder's inequality, (5.27) and (5.22), there is a constant $C > 0$ such that for all $0 \leqslant s < t \leqslant T$ and $N \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}|X_t^{N,1} - X_s^{N,1}|^4 &= \mathbb{E}\big(\mathscr{E}_T^N |\widetilde{X}_t^{N,1} - \widetilde{X}_s^{N,1}|^4\big)\\ &\leqslant \big(\mathbb{E}(\mathscr{E}_T^N)^2\big)^{1/2}\big(\mathbb{E}|\widetilde{X}_t^{N,1} - \widetilde{X}_s^{N,1}|^8\big)^{1/2} \leqslant C|t-s|^2,\end{aligned}$$

which, together with (5.18), implies the tightness by Kolmogorov's criterion.

(ii) Let $\gamma > 1$ be such that $(\frac{q}{\gamma}, \frac{\boldsymbol{p}}{\gamma}) \in \mathscr{I}_2$. By (5.26), Hölder's inequality, (5.27) and (3.27), we have

$$\begin{aligned}\mathbb{E}\left(\int_0^T f(t, X_t^{N,1})\mathrm{d}t\right) &= \mathbb{E}\left(\mathscr{E}_T^N \int_0^T f(t, \widetilde{X}_t^{N,1})\mathrm{d}t\right)\\ &\leqslant \left[\mathbb{E}(\mathscr{E}_T^N)^{\frac{\gamma}{\gamma-1}}\right]^{1-1/\gamma}\left[\mathbb{E}\left(\int_0^T |f(t, \widetilde{X}_t^{N,1})|^\gamma \mathrm{d}t\right)\right]^{1/\gamma}\\ &\leqslant C\|f^\gamma\|^{1/\gamma}_{\widetilde{\mathbb{L}}_T^{q/\gamma}(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}/\gamma})} = C\|f\|_{\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})}.\end{aligned}$$

(5.29) follows by the same method and (3.46).

(iii) Let $\gamma \in (1, \min_i(p_{1i}, p_{2i}) \wedge q)$ be such that $|\frac{1}{\boldsymbol{p}_1/\gamma}| + |\frac{1}{\boldsymbol{p}_2/\gamma}| + \frac{2}{q/\gamma} < 2$. By (5.26), Hölder's inequality and (5.27), we have

$$\begin{aligned}\mathbb{E}\left(\int_0^T f(t, X_t^{N,1}, X_t^{N,2})\mathrm{d}t\right) &= \mathbb{E}\left(\mathscr{E}_T^N \int_0^T f(t, \widetilde{X}_t^{N,1}, \widetilde{X}_t^{N,2})\mathrm{d}t\right)\\ &\leqslant \left[\mathbb{E}(\mathscr{E}_T^N)^{\frac{\gamma}{\gamma-1}}\right]^{\frac{\gamma-1}{\gamma}}\left[\mathbb{E}\left(\int_0^T |f(t, \widetilde{X}_t^{N,1}, \widetilde{X}_t^{N,2})|^\gamma \mathrm{d}t\right)\right]^{\frac{1}{\gamma}}\\ &\lesssim \sup_x \left[\mathbb{E}\left(\int_0^T |f(t, Z_t(x), \widetilde{X}_t^{N,2})|^\gamma \mathrm{d}t\right)\right]^{\frac{1}{\gamma}}.\end{aligned}$$

By $|\frac{1}{\boldsymbol{p}_1/\gamma}| + |\frac{1}{\boldsymbol{p}_2/\gamma}| + \frac{2}{q/\gamma} < 2$, one can choose $q_1, q_2 > \gamma$ so that $\frac{1}{q_1/\gamma} + \frac{1}{q_2/\gamma} = 1 + \frac{1}{q/\gamma}$ and $(q_i/\gamma, \boldsymbol{p}_i/\gamma) \in \mathscr{I}_2$, $i = 1, 2$. Since $Z_\cdot(x)$ and $\widetilde{X}^{N,2}$ are independent by (5.20) and (5.21), and satisfy the Krylov estimate (5.28), the desired estimate now follows by using [89, Lemma 2.6]. □

In the following, in order to take weak limits, we need to mollify the coefficients. For $\varepsilon \in (0, 1)$ and $k \in \mathbb{N}$, we define

$$b_{\varepsilon,k}(t, x, \mu) := F_\varepsilon(t, x, (\phi_t^k \circledast \mu)(x)), \tag{5.31}$$

where

$$F_\varepsilon(t, x, r) := (-\varepsilon^{-1}) \vee \big((F(t, \cdot, r) * \Gamma_\varepsilon)(x)\big) \wedge \varepsilon^{-1}$$

and

$$\phi_t^k(x,y) := (-k) \vee \big((\phi_t * \Gamma_{1/k})(x,y)\big) \wedge k.$$

We have the following properties for the above approximation.

Lemma 5.8. *(i)* $b_{\varepsilon,k} \in L_T^\infty(C_b(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)))$ *and*

$$|b_{\varepsilon,k}(t,x,\mu)| \leqslant h_t * \Gamma_\varepsilon(x) + \kappa_0(\phi_t^k \circledast \mu) * \Gamma_\varepsilon(x)$$

and

$$|b - b_{\varepsilon,k}|\big(t,x,\mu\big) \leqslant \sup_{|r|\leqslant k} |F_\varepsilon - F|(t,x,r) + \kappa_0|(\phi_t^k - \phi_t) \otimes \mu|(x).$$

(ii) For any $T > 0$,

$$\lim_{k\to\infty} \lim_{\varepsilon\to 0} \sup_N \mathbb{E}\left(\int_0^T |b - b_{\varepsilon,k}|\big(s, X_s^{N,1}, \eta_{\mathbf{X}_s^N}\big)\mathrm{d}s\right) = 0. \tag{5.32}$$

Proof. (i) is obvious by definition and the assumptions. We now show (ii). Note that

$$|b - b_{\varepsilon,k}|\big(s, X_s^{N,1}, \eta_{\mathbf{X}_s^N}\big) \leqslant \sup_{|r|\leqslant k} |F_\varepsilon - F|(s, X_s^{N,1}, r) + \frac{\kappa_0}{N}\sum_{j=1}^N |\phi_s^k - \phi_s|(X_s^{N,1}, X_s^{N,j}). \tag{5.33}$$

We first show that for fixed $r \in \mathbb{R}^m$,

$$\lim_{\varepsilon\to 0} \sup_N \mathbb{E}\left(\int_0^T |F_\varepsilon - F|(s, X_s^{N,1}, r)\mathrm{d}s\right) = 0. \tag{5.34}$$

Let $R > 0$. Since $(\frac{q}{2}, \frac{\boldsymbol{p}}{2}) \in \mathscr{I}_2$ and $\|F(\cdot,r)\|^2_{\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})} < \infty$ by (5.6) and (5.7), by Hölder's inequality and (5.28), (2.12), we have

$$\begin{aligned}
&\mathbb{E}\left(\int_0^T |F_\varepsilon - F|(s, X_s^{N,1}, r)\mathbb{1}_{|X_s^{N,1}|>R}\mathrm{d}s\right) \\
&\leqslant \left[\mathbb{E}\left(\int_0^T |F_\varepsilon - F|^2(s, X_s^{N,1}, r)\mathrm{d}s\right)\right]^{\frac{1}{2}} \left[\int_0^T \mathbb{P}(|X_s^{N,1}| > R)\mathrm{d}s\right]^{\frac{1}{2}} \\
&\lesssim \||F_\varepsilon - F|^2(\cdot,r)\|^{1/2}_{\widetilde{\mathbb{L}}_T^{q/2}(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}/2})} \left[\int_0^T \Big(\mathbb{P}\big(|X_s^{N,1} - X_0^{N,1}| > \tfrac{R}{2}\big) + \mathbb{P}\big(|X_0^{N,1}| > \tfrac{R}{2}\big)\Big)\mathrm{d}s\right]^{\frac{1}{2}} \\
&\lesssim \|F(\cdot,r)\|_{\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})} \left[\int_0^T \left(\frac{\mathbb{E}|X_s^{N,1} - X_0^{N,1}|}{R} + \mathbb{P}\big(|X_0^{N,1}| > \tfrac{R}{2}\big)\right)\mathrm{d}s\right]^{\frac{1}{2}} \\
&\lesssim \|F(\cdot,r)\|_{\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})} \left[\frac{C}{R} + \mathbb{P}\big(|\xi_1^N| > \tfrac{R}{2}\big)\right]^{\frac{1}{2}} \to 0, \quad R \to \infty.
\end{aligned} \tag{5.35}$$

On the other hand, for each $R > 0$, by (5.28) again, we have

$$\mathbb{E}\left(\int_0^T |F_\varepsilon - F|(s, X_s^{N,1}, r)\mathbb{1}_{|X_s^{N,1}|\leqslant R}\mathrm{d}s\right) \lesssim \|(F_\varepsilon - F)(\cdot, r)\mathbb{1}_{B_R}\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \overset{(2.13)}{\to} 0, \quad \varepsilon \to 0,$$

which together with (5.35) yields (5.34).

Since $|F_\varepsilon(t, x, r) - F_\varepsilon(t, x, r')| \leqslant \kappa_0|r - r'|$, by (5.34) and a finite covering technique, for each $k \in \mathbb{N}$, we further have

$$\lim_{\varepsilon\to 0}\sup_N \mathbb{E}\left(\int_0^T \sup_{|r|\leqslant k} |F_\varepsilon - F|(s, X_s^{N,1}, r)\mathrm{d}s\right) = 0. \tag{5.36}$$

Indeed, for any given $\delta > 0$, one can find M-balls in $\mathbb{R}^m$ with centers in $\{r_i, i = 1, \cdots, M\}$ and radius δ such that

$$\{r : |r| \leqslant k\} \subset \cup_{i=1,\cdots,M} B_\delta(r_i).$$

Thus,

$$\mathbb{E}\left(\int_0^T \sup_{|r|\leqslant k} |F_\varepsilon - F|(s, X_s^{N,1}, r)\mathrm{d}s\right) \leqslant \sum_{i=1}^M \mathbb{E}\left(\int_0^T |F_\varepsilon - F|(s, X_s^{N,1}, r_i)\mathrm{d}s\right) + \kappa_0\delta.$$

By (5.34) and firstly letting $\varepsilon \to 0$ and then $\delta \to 0$, we get (5.36).

Moreover, for $j \neq 1$, since

$$\mathbb{E}\left(\int_0^T |\phi_s^k - \phi_s|(X_s^{N,1}, X_s^{N,j})\mathrm{d}s\right) = \mathbb{E}\left(\int_0^T |\phi_s^k - \phi_s|(X_s^{N,1}, X_s^{N,2})\mathrm{d}s\right),$$

as in proving (5.34) and by (5.7) and (5.30), we also have

$$\lim_{k\to\infty}\sup_N \mathbb{E}\left(\int_0^T |\phi_s^k - \phi_s|(X_s^{N,1}, X_s^{N,2})\mathrm{d}s\right) = 0,$$

and because $\phi_s(x, x) = 0$ and (5.28),

$$\lim_{k\to\infty}\sup_N \mathbb{E}\left(\int_0^T |\phi_s^k|(X_s^{N,1}, X_s^{N,1})\mathrm{d}s\right) = 0.$$

Hence,

$$\lim_{k\to\infty}\sup_N \sup_{j=1,\cdots,N} \mathbb{E}\left(\int_0^T |\phi_s^k - \phi_s|(X_s^{N,1}, X_s^{N,j})\mathrm{d}s\right) = 0,$$

which together with (5.36) and (5.33) yields (5.32). □

Now we are ready to give the

Proof of Theorem 5.5. Consider the following random measure with values in $\mathcal{P}(\mathbb{C}_T)$,

$$\omega \to \Pi_N(\omega, \mathrm{d}w) := \frac{1}{N}\sum_{i=1}^N \delta_{X_\cdot^{N,i}(\omega)}(\mathrm{d}w).$$

By (i) of Lemma 5.7 and [98, (ii) of Proposition 2.2], the laws of Π_N, $N \in \mathbb{N}$, are tight in $\mathcal{P}(\mathcal{P}(\mathbb{C}_T))$. Without loss of generality, we assume that the laws of Π_N weakly converge to some $\Pi_\infty \in \mathcal{P}(\mathcal{P}(\mathbb{C}_T))$. By (5.28) and (5.30), it is standard to derive that for any $(q, \boldsymbol{p}) \in \mathscr{I}_2$ and $f \in \widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})$ (see [110, Remark 3.4]),

$$\left|\int_{\mathcal{P}(\mathbb{C}_T)}\int_{\mathbb{C}_T}\left(\int_0^T f(s, w_s)\mathrm{d}s\right)\nu(\mathrm{d}w)\Pi_\infty(\mathrm{d}\nu)\right| \leqslant C|\!|\!|f|\!|\!|_{\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}}^{\boldsymbol{p}})}, \tag{5.37}$$

and for any $\boldsymbol{p}_1, \boldsymbol{p}_2 \in (1,\infty)^d$ and $q \in (1,\infty)$ with $|\frac{1}{\boldsymbol{p}_1}| + |\frac{1}{\boldsymbol{p}_2}| + \frac{2}{q} < 2$, and $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2 \in S_d$, $f \in \widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}_1}^{\boldsymbol{p}_1}(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}_2}^{\boldsymbol{p}_2}))$,

$$\left|\int_{\mathcal{P}(\mathbb{C}_T)}\int_{\mathbb{C}_T}\int_{\mathbb{C}_T}\left(\int_0^T f(s, w_s, w'_s)\mathrm{d}s\right)\nu(\mathrm{d}w')\nu(\mathrm{d}w)\Pi_\infty(\mathrm{d}\nu)\right| \leqslant C|\!|\!|f|\!|\!|_{\widetilde{\mathbb{L}}_T^q(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}_1}^{\boldsymbol{p}_1}(\widetilde{\mathbb{L}}_{\boldsymbol{\pi}_2}^{\boldsymbol{p}_2}))}. \tag{5.38}$$

Our aim below is to show that Π_∞ is a Dirac measure, i.e.,

$$\Pi_\infty(\mathrm{d}\nu) = \delta_\mu(\mathrm{d}\nu), \quad \Pi_\infty - a.s.,$$

where $\mu \in \mathcal{M}_{\mu_0}^{\sigma,b}$ is the unique martingale solution of dDDSDE with initial distribution μ_0.

We divide the proofs into two steps.

(Step 1) For given $f \in C_0^2(\mathbb{R}^d)$ and $\nu \in \mathcal{P}(\mathbb{C}_T)$, we define a functional on $\mathbb{C}_T$ by

$$M_{f,\nu}^{\sigma,b}(t,w) := f(w_t) - f(w_0) - \int_0^t \mathscr{L}_\nu^{\sigma,b} f(s, w_s)\mathrm{d}s, \ t \in [0,T],$$

where

$$\mathscr{L}_\nu^{\sigma,b} f(s,x) := \tfrac{1}{2}\mathrm{tr}(\sigma\sigma^* \cdot \nabla^2 f)(s,x) + b(s,x,\nu_s)\cdot\nabla f(x),$$

and

$$\nu_s := \nu \circ w_s^{-1} \text{ is the marginal distribution of } \nu \text{ at time } s.$$

Fix $n \in \mathbb{N}$. For given $g \in C_0(\mathbb{R}^{nd})$ and $0 \leqslant s_1 < \cdots < s_n \leqslant s$, we also introduce a functional Ξ_f^g on $\mathcal{P}(\mathbb{C}_T)$ by

$$\Xi_f^g(\nu) := \int_{\mathbb{C}_T}\big(M_{f,\nu}^{\sigma,b}(t,w) - M_{f,\nu}^{\sigma,b}(s,w)\big)g(w_{s_1},\cdots,w_{s_n})\nu(\mathrm{d}w).$$

In particular,

$$\Xi_f^g(\Pi_N) = \frac{1}{N}\sum_{i=1}^N \big(M_{f,\Pi_N}^{\sigma,b}(t, X_\cdot^{N,i}) - M_{f,\Pi_N}^{\sigma,b}(s, X_\cdot^{N,i})\big) g\big(X_{s_1}^{N,i}, \cdots, X_{s_n}^{N,i}\big) \tag{5.39}$$

and

$$\Pi_N \circ w_s^{-1} = \eta_{\mathbf{X}_s^N}.$$

Noting that by Itô's formula,

$$\begin{aligned} M_{f,\Pi_N}^{\sigma,b}(t, X_\cdot^{N,i}) &= f(X_t^{N,i}) - f(X_0^{N,i}) - \int_0^t \mathscr{L}_{\Pi_N}^{\sigma,b} f(s, X_s^{N,i}) \mathrm{d}s \\ &= \int_0^t (\sigma^* \cdot \nabla f)\big(s, X_s^{N,i}\big) \mathrm{d}W_s^i, \end{aligned}$$

by (5.39) and the Itô isometry for stochastic integrals, we have

$$\begin{aligned} \mathbb{E}|\Xi_f^g(\Pi_N)|^2 &= \frac{1}{N^2}\mathbb{E}\left|\sum_{i=1}^N \int_s^t (\sigma^* \cdot \nabla f)\big(r, X_r^{N,i}\big) g\big(X_{s_1}^{N,i}, \cdots, X_{s_n}^{N,i}\big) \mathrm{d}W_r^i\right|^2 \\ &= \frac{1}{N^2}\sum_{i=1}^N \int_s^t \mathbb{E}\big|(\sigma^* \cdot \nabla f)\big(r, X_r^{N,i}\big) g\big(X_{s_1}^{N,i}, \cdots, X_{s_n}^{N,i}\big)\big|^2 \mathrm{d}r \\ &\leqslant \frac{1}{N}(t-s)\|\sigma^* \cdot \nabla f\|_\infty^2 \|g\|_\infty^2. \end{aligned} \tag{5.40}$$

Suppose that we have proven

$$\lim_{N\to\infty} \mathbb{E}|\Xi_f^g(\Pi_N)| = \int_{\mathcal{P}(\mathbb{C}_T)} |\Xi_f^g(\nu)| \Pi_\infty(\mathrm{d}\nu). \tag{5.41}$$

Then by (5.40) and (5.41), for each $f \in C_0^2(\mathbb{R}^d)$ and $n \in \mathbb{N}$, $g \in C_0(\mathbb{R}^{nd})$,

$$\int_{\mathcal{P}(\mathbb{C}_T)} |\Xi_f^g(\nu)| \Pi_\infty(\mathrm{d}\nu) = 0 \Rightarrow \Xi_f^g(\nu) = 0 \text{ for } \Pi_\infty\text{-a.s. } \nu \in \mathcal{P}(\mathbb{C}_T).$$

Since $C_0^2(\mathbb{R}^d)$ and $C_0(\mathbb{R}^{nd})$ are separable, one can find a common Π_∞-null set $\mathcal{N} \subset \mathcal{P}(\mathbb{C}_T)$ such that for all $\nu \notin \mathcal{N}$ and for all $0 \leqslant s < t \leqslant T$, $f \in C_0^2(\mathbb{R}^d)$ and $n \in \mathbb{N}$, $g \in C_0(\mathbb{R}^{nd})$,

$$\Xi_f^g(\nu) = \int_{\mathbb{C}_T} \big(M_{f,\nu}^{\sigma,b}(t, w) - M_{f,\nu}^{\sigma,b}(s, w)\big) g(w_{s_1}, \cdots, w_{s_n}) \nu(\mathrm{d}w) = 0.$$

Moreover, by (5.18) and (1.7), we also have

$$\Pi_\infty\{\nu \in \mathcal{P}(\mathbb{C}_T) : \nu_0 = \mu_0\} = 1.$$

Hence, for Π_∞-almost all ν,

$$\nu \in \mathcal{M}^{\sigma,b}_{\mu_0}.$$

Since $\mathcal{M}^{\sigma,b}_{\mu_0}$ only contains one point by uniqueness (see Theorem 4.1), all the points $\nu \notin \mathcal{N}$ are the same. Hence, Π_N weakly converges to a one-point measure. By [98, (ii) of Proposition 2.2], we conclude (5.19). Thus it remains to show (5.41).

(Step 2) Let $b_{\varepsilon,k}$ be defined by (5.31) and define

$$\Xi_{\varepsilon,k}(\nu) := \int_{\mathbb{C}_T} \big(M^{\sigma,b_{\varepsilon,k}}_{f,\nu}(t,w) - M^{\sigma,b_{\varepsilon,k}}_{f,\nu}(s,w)\big) g(w_{s_1},\cdots,w_{s_n})\nu(\mathrm{d}w).$$

By $b_{\varepsilon,k} \in L^\infty_T(C_b(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)))$, we have

$$\Xi_{\varepsilon,k} \in C_b(\mathcal{P}(\mathbb{C}_T)), \quad \forall \varepsilon > 0, k \in \mathbb{N}. \tag{5.42}$$

Indeed, note that

$$\begin{aligned}\Xi_{\varepsilon,k}(\nu) &= \int_{\mathbb{C}_T} \left(f(w_t) - f(w_s) + \frac{1}{2}\int_s^t \mathrm{tr}(\sigma\sigma^* \cdot \nabla^2 f)(r,w_r)\mathrm{d}r \right) g(w_{s_1},\cdots,w_{s_n})\nu(\mathrm{d}w)\\ &\quad + \int_{\mathbb{C}_T} \left(\int_s^t (b_{\varepsilon,k}\cdot\nabla f)(r,w_r,\nu_r)\mathrm{d}r \right) g(w_{s_1},\cdots,w_{s_n})\nu(\mathrm{d}w) =: \Xi^{(1)}_{\varepsilon,k}(\nu) + \Xi^{(2)}_{\varepsilon,k}(\nu).\end{aligned}$$

Since $f \in C^2_b$ and σ, g are bounded continuous, we have $\Xi^{(1)}_{\varepsilon,k} \in C_b(\mathcal{P}(\mathbb{C}_T))$. For $\Xi^{(2)}_{\varepsilon,k}$, since it is a non-linear functional of ν, we have to take some care for the continuity of $\nu \mapsto \Xi^{(2)}_{\varepsilon,k}(\nu)$. Suppose that $\nu_m \in \mathcal{P}(\mathbb{C}_T)$ weakly converges to $\nu \in \mathcal{P}(\mathbb{C}_T)$. By definition, we have

$$\begin{aligned}|\Xi^{(2)}_{\varepsilon,k}(\nu_m) - \Xi^{(2)}_{\varepsilon,k}(\nu)| &\leqslant \left| \int_{\mathbb{C}_T} \left(\int_s^t (b_{\varepsilon,k}\cdot\nabla f)(r,w_r,\nu_r)\mathrm{d}r \right) g(w_{s_1},\cdots,w_{s_n})(\nu_m-\nu)(\mathrm{d}w) \right|\\ &\quad + \kappa_0\|\nabla f\|_\infty\|g\|_\infty \int_{\mathbb{C}_T} \left(\int_s^t |\phi^k_r \otimes (\nu_m-\nu)_r|(w_r)\mathrm{d}r \right) \nu_m(\mathrm{d}w)\\ &=: I^{(1)}_m + I^{(2)}_m,\end{aligned}$$

where we have used that

$$|F_\varepsilon(r,x,s_1) - F_\varepsilon(r,x,s_2)| \leqslant \kappa_0|s_1 - s_2|.$$

For $I^{(1)}_m$, we clearly have

$$\lim_{m\to\infty} I^{(1)}_m = 0.$$

For $I^{(2)}_m$, by the dominated convergence theorem, it suffices to show that for each $r \in [s,t]$,

$$\lim_{m\to\infty} \int_{\mathbb{C}_T} |\phi^k_r \otimes (\nu_m-\nu)_r|(w_r)\nu_m(\mathrm{d}w) = 0,$$

which follows by noting that (see the proof of (5.36))

$$\lim_{m\to\infty} |\phi_r^k \otimes (\nu_m - \nu)_r|(x) = 0, \ x \in \mathbb{R}^d,$$

and

$$\lim_{|x-y|\to 0} \sup_m |(\phi_r^k \otimes \nu_{m,r})(x) - (\phi_r^k \otimes \nu_{m,r})(y)| = 0.$$

Thus we get (5.42), and so,

$$\lim_{N\to\infty} \mathbb{E}|\Xi_{\varepsilon,k}(\Pi_N)| = \int_{\mathcal{P}(\mathbb{C}_T)} |\Xi_{\varepsilon,k}(\nu)|\Pi_\infty(\mathrm{d}\nu).$$

On the other hand, we note that

$$\Xi_{\varepsilon,k}(\nu) - \Xi_f^g(\nu) = \int_{\mathbb{C}_T} \left(\int_s^t (b - b_{\varepsilon,k})(r, w_r, \nu_r) \cdot \nabla f(w_r) \mathrm{d}r \right) g(w_{s_1}, \cdots, w_{s_n}) \nu(\mathrm{d}w),$$

and

$$\Xi_{\varepsilon,k}(\Pi_N) - \Xi_f^g(\Pi_N) = \frac{1}{N} \sum_{i=1}^N \left(\int_s^t ((b - b_{\varepsilon,k}) \cdot \nabla f)\big(r, X_r^{N,i}, \eta_{\mathbf{X}_r^N}\big) \mathrm{d}r \right) g\big(X_{s_1}^{N,i}, \cdots, X_{s_n}^{N,i}\big).$$

By (5.32), we have

$$\begin{aligned} &\lim_{k\to\infty} \lim_{\varepsilon\to 0} \sup_N \mathbb{E}|\Xi_{\varepsilon,k}(\Pi_N) - \Xi_f^g(\Pi_N)| \\ &\quad \leqslant \|\nabla f\|_\infty \|g\|_\infty \lim_{k\to\infty} \lim_{\varepsilon\to 0} \sup_N \mathbb{E} \left(\int_s^t |b - b_{\varepsilon,k}|\big(r, X_r^{N,1}, \eta_{\mathbf{X}_r^N}\big) \mathrm{d}r \right) = 0, \end{aligned}$$

and by (5.37) and (5.38), as in showing (5.32),

$$\begin{aligned} &\lim_{k\to\infty} \lim_{\varepsilon\to 0} \int_{\mathcal{P}(\mathbb{C}_T)} |\Xi_{\varepsilon,k}(\nu) - \Xi_f^g(\nu)|\Pi_\infty(\mathrm{d}\nu) \\ &\leqslant \|\nabla f\|_\infty \|g\|_\infty \lim_{k\to\infty} \lim_{\varepsilon\to 0} \int_{\mathcal{P}(\mathbb{C}_T)} \int_{\mathbb{C}_T} \left(\int_0^T |b - b_{\varepsilon,k}|(s, w_s, \nu_s) \mathrm{d}s \right) \nu(\mathrm{d}w) \Pi_\infty(\mathrm{d}\nu) = 0. \end{aligned}$$

Thus we obtain (5.41) and the proof is complete. □

5.2.2 Entropy method

In this section we recall the entropy method used in [58] to show a quantitative result for weak convergence when the interaction kernel is bounded measurable, which is essentially contained in [58]. For the completeness of the paper, we provide a detailed proof. We first prepare the following lemma.

Lemma 5.9. *Let $\phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a bounded measurable function with $\phi(x,x) = 0$ and $\boldsymbol{\xi} := (\xi_1, \cdots, \xi_N)$ be a sequence of independent identical distributed random variables. Set*

$$\bar\phi(x,y) := \phi(x,y) - (\phi \circledast \mu)(x).$$

Then for any $\lambda \leqslant \frac{1}{16e^2\|\phi\|_\infty^2}$,

$$\mathbb{E}e^{\lambda N|(\bar\phi \circledast \eta_{\boldsymbol{\xi}})(\xi_1)|^2} \leqslant 6,$$

where $\eta_{\boldsymbol{\xi}}(dy) := \frac{1}{N}\sum_{i=1}^N \delta_{\xi_i}(dy)$.

Proof. Note that by Taylor's expansion,

$$\begin{aligned} e^{\lambda N|(\bar\phi \circledast \eta_{\boldsymbol{\xi}})(\xi_1)|^2} &= \sum_{m=0}^\infty \frac{\lambda^m N^m}{m!}|(\bar\phi \circledast \eta_{\boldsymbol{\xi}})(\xi_1)|^{2m} = \sum_{m=0}^\infty \frac{\lambda^m}{m!N^m}\Big|\sum_{j=1}^N \bar\phi(\xi_1,\xi_j)\Big|^{2m} \\ &\leqslant \sum_{m=0}^\infty \frac{\lambda^m}{m!N^m} 2^{2m}\Big(|\bar\phi(\xi_1,\xi_1)|^{2m} + \Big|\sum_{j=2}^N \bar\phi(\xi_1,\xi_j)\Big|^{2m}\Big) \\ &\leqslant \sum_{m=0}^\infty \frac{(4\lambda)^m}{m!N^m}\left(\|\bar\phi\|_\infty^{2m} + \sum_{j_1,\cdots,j_{2m}=2}^N \bar\phi(\xi_1,\xi_{j_1})\cdots\bar\phi(\xi_1,\xi_{j_{2m}})\right). \end{aligned}$$

Let $\mathbf{J}$ be the set of all indices $(j_1, \cdots, j_{2m}) \in \{2, \cdots, N\}^{2m}$ such that there is at least one index j_k different from all others. Since for $j \in \{2, \cdots, N\}$ and $x \in \mathbb{R}^d$,

$$\mathbb{E}\bar\phi(x,\xi_j) = 0,$$

by the independence of the components of $\boldsymbol{\xi}$, we have for any $(j_1, \cdots, j_{2m}) \in \mathbf{J}$,

$$\mathbb{E}\Big[\bar\phi(\xi_1,\xi_{j_1})\cdots\bar\phi(\xi_1,\xi_{j_{2m}})\Big] = \mathbb{E}\Big[\mathbb{E}\big[\bar\phi(x,\xi_{j_1})\cdots\bar\phi(x,\xi_{j_{2m}})\big]|_{x=\xi_1}\Big] = 0.$$

Hence,

$$\mathbb{E}e^{\lambda N|(\bar\phi \circledast \eta_{\boldsymbol{\xi}})(\xi_1)|^2} \leqslant \sum_{m=0}^\infty \frac{(4\lambda)^m}{m!N^m}\|\bar\phi\|_\infty^{2m}(1 + \sharp\mathbf{J}^c),$$

where $\sharp\mathbf{J}^c$ stands for the cardinality of the complement set $\mathbf{J}^c$.

Suppose $2m \leqslant N$. It is easy to see that $(j_1, \cdots, j_{2m}) \in \mathbf{J}^c$ if and only if each j_k appears at least twice and there are at most m-distinct j_k. Thus one has

$$\mathbf{J}^c = \cup_{n=1}^m \mathbf{J}_n,$$

where $\mathbf{J}_n$ is the set of $(j_1, \cdots, j_{2m})$ such that each j_k appears at least twice and exactly n-integers appear. Clearly, by Stirling's formula $n^n \leqslant e^n n! \leqslant e^{2n}n^n$, we have

$$\sharp\mathbf{J}_n \leqslant \binom{N-1}{n} n^{2m} = \frac{(N-1)^n}{n!}n^{2m} \leqslant \frac{e^n(N-1)^n}{n^n}n^{2m} \leqslant (Ne)^n n^m.$$

Thus, for $2m \leqslant N$,

$$\sharp \mathbf{J}^c \leqslant \sum_{n=1}^{m} (N\mathrm{e})^n n^m \leqslant 2(N\mathrm{e})^m m^m \leqslant 2(N\mathrm{e})^m \mathrm{e}^m m!.$$

Moreover, for $2m > N$, we obviously have

$$\sharp \mathbf{J}^c \leqslant N^{2m} \leqslant N^m (2m)^m \leqslant N^m (2\mathrm{e})^m m!.$$

So, for $\lambda \leqslant \frac{1}{16\mathrm{e}^2\|\phi\|_\infty^2}$,

$$\mathbb{E}\mathrm{e}^{\lambda N|(\bar\phi \circledast \eta_{\boldsymbol{\xi}})(\xi_1)|^2} \leqslant \sum_{m=0}^{\infty} (4\lambda)^m \|\bar\phi\|_\infty^{2m} \Big(\frac{1}{m! N^m} + (2\mathrm{e})^m\Big) \leqslant 2\sum_{m=0}^{\infty} 2^{-m} = 6.$$

The proof is complete. □

Now we can use the entropy formula in Lemma 3.26 to show the following result.

Theorem 5.10. *Suppose that* $(\mathbf{H}^\sigma)$ *and* $(\mathbf{H}^b)$ *hold and* ϕ *is bounded measurable. Let* μ_t^N *be the law of* $\mathbf{X}_t^N$ *in* $\mathbb{R}^{dN}$ *and* μ_t *be the law of* X_t *in* $\mathbb{R}^d$*. Then there is a constant* $C = C(\kappa_0, \kappa_1) > 0$ *independent of* ϕ *such that for any* $t > 0$,

$$\mathcal{H}\big(\mu_t^N|\mu_t^{\otimes N}\big) \leqslant \mathrm{e}^{C\|\phi\|_\infty^2 t}\Big(\mathcal{H}\big(\mu_0^N|\mu_0^{\otimes N}\big) + C\|\phi\|_\infty^2 t\Big).$$

Proof. Let $\eta_{\boldsymbol{w}_s} := \frac{1}{N}\sum_{i=1}^N \delta_{w_s^i}$ and $B, \boldsymbol{\sigma}$ be defined by (5.15) and (5.16), respectively. By Lemma 3.26 and (5.6), we have

$$\begin{aligned}
\mathcal{H}\big(\mu_t^N|\mu_t^{\otimes N}\big) &\leqslant \mathcal{H}\big(\mu_0^N|\mu_0^{\otimes N}\big) + \frac{1}{2}\int_0^t \mathbb{E}^{\mu_s^N}|\boldsymbol{\sigma}(s,\boldsymbol{w}_s)^{-1}(B(s,\boldsymbol{w}_s,\mu_s) - B(s,\boldsymbol{w}_s,\eta_{\boldsymbol{w}_s}))|^2 \mathrm{d}s \\
&\leqslant \mathcal{H}\big(\mu_0^N|\mu_0^{\otimes N}\big) + \frac{\kappa_0}{2}\int_0^t \mathbb{E}^{\mu_s^N}|B(s,\boldsymbol{w}_s,\mu_s) - B(s,\boldsymbol{w}_s,\eta_{\boldsymbol{w}_s})|^2 \mathrm{d}s \\
&\leqslant \mathcal{H}\big(\mu_0^N|\mu_0^{\otimes N}\big) + \frac{\kappa_0\kappa_1}{2}\sum_{i=1}^N \int_0^t \mathbb{E}^{\mu_s^N}|(\phi_s \circledast \mu_s)(w_s^i) - (\phi_s \circledast \eta_{\boldsymbol{w}_s})(w_s^i)|^2 \mathrm{d}s \\
&= \mathcal{H}\big(\mu_0^N|\mu_0^{\otimes N}\big) + \frac{\kappa_0\kappa_1}{2}\int_0^t N\mathbb{E}^{\mu_s^N}|(\bar\phi_s \circledast \eta_{\boldsymbol{w}_s})(w_s^1)|^2 \mathrm{d}s.
\end{aligned}$$

Now by the variational representation (2.60) and Lemma 5.9 with $\lambda = \frac{1}{16\mathrm{e}^2\|\phi\|_\infty^2}$, we further have

$$\begin{aligned}
\mathcal{H}\big(\mu_t^N|\mu_t^{\otimes N}\big) &\leqslant \mathcal{H}\big(\mu_0^N|\mu_0^{\otimes N}\big) + \frac{\kappa_0\kappa_1}{2\lambda}\int_0^t \Big[\mathcal{H}\big(\mu_s^N|\mu_s^{\otimes N}\big) + \log \mathbb{E}^{\mu_s^{\otimes N}} \mathrm{e}^{\lambda N|(\bar\phi_s \circledast \eta_{\boldsymbol{w}_s})(w_s^1)|^2}\Big]\mathrm{d}s \\
&\leqslant \mathcal{H}\big(\mu_0^N|\mu_0^{\otimes N}\big) + C\|\phi\|_\infty^2 \int_0^t \Big[\mathcal{H}\big(\mu_s^N|\mu_s^{\otimes N}\big) + \log 6\Big]\mathrm{d}s,
\end{aligned}$$

which yields the desired estimate by Gronwall's inequality. □

Remark 5.11. By the Pinsker inequalities (2.58) and (2.61), we have for any $k \leqslant N$,

$$\|\mu_t^{N,k} - \mu_t^{\otimes k}\|_{\mathrm{var}} \leqslant \sqrt{2\mathcal{H}\big(\mu_t^{N,k}|\mu_t^{\otimes k}\big)} \leqslant \sqrt{\frac{\mathrm{e}^{C\|\phi\|_\infty^2 t}k}{N}}\Big(\mathcal{H}\big(\mu_0^N|\mu_0^{\otimes N}\big) + C\|\phi\|_\infty^2 t\Big).$$

Note that when $F(t,x,r) = r$ is linear and $\mathcal{H}\big(\mu_t^{N,k}|\mu_t^{\otimes k}\big) \leqslant C_0 k^2/N^2$, by a delicate analysis of the BBGKY hierarchy, the following sharp estimate is obtained by Lacker (see Theorem 2.10 of [70]):

$$\|\mu_t^{N,k} - \mu_t^{\otimes k}\|_{\mathrm{var}} \leqslant \sqrt{2\mathcal{H}\big(\mu_t^{N,k}|\mu_t^{\otimes k}\big)} \leqslant Ck/N.$$

5.3 From weak convergence to strong convergence : Proof of Theorem 5.1

In this section we show how to use the previous weak convergence result to derive the strong convergence of the particle system based on the Zvonkin's transformation. The following lemma is the key point.

Lemma 5.12. *Let $\phi : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Set*

$$\bar\phi_t(x,y) := \phi_t(x,y) - (\phi_t \circledast \mu_{X_t})(x).$$

(i) If ϕ is bounded measurable, then there is a constant $C = C(\kappa_0, \kappa_1) > 0$ such that for all $t > 0$,

$$\mathbb{E}|(\bar\phi_t \circledast \eta_{\mathbf{X}_t^N})(X_t^{N,1})|^2 \leqslant C\|\phi\|_\infty^2 \mathrm{e}^{C\|\phi\|_\infty^2 t}\Big(\mathcal{H}\big(\mu_0^N|\mu_0^{\otimes N}\big) + 1\Big)/N. \tag{5.43}$$

(ii) If ϕ satisfies (5.7), then for any $T > 0$,

$$\lim_{N\to\infty} \mathbb{E}\left(\int_0^T |(\bar\phi_t \circledast \eta_{\mathbf{X}_t^N})(X_t^{N,1})|^2 \mathrm{d}t\right) = 0. \tag{5.44}$$

Proof. (i) By the variational representation (2.60), for any $\varepsilon > 0$, we have

$$\varepsilon N \mathbb{E}|(\bar\phi_t \circledast \eta_{\mathbf{X}_t^N})(X_t^{N,1})|^2 = \varepsilon N \mathbb{E}^{\mu_t^N}|\bar\phi_t(w_t^1, \eta_{\boldsymbol{w}_t})|^2 \leqslant \mathcal{H}(\mu_t^N|\mu_t^{\otimes N}) + \log \mathbb{E}^{\mu_t^{\otimes N}} \mathrm{e}^{\varepsilon N|\bar\phi_t(w_t^1, \eta_{\boldsymbol{w}_t})|^2},$$

which in turn implies (5.43) by Lemma 5.9 with $\varepsilon = \frac{1}{16\mathrm{e}^2\|\phi\|_\infty^2}$ and Theorem 5.10.

(ii) By definition we have

$$\mathbb{E}\left(\int_0^T |(\bar\phi_t \circledast \eta_{\mathbf{X}_t^N})(X_t^{N,1})|^2 \mathrm{d}t\right) = \frac{1}{N^2}\sum_{j,k=1}^N \mathbb{E}\left(\int_0^T \Gamma_t\big(X_t^{N,1}, X_t^{N,j}, X_t^{N,k}\big)\mathrm{d}t\right), \tag{5.45}$$

where

$$\Gamma_t(x,y,z) := \bar{\phi}_t(x,y)\bar{\phi}_t(x,z).$$

Let $\phi^\varepsilon_t(x,y) := (\phi_t * \varGamma_\varepsilon)(x,y)$ be the mollifying approximation of ϕ_t and

$$\bar{\phi}^\varepsilon_t(x,y) := \phi^\varepsilon_t(x,y) - (\phi^\varepsilon_t \circledast \mu_{X_t})(x),$$

and

$$\Gamma^\varepsilon_t(x,y,z) := \bar{\phi}^\varepsilon_t(x,y)\bar{\phi}^\varepsilon_t(x,z).$$

Noting that

$$(\Gamma_t - \Gamma^\varepsilon_t)(x,y,z) = (\bar{\phi}_t - \bar{\phi}^\varepsilon_t)(x,y)\bar{\phi}^\varepsilon_t(x,z) + \bar{\phi}_t(x,y)(\bar{\phi}_t - \bar{\phi}^\varepsilon_t)(x,z),$$

by Hölder's inequality, we have

$$\begin{aligned} I^N_{j,k}(\varepsilon) &:= \left|\mathbb{E}\left(\int_0^T (\Gamma_t - \Gamma^\varepsilon_t)\big(X^{N,1}_t, X^{N,j}_t, X^{N,k}_t\big)\mathrm{d}t\right)\right| \\ &\leqslant \left(\mathbb{E}\int_0^T (\bar{\phi}_t - \bar{\phi}^\varepsilon_t)^2\big(X^{N,1}_t, X^{N,j}_t\big)\mathrm{d}t\right)^{1/2}\left(\mathbb{E}\int_0^T \bar{\phi}^\varepsilon_t\big(X^{N,1}_t, X^{N,k}_t\big)^2\mathrm{d}t\right)^{1/2} \\ &\quad + \left(\mathbb{E}\int_0^T \bar{\phi}_t\big(X^{N,1}_t, X^{N,j}_t\big)^2\mathrm{d}t\right)^{1/2}\left(\mathbb{E}\int_0^T (\bar{\phi}_t - \bar{\phi}^\varepsilon_t)^2\big(X^{N,1}_t, X^{N,k}_t\big)\mathrm{d}t\right)^{1/2}. \end{aligned}$$

Using the Krylov estimate (5.30) and as in showing (5.34), we get

$$\lim_{\varepsilon\to 0}\sup_N \sup_{j,k} I^N_{j,k}(\varepsilon) = 0. \tag{5.46}$$

On the other hand, for fixed ε, by (5.19) we have

$$\begin{aligned} &\lim_{N\to\infty}\sup_{j\neq k\neq 1}\mathbb{E}\left(\int_0^T \Gamma^\varepsilon_t\big(X^{N,1}_t, X^{N,j}_t, X^{N,k}_t\big)\mathrm{d}t\right) \\ &= \lim_{N\to\infty}\mathbb{E}\left(\int_0^T \Gamma^\varepsilon_t\big(X^{N,1}_t, X^{N,2}_t, X^{N,3}_t\big)\mathrm{d}t\right) \\ &= \mathbb{E}\left(\int_0^T \Gamma^\varepsilon_t\big(X^1_t, X^2_t, X^3_t\big)\mathrm{d}t\right) = 0, \end{aligned} \tag{5.47}$$

where the last step is due to the fact that

$$\mathbb{E}\Gamma^\varepsilon_t\big(X^1_t, X^2_t, X^3_t\big) = \mathbb{E}\left[\mathbb{E}\bar{\phi}^\varepsilon_t(x, X^2_t)\mathbb{E}\bar{\phi}^\varepsilon_t(x, X^3_t); x = X^1_t\right] = 0.$$

Thus by (5.46) and (5.47),

$$\lim_{N\to\infty}\sup_{j\neq k\neq 1}\mathbb{E}\left(\int_0^T \Gamma_t(X^{N,1}_t, X^{N,j}_t, X^{N,k}_t)\mathrm{d}t\right) = 0. \tag{5.48}$$

Moreover, by the Krylov estimate (5.30) we also have

$$\begin{aligned}
&\sup_{j,k} \mathbb{E}\left(\int_0^T \Gamma_t\big(X_t^{N,1}, X_t^{N,j}, X_t^{N,k}\big)\mathrm{d}t\right) \\
&\leqslant \sup_{j,k} \mathbb{E}\left(\int_0^T \bar{\phi}_t\big(X_t^{N,1}, X_t^{N,j}\big)^2\mathrm{d}t\right)^{\frac{1}{2}} \mathbb{E}\left(\int_0^T \bar{\phi}_t\big(X_t^{N,1}, X_t^{N,k}\big)^2\mathrm{d}t\right)^{\frac{1}{2}} \\
&= \sup_{j} \mathbb{E}\left(\int_0^T \bar{\phi}_t\big(X_t^{N,1}, X_t^{N,j}\big)^2\mathrm{d}t\right) < \infty. \qquad (5.49)
\end{aligned}$$

By (5.45), (5.48) and (5.49), we obtain (5.44). □

Now we can give the

Proof of Theorem 5.1. Let X_t be the unique strong solution of dDDSDE (5.2) starting from X_0 (see Theorem 4.1). Define

$$\bar{b}(t,x) := b(t,x,\mu_{X_t}) = F(t,x,(\phi_t \circledast \mu_{X_t})(x)).$$

By $(\mathbf{H}^b)$, it is easy to see that

$$\flat := \|\bar{b}\|_{\mathbb{L}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} < \infty.$$

Consider the following backward PDE

$$\partial_t \mathbf{u} + \tfrac{1}{2}\mathrm{tr}(\sigma\sigma^* \cdot \nabla^2 \mathbf{u}) + \bar{b}\cdot\nabla\mathbf{u} - \lambda\mathbf{u} + \bar{b} = 0, \quad \mathbf{u}(T) = 0.$$

By reversing the time variable and Theorem 2.19, there is a unique solution $\mathbf{u}$ satisfying the following estimate: for any $\beta \in (0,\vartheta)$, where $\vartheta := 1 - |\frac{1}{\boldsymbol{p}}| - \frac{2}{q}$, there is a constant $C_0 = C_0(T,\kappa_0,d,\boldsymbol{p},q,\beta) \geqslant 1$ such that for all $\lambda \geqslant C_0\flat^{2/\vartheta}$,

$$\lambda^{\frac{1}{2}(\vartheta-\beta)}\|\mathbf{u}\|_{\mathbb{L}^\infty_T(\mathcal{C}^{1+\beta})} + \|\nabla^2\mathbf{u}\|_{\widetilde{\mathbb{L}}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \leqslant C_0\flat. \qquad (5.50)$$

In particular, one can choose $\lambda = (2C_0\flat)^{2/\vartheta}$ so that

$$\|\nabla\mathbf{u}\|_{\mathbb{L}^\infty_T} \leqslant \tfrac{1}{2}. \qquad (5.51)$$

Now if we define

$$\Phi(t,x) := x + \mathbf{u}(t,x),$$

then for each t,

$$x \mapsto \Phi(t,x) \text{ is a } C^1\text{-diffeomorphism on } \mathbb{R}^d,$$

and

$$\|\nabla\Phi\|_{\mathbb{L}^\infty_T} + \|\nabla\Phi^{-1}\|_{\mathbb{L}^\infty_T} \leqslant 2. \qquad (5.52)$$

Define

$$Y_t := \Phi(t, X_t), \quad Y_t^{N,1} := \Phi(t, X_t^{N,1}).$$

By Itô's formula (see the proof in Lemma 3.12), we have

$$\mathrm{d}Y_t = \lambda \mathbf{u}(t, X_t)\mathrm{d}t + \tilde{\sigma}(t, X_t)\mathrm{d}W_t^1$$

and

$$\mathrm{d}Y_t^{N,1} = \lambda \mathbf{u}(t, X_t^{N,1})\mathrm{d}t + \big(B \cdot \nabla \mathbf{u}\big)\big(t, X_t^{N,1}\big)\mathrm{d}t + \tilde{\sigma}(t, X_t^{N,1})\mathrm{d}W_t^1,$$

where $\tilde{\sigma} := \sigma^* \nabla \Phi$ and

$$B(t,x) := b(t, x, \eta_{\mathbf{X}_t^N}) - b(t, x, \mu_{X_t}).$$

In particular, we have

$$\begin{aligned} Y_t^{N,1} - Y_t = \Phi(0, X_0^{N,1}) - \Phi(0, X_0) + \lambda \int_0^t \Big[\mathbf{u}(s, X_s^{N,1}) - \mathbf{u}(s, X_s)\Big]\mathrm{d}s \\ + \int_0^t \big(B \cdot \nabla \mathbf{u}\big)\big(s, X_s^{N,1}\big)\mathrm{d}s + \int_0^t \Big[\tilde{\sigma}(s, X_s^{N,1}) - \tilde{\sigma}(s, X_s)\Big]\mathrm{d}W_s^1. \end{aligned}$$

By Itô's formula and (5.51), (5.52), we further have

$$\begin{aligned} |Y_t^{N,1} - Y_t|^2 \leqslant 4|X_0^{N,1} - X_0|^2 + \int_0^t |Y_s^{N,1} - Y_s|\big(\lambda |X_s^{N,1} - X_s| + |B\big(s, X_s^{N,1}\big)|\big)\mathrm{d}s \\ + \int_0^t |\tilde{\sigma}(s, X_s^{N,1}) - \tilde{\sigma}(s, X_s)|^2 \mathrm{d}s + M_t, \end{aligned} \tag{5.53}$$

where M_t is a continuous local martingale. Note that by (2.15),

$$|\tilde{\sigma}(s, X_s^{N,1}) - \tilde{\sigma}(s, X_s)|^2 \leqslant 2\ell_{N,0}(s)|X_s^{N,1} - X_s|^2,$$

where

$$\ell_{N,\lambda}(s) := \mathcal{M}|\nabla \tilde{\sigma}(s, \cdot)|^2(X_s^{N,1}) + \mathcal{M}|\nabla \tilde{\sigma}(s, \cdot)|^2(X_s) + \|\tilde{\sigma}\|_\infty^2 + \lambda + 1.$$

Thus, by (5.53) and (5.52) we have

$$\begin{aligned} |X_t^{N,1} - X_t|^2 \leqslant C\bigg(|X_0^{N,1} - X_0|^2 + \int_0^t \ell_{N,\lambda}(s)|X_s^{N,1} - X_s|^2 \mathrm{d}s \\ + \int_0^t |B\big(s, X_s^{N,1}\big)|^2 \mathrm{d}s\bigg) + M_t, \end{aligned} \tag{5.54}$$

where $C > 0$ is an absolute constant. By the chain rule, we have

$$\mathcal{M}|\nabla \tilde{\sigma}|^2 \leqslant 4\mathcal{M}|\nabla \sigma|^2 + \|\sigma\|_\infty^2 \mathcal{M}|\nabla^2 \mathbf{u}|^2.$$

By (2.16) and (3.2), we have

$$\|\mathcal{M}|\nabla\sigma|^2\|_{\mathbb{L}^{q_0/2}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_0/2}_{\boldsymbol{\pi}})} \lesssim \||\nabla\sigma|^2\|_{\mathbb{L}^{q_0/2}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_0/2}_{\boldsymbol{\pi}})} = \|\nabla\sigma\|^2_{\mathbb{L}^{q_0}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}_0}_{\boldsymbol{\pi}})} \leqslant \kappa_0,$$

and by (5.50),

$$\|\mathcal{M}|\nabla^2\mathbf{u}|^2\|_{\mathbb{L}^{q/2}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}/2}_{\boldsymbol{\pi}})} \lesssim \||\nabla^2\mathbf{u}|^2\|_{\mathbb{L}^{q/2}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}/2}_{\boldsymbol{\pi}})} = \|\nabla^2\mathbf{u}\|^2_{\mathbb{L}^{q}_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \leqslant (C_0\flat)^2.$$

Since $(\frac{q_0}{2}, \frac{\boldsymbol{p}_0}{2}), (\frac{q}{2}, \frac{\boldsymbol{p}}{2}) \in \mathscr{I}_2$, by (5.29) and (3.46) we have for any $\gamma > 0$,

$$A_\gamma := \sup_N \mathbb{E}\exp\left\{\gamma\int_0^T \ell_{N,\lambda}(s)\mathrm{d}s\right\} < \infty.$$

Thus by (5.54) and Lemma A.5, we get for any $\gamma \in (0,1)$,

$$\mathbb{E}\left(\sup_{t\in[0,T]}|X^{N,1}_t - X_t|^{2\gamma}\right) \leqslant C_\gamma A_{\frac{\gamma+1}{\gamma-1}}\left(\mathbb{E}|X^{N,1}_0 - X_0|^2 + \mathbb{E}\int_0^T |B\big(s, X^{N,1}_s\big)|^2\mathrm{d}s\right)^\gamma. \quad (5.55)$$

Noting that by (5.6),

$$|B(t,x)| \leqslant \kappa_1|(\phi_t \circledast \eta_{\mathbf{X}^N_t})(x) - (\phi_t \circledast \mu_{X_t})(x)| = \kappa_1|(\bar\phi_t \circledast \eta_{\mathbf{X}^N_t})(x)|,$$

where

$$\bar\phi_t(x,y) := \phi_t(x,y) - (\phi_t \circledast \mu_{X_t})(x),$$

we further have for any $\gamma \in (0,1)$,

$$\mathbb{E}\left(\sup_{t\in[0,T]}|X^{N,1}_t - X_t|^{2\gamma}\right) \leqslant C_\gamma A_{\frac{\gamma+1}{\gamma-1}}\left(\mathbb{E}|X^{N,1}_0 - X_0|^2 + \kappa_1^2\mathbb{E}\int_0^T |(\bar\phi_s \circledast \eta_{\mathbf{X}^N_s})(X^{N,1}_s)|^2\mathrm{d}s\right)^\gamma.$$

Now, (i) follows by (5.44) and the above estimate.

(ii) When h and ϕ are bounded, by (5.6) one has

$$|\bar b(t,x)| \leqslant \|h\|_\infty + \kappa_1\|\phi\|_\infty.$$

Thus for any $\delta > 2$, one can choose $q, \boldsymbol{p}$ in (5.50) close to ∞ so that $\vartheta = \frac{2}{\delta} = 1 - \frac{2}{q} - |\frac{1}{\boldsymbol{p}}|$ and

$$\flat := \|\bar b\|_{\mathbb{L}^q_T(\widetilde{\mathbb{L}}^{\boldsymbol{p}}_{\boldsymbol{\pi}})} \leqslant C(1 + \|\phi\|_\infty).$$

By (5.29), (3.47) and for $\lambda = (2C_0\flat)^{2/\vartheta}$, we have

$$A_\gamma = \sup_N \mathbb{E}\exp\left\{\gamma\int_0^T \ell_{N,\lambda}(s)\mathrm{d}s\right\} \leqslant C\mathrm{e}^{C\flat^{2/\vartheta}} \leqslant C\mathrm{e}^{C\|\phi\|_\infty^{2/\vartheta}}.$$

Estimate (5.10) now follows by the above estimates and (5.43). □

5.4 Moderately interacting particle systems: Proof of Theorem 5.3

We consider the following McKean-Vlasov type approximation for density-dependent SDE (5.4):

$$\mathrm{d}X_t^\varepsilon = F(t, X_t^\varepsilon, (\phi_\varepsilon * \rho_t^\varepsilon)(X_t^\varepsilon))\mathrm{d}t + \sigma(t, X_t^\varepsilon)\mathrm{d}W_t^1, \ X_0^\varepsilon = X_0,$$

where $\phi_\varepsilon(x) = \varepsilon^{-d}\phi(x/\varepsilon)$, and ϕ is a bounded probability density function with support in the unit ball, F is bounded measurable and ρ_t^ε is the density of X_t^ε.

We first show the following lemma.

Lemma 5.13. *For any $T > 0$, $\beta \in (0, \gamma_0)$ and $\gamma \in (0, 1)$, there is a constant $C = C(T, \beta, \gamma, \Theta) > 0$ such that for all $\varepsilon \in (0, 1)$,*

$$\mathbb{E}\left(\sup_{t\in[0,T]} |X_t^\varepsilon - X_t|^{2\gamma}\right) \leqslant C\varepsilon^{2\beta\gamma}.$$

Proof. Let X_t be the unique strong solution of DDSDE (5.4) starting from X_0. Define

$$\bar{b}(t, x) := F(t, x, \rho_t(x)).$$

By assumption we have

$$\|\bar{b}\|_{\mathbb{L}_T^\infty} \leqslant \|F\|_{\mathbb{L}_T^\infty}.$$

Consider the following backward PDE

$$\partial_t \mathbf{u} + \tfrac{1}{2}\mathrm{tr}(\sigma\sigma^* \cdot \nabla^2 \mathbf{u}) + \bar{b} \cdot \nabla \mathbf{u} - \lambda \mathbf{u} + \bar{b} = 0, \quad \mathbf{u}(T) = 0.$$

As in the proof of Theorem 5.1 we construct a C^1-diffeomorphism

$$\Phi(t, x) := x + \mathbf{u}(t, x),$$

and define

$$Y_t^\varepsilon := \Phi(t, X_t^\varepsilon), \quad Y_t := \Phi(t, X_t).$$

By the generalized Itô formula, we have

$$\mathrm{d}Y_t = \lambda \mathbf{u}(t, X_t)\mathrm{d}t + \tilde{\sigma}(t, X_t)\mathrm{d}W_t^1$$

and

$$\mathrm{d}Y_t^\varepsilon = \lambda \mathbf{u}(t, X_t^\varepsilon)\mathrm{d}t + \big(B_\varepsilon \cdot \nabla \mathbf{u}\big)\big(t, X_t^\varepsilon\big)\mathrm{d}t + \tilde{\sigma}(t, X_t^\varepsilon)\mathrm{d}W_t^1,$$

where $\tilde{\sigma} = \sigma^* \nabla \Phi$ and

$$B_\varepsilon(t, x) := F(t, x, (\phi_\varepsilon * \rho_t^\varepsilon)(x)) - F(t, x, \rho_t(x)).$$

In particular, we have

$$Y_t^\varepsilon - Y_t = \lambda \int_0^t \Big[\mathbf{u}(s, X_s^\varepsilon) - \mathbf{u}(s, X_s)\Big] \mathrm{d}s + \int_0^t \big(B_\varepsilon \cdot \nabla \mathbf{u}\big)\big(s, X_s^\varepsilon\big)\mathrm{d}s$$
$$+ \int_0^t \Big[\tilde\sigma(s, X_s^\varepsilon) - \tilde\sigma(s, X_s)\Big]\mathrm{d}W_s^1.$$

By Itô's formula and (5.51), we further have

$$\begin{aligned} |Y_t^\varepsilon - Y_t|^2 \leqslant & \int_0^t |Y_s^\varepsilon - Y_s|\big(\lambda |X_s^\varepsilon - X_s| + |B_\varepsilon\big(s, X_s^\varepsilon\big)|\big)\mathrm{d}s \\ & + \int_0^t |\tilde\sigma(s, X_s^\varepsilon) - \tilde\sigma(s, X_s)|^2 \mathrm{d}s + M_t, \end{aligned} \tag{5.56}$$

where M_t is a continuous local martingale. Completely the same way as in proving (5.55), we have

$$\mathbb{E}|X_t^{N,1} - X_t|^{2\gamma} \lesssim \left(\mathbb{E}\int_0^T |B_\varepsilon\big(s, X_s^\varepsilon\big)|^2 \mathrm{d}s\right)^\gamma. \tag{5.57}$$

On the other hand, for any $p > d$, by Lemma 4.4 we have

$$\|\rho_t^\varepsilon - \rho_t\|_{\mathbb{L}^\infty} \lesssim_C \int_0^t (t-s)^{-\frac{1}{2}(1+\frac{d}{p})} \|B_\varepsilon(s)\|_{\widetilde{\mathbb{L}}^p} \mathrm{d}s.$$

By the Lipschitz assumption on F in r, we have

$$\|B_\varepsilon(s)\|_{\widetilde{\mathbb{L}}^p} \leqslant \|B_\varepsilon(s)\|_{\mathbb{L}^\infty} \lesssim \|\phi_\varepsilon * \rho_s^\varepsilon - \rho_s\|_{\mathbb{L}^\infty} \leqslant \|\rho_s^\varepsilon - \rho_s\|_{\mathbb{L}^\infty} + \|\phi_\varepsilon * \rho_s - \rho_s\|_{\mathbb{L}^\infty}.$$

For any $\beta \in (0, \gamma_0)$, noting that by (3.42),

$$\|\rho_s(\cdot + y) - \rho_s\|_{\mathbb{L}^\infty} \leqslant C\|\rho_0\|_\infty |y|^\beta s^{-\beta/2},$$

we have

$$\begin{aligned} \|\phi_\varepsilon * \rho_s - \rho_s\|_{\mathbb{L}^\infty} & \leqslant \int_{\mathbb{R}^d} \|\rho_s(\cdot + y) - \rho_s\|_{\mathbb{L}^\infty} \cdot |\phi_\varepsilon(y)|\mathrm{d}y \\ & \lesssim s^{-\beta/2} \int_{\mathbb{R}^d} |y|^\beta \cdot |\phi_\varepsilon(y)|\mathrm{d}y \lesssim s^{-\beta/2}\varepsilon^\beta. \end{aligned} \tag{5.58}$$

Hence,

$$\|\rho_t^\varepsilon - \rho_t\|_{\mathbb{L}^\infty} \lesssim_C \int_0^t (t-s)^{-\frac{1}{2}(1+\frac{d}{p})} (\|\rho_s^\varepsilon - \rho_s\|_{\mathbb{L}^\infty} + s^{-\frac{\beta}{2}}\varepsilon^\beta)\mathrm{d}s.$$

By Lemma A.4, we have

$$\|\rho_t^\varepsilon - \rho_t\|_{\mathbb{L}^\infty} \leqslant Ct^{\frac{1}{2}-\frac{d}{2p}-\frac{\beta}{2}}\varepsilon^\beta \leqslant Ct^{-\frac{\beta}{2}}\varepsilon^\beta. \tag{5.59}$$

Note that by (5.6), (5.58) and (5.59),

$$\begin{aligned}\mathbb{E}|B_\varepsilon\big(s, X_s^\varepsilon\big)|^2 &\leqslant \kappa_1^2 \int_{\mathbb{R}^d} |\phi_\varepsilon * \rho_s^\varepsilon(x) - \rho_s(x)|^2 \rho_s^\varepsilon(x)\mathrm{d}x \\ &\leqslant \kappa_1^2 \|\phi_\varepsilon * \rho_s^\varepsilon - \rho_s\|_{\mathbb{L}^\infty}^2 \leqslant Cs^{-\beta}\varepsilon^{2\beta}.\end{aligned}$$

Substituting this into (5.57), we obtain the desired estimate. □

Now we can give the

Proof of Theorem 5.3. This is a direct combination of Lemma 5.13 and (ii) of Theorem 5.1. □

Chapter 6

Strong and weak convergence rate of averaging principle for McKean-Vlasov SDEs with localized L^p drift

In this chapter, we consider the averaging principle of the following DDSDE with highly oscillating time component

$$\mathrm{d}X_t^\varepsilon = b\left(\frac{t}{\varepsilon}, X_t^\varepsilon, \mu_t^\varepsilon\right)\mathrm{d}t + \sigma(X_t^\varepsilon)\mathrm{d}W_t, \quad X_0^\varepsilon = \xi \in \mathscr{F}_0, \tag{6.1}$$

where $b : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ are measurable functions, $\mu_t^\varepsilon := \mathcal{L}(X_t^\varepsilon)$ is the time marginal law of X_t^ε and the time scale $0 < \varepsilon \ll 1$.

Throughout this chapter we need the following conditions.

($\mathbf{H}_b^1$) Let $p_0 \in (d \vee 2, \infty)$ and assume that there is a nonnegative constant κ_0 such that for all $t \geqslant 0$ and $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$

$$|\!|\!| b(t,\cdot,\mu) |\!|\!|_{p_0} + \frac{|\!|\!| b(t,\cdot,\mu) - b(t,\cdot,\nu) |\!|\!|_{p_0}}{\|\mu - \nu\|_{var}} \leqslant \kappa_0,$$

where $\|\mu - \nu\|_{var} := \sup_{A \in \mathscr{B}(\mathbb{R}^d)} |\mu(A) - \nu(A)|$ is total variation.

($\mathbf{H}_b^2$) There are functions $\bar{b} : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$, $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ and $H : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}_+$ such that for all $(T, t, x, \mu) \in \mathbb{R}_+^2 \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$

$$\left| \frac{1}{T} \int_t^{T+t} (b(s,x,\mu) - \bar{b}(x,\mu))\mathrm{d}s \right| \leqslant \omega(T) H(x,\mu), \tag{6.2}$$

where $\lim_{t\to\infty} \omega(t) = 0$ and $\sup_\mu |\!|\!| H(\cdot,\mu) |\!|\!|_{p_0} < \kappa_0$. Here p_0 and κ_0 are as in **($\mathbf{H}_b^1$)**.

$(\mathbf{H}^\sigma)$ There are constants $p > d \vee 2$, $\kappa_1 > 1$ and $\beta \in (0,1)$ such that for all $x, y, \xi \in \mathbb{R}^d$,

$$\kappa_1^{-1}|\xi| \leqslant |\sigma(x)\xi| \leqslant \kappa_1|\xi|, \quad \|\nabla\sigma\|_p \leqslant \kappa_1,$$

and

$$\|\sigma(x) - \sigma(y)\|_{HS} \leqslant \kappa_1|x-y|^\beta,$$

where $\|\cdot\|_{HS}$ is the Hilbert-Schmidt norm.

Remark 6.1. We Note that

$$\begin{aligned}\|\bar{b}(\cdot,\mu)\|_{p_0} &\leqslant \frac{1}{T}\int_0^T \|b(s,\cdot,\mu)\|_{p_0}\mathrm{d}s + \|\frac{1}{T}\int_0^T \big(b(s,\cdot,\mu) - \bar{b}(\cdot,\mu)\mathrm{d}s\big)\|_{p_0} \\ &\leqslant \kappa_0 + \omega(T)\|H(\cdot,\mu)\|_{p_0},\end{aligned}$$

provided conditions $(\mathbf{H}_b^1)$ and $(\mathbf{H}_b^2)$ hold. Taking $T \to \infty$, we have

$$\|\bar{b}(\cdot,\mu)\|_{p_0} \leqslant \kappa_0.$$

Similarly, we have

$$\|\bar{b}(\cdot,\mu) - \bar{b}(\cdot,\nu)\|_{p_0} \leqslant \kappa_0\|\mu-\nu\|_{var}.$$

Thus, the function $\bar{b}$ satisfies the condition $(\mathbf{H}_b^1)$ with the same constant κ_0 as well.

Then, it is expected that the averaging principle holds. That is, as the time scale ε goes to zero, the solution of the original equation (6.1) converges to that of the following averaged equation on any finite time interval

$$\mathrm{d}X_t = \bar{b}(X_t,\mu_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \quad X_0 = \xi, \tag{6.3}$$

where $\mu_t := \mathscr{L}(X_t)$ stands for the distribution of X_t.

Under assumptions $(\mathbf{H}^\sigma)$ and $(\mathbf{H}_b^1)$ (respectively Remark 6.1), for any initial value $\xi \in \mathscr{F}_0$, it is well-known that there is a unique strong solution to DDSDE (6.1) (respectively, (6.3)); see [117] and [45]. The main result in this chapter is the following strong and weak convergence of the averaging principle for DDSDE and SDE with L^p drift.

Theorem 6.2. *Under $(\mathbf{H}_b^1)$, $(\mathbf{H}_b^2)$ and $(\mathbf{H}^\sigma)$, for any $T > 0$ and $\ell \in (0,1)$, there is a constant C, depending only on $\kappa_0, \kappa_1, T, d, \beta, p_0, p, \ell$, such that for any $\varepsilon > 0$*

$$\sup_{t\in[0,T]} \|\mu_t^\varepsilon - \mu_t\|_{var} \leqslant C \inf_{h>0}\left(h^{\frac{1}{2}-\frac{d}{2p_0}} + \omega\left(\frac{h}{\varepsilon}\right)\right) \tag{6.4}$$

and

$$\mathbb{E}\left(\sup_{t\in[0,T]}|X_t^\varepsilon - X_t|^{2\ell}\right) \leqslant C\inf_{h>0}\left((\omega(h/\varepsilon))^2 + h^{1-\frac{d}{p_0}}\right)^\ell. \tag{6.5}$$

When the drift b is independent of the distribution, the convergence rate is independent of p_0.

Theorem 6.3. *Assume that*

$$b(t,x,\mu)\equiv b(t,x).$$

Under $(\mathbf{H}_b^1)$, $(\mathbf{H}_b^2)$ *and* $(\mathbf{H}^\sigma)$, *for any* $T>0$, $\delta>0$ *and* $\ell\in(0,1)$, *there is a constant* C, *depending only on* $\kappa_0,\kappa_1,T,d,p_0,p,\delta,\ell$, *such that for any* $\varepsilon>0$

$$\mathbb{E}\left(\sup_{t\in[0,T]}|X_t^\varepsilon-X_t|^{2\ell}\right)\leqslant C\inf_{h>0}\left(\left(\omega\left(\frac{h}{\varepsilon}\right)\right)^2+h^{1-\delta}\right)^\ell.$$

Remark 6.4. (i) Since we use the Zvonkin transformation using the parabolic equation, when $\bar{b}=\bar{b}(t)=\bar{b}(\cdot,\mu_t)$ depends on the time variable t, the time regularity for solutions to this parabolic equation affects the convergence rate (see (6.23) and Lemma 6.5 for more details). When $\bar{b}$ is independent of time, we can construct the Zvonkin transformation using the elliptic equation. Hence, the convergence rates in Theorems 6.2 and 6.3 are different.

(ii) Noting that $|\!|\!|f|\!|\!|_{p_0}\lesssim\|f\|_\infty$ for all $p_0\in(1,\infty)$, these results are valid for $p_0=\infty$, in which case the rate of convergence in (6.5) is

$$\inf_{h>0}\left(\left(\omega\left(\frac{h}{\varepsilon}\right)\right)^2+h^{1-\delta}\right)^\ell$$

for any $\delta>0$. In particular, we obtain the convergence rate $\varepsilon^{\frac{1}{3}-\delta}$ for a large number of examples (see e.g. Example 6.8 below), which is faster than $\varepsilon^{\frac{1}{6}}$ in [54].

6.1 Outline

In this section, we give a brief outline of the proof to Theorem 6.2 and Theorem 6.3. First, we note that since the drift of both the DDSDE and SDE in this paper is locally $L_x^{p_0}$ integrable, we cannot use the Gronwall lemma or the generalized Gronwall lemma directly to prove the convergence of X^ε to X as in [54, 92]. On the other hand, our system (6.1) can be rewritten in the following slow-fast system:

$$\begin{cases}\mathrm{d}X_t^\varepsilon=b\left(Y_t^\varepsilon,X_t^\varepsilon,\mu_t^\varepsilon\right)\mathrm{d}t+\sigma(X_t^\varepsilon)\mathrm{d}W_t,\\ \mathrm{d}Y_t^\varepsilon=\dfrac{1}{\varepsilon}\mathrm{d}t.\end{cases}$$

Since the Kolmogorov operator of the fast process $Y_t^\varepsilon=\frac{t}{\varepsilon}$, $t\geqslant0$, does not have a second order elliptic part, we cannot use the technique based on the Poisson equation as in [88].

To overcome these difficulties, we use Zvonkin's transformation to remove the drift b and employ the classical technique of time discretization.

More precisely, consider the following backward PDE for $t \in [0,T]$ related to (6.3)

$$\partial_t u + a_{ij}\partial_i\partial_j u - \lambda u + B \cdot \nabla u + B = 0, \quad u(T) = 0,$$

where $B(t,x) := \bar{b}(x,\mu_t)$, $\lambda \geqslant 0$, is the dissipative term. Under $(\mathbf{H}^\sigma)$ and $(\mathbf{H}_b^1)$, by Lemma 2.8, for a sufficiently large number λ, there is a solution u such that

$$|\nabla u(t,x)| \leqslant \frac{1}{2}, \quad t \in [0,T], x \in \mathbb{R}^d.$$

Hence, if we define $\Phi_t(x) := x + u(t,x)$, then $x \to \Phi_t(x)$ is a C^1 diffeomorphism of $\mathbb{R}^d$. By Itô's formula (3.52), $Y_t^\varepsilon := \Phi_t(X_t^\varepsilon)$ and $Y_t := \Phi_t(X_t)$ solve the following new SDEs:

$$\begin{aligned} \mathrm{d}Y_t^\varepsilon =& \lambda u(t, \Phi_t^{-1}(Y_t^\varepsilon))\mathrm{d}t + (\sigma^*\nabla\Phi_t(\Phi_t^{-1}(Y_t^\varepsilon)))\mathrm{d}W_t \\ &+ \big(b(t/\varepsilon, X_t^\varepsilon, \mu_t^\varepsilon) - \bar{b}(X_t^\varepsilon, \mu_t)\big) \cdot \nabla\Phi_t\big(X_t^\varepsilon\big)\mathrm{d}t \end{aligned}$$

and

$$\mathrm{d}Y_t = \lambda u(t, \Phi_t^{-1}(Y_t))\mathrm{d}t + (\sigma^*\nabla\Phi_t)(\Phi_t^{-1}(Y_t))\mathrm{d}W_t,$$

where σ^* is the transpose of σ and Φ_t^{-1} is the inverse of $x \to \Phi_t(x)$. Since these new systems have differentiable diffusion coefficients and the drifts are Lipshitz continuous, we can use the stochastic Gronwall inequality Lemma A.5.

The remaining part of the proof is about how to estimate the following crucial term

$$\mathbb{E}\Big[\sup_{t\in[0,T]}\Big|\int_0^t \big(b(\frac{s}{\varepsilon}, X_s^\varepsilon, \mu_s^\varepsilon) - \bar{b}(X_s^\varepsilon, \mu_s)\big) \cdot \nabla\Phi_s\big(X_s^\varepsilon\big)\mathrm{d}s\Big|^2\Big]. \tag{6.6}$$

In particular, we need to estimate

$$\|\mu_t - \mu_t^\varepsilon\|_{var} \tag{6.7}$$

and

$$\mathbb{E}\Big[\sup_{t\in[0,T]}\Big|\int_0^t \big(b(\frac{s}{\varepsilon}, X_s^\varepsilon, \mu_s^\varepsilon) - \bar{b}(X_s^\varepsilon, \mu_s^\varepsilon)\big) \cdot \nabla\Phi_s\big(X_s^\varepsilon\big)\mathrm{d}s\Big|^2\Big]. \tag{6.8}$$

In Section 6.2, we will use (3.91) and classical method of time discretization in averaging principle to estimate (6.8).

To estimate (6.7), we employ a method based on the Kolmogorov equation which is also used in [87]. Then, again by time discretization, we estimate the difference (6.7) and obtain (6.4) (see Section 6.3).

In the following, we will first show the crucial lemma estimating (6.8) in Section 6.2. Then, we will show the weak convergence rate (6.4) and strong convergence rate (6.5) in Section 6.3 and 6.4 respectively. Finally, some examples will be given in Section 6.5.

6.2 Crucial lemma

In this section, we prove the following crucial lemma.

Lemma 6.5. *Let $T > 0$ and $g : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ be a bounded function satisfying*

$$c_g := \sup_{t \neq s \in [0,T]} \sup_{x \in \mathbb{R}^d} \Big(|g(t,x)| + |g(t,x) - g(s,x)|/|t-s|^\alpha \Big) < \infty \tag{6.9}$$

for some $\alpha > 0$. Assume $(\mathbf{H}^\sigma)$, $(\mathbf{H}_b^1)$ and $(\mathbf{H}_b^2)$ hold. Then for any $\delta > 0$, there is a constant $C = C(\Xi, \kappa_0, \delta, c_g)$ such that for any $\varepsilon > 0$,

$$\begin{aligned} &\mathbb{E}\left(\sup_{t \in [0,T]} \left| \int_0^t g(s, X_s^\varepsilon)(b(\frac{s}{\varepsilon}, X_s^\varepsilon, \mu_s^\varepsilon) - \bar{b}(X_s^\varepsilon, \mu_s^\varepsilon)) \mathrm{d}s \right|^2 \right) \\ &\leqslant C \inf_{h>0} \left(h^{1-\delta} + h^{2\alpha} + (\omega(h/\varepsilon))^2 \right). \end{aligned} \tag{6.10}$$

Proof. For simplicity, we drop the superscript ε from X^ε and μ^ε. Set $b_\varepsilon(t) := b(t/\varepsilon)$ and $\widetilde{X}_\cdot := (X_\cdot, \mu_\cdot)$. For any $f = f(t,x,\mu)$ and $h > 0$, define $F_h^f(t) := f(s, \widetilde{X}_s) - f(s, \widetilde{X}_{\pi_h(s)})$. Then, for any $h \in (0,1)$, the left hand side of (6.10) is dominated by

$$\begin{aligned} &\mathbb{E}\left(\sup_{t \in [0,T]} \left| \int_0^t F_h^{gb_\varepsilon}(s) \mathrm{d}s \right|^2 \right) + \mathbb{E}\left(\sup_{t \in [0,T]} \left| \int_0^t F_h^{g\bar{b}}(s) \mathrm{d}s \right|^2 \right) \\ &\quad + \mathbb{E}\left(\sup_{t \in [0,T]} \left| \int_0^t g(s, X_{\pi_h(s)})(b_\varepsilon(s, \widetilde{X}_{\pi_h(s)}) - \bar{b}(\widetilde{X}_{\pi_h(s)})) \mathrm{d}s \right|^2 \right) =: \mathscr{I}_1^{\varepsilon,h} + \mathscr{I}_2^{\varepsilon,h} + \mathscr{I}_3^{\varepsilon,h}. \end{aligned}$$

By $(\mathbf{H}_b^1)$ and (3.91), for any $\delta > 0$, we have $\mathscr{I}_1^{\varepsilon,h} + \mathscr{I}_2^{\varepsilon,h} \lesssim \|g\|_\infty^2 \kappa_0^2 h^{1-\delta}$.

For $\mathscr{I}_3^{\varepsilon,h}$, we note that by (6.9) and (3.77) with $p_0 > (d/2) \vee 1$,

$$\begin{aligned} \mathscr{I}_3^{\varepsilon,h} &\lesssim \mathbb{E}\left(\sup_{t \in [0,T]} \left| \int_0^t g(\pi_h(s), X_{\pi_h(s)})(b_\varepsilon(s, \widetilde{X}_{\pi_h(s)}) - \bar{b}(\widetilde{X}_{\pi_h(s)})) \mathrm{d}s \right|^2 \right) \\ &\quad + \mathbb{E}\left(\sup_{t \in [0,T]} \int_0^t |s - \pi_h(s)|^{2\alpha} \Big(|b_\varepsilon(s, \widetilde{X}_{\pi_h(s)})|^2 + |\bar{b}(\widetilde{X}_{\pi_h(s)}))|^2 \Big) \mathrm{d}s \right) \lesssim \mathscr{I}_{31}^{\varepsilon,h} + h^{2\alpha}, \end{aligned}$$

where $\mathscr{I}_{31}^{\varepsilon,h} := \mathbb{E}\left(\sup_{t\in[0,T]} \left| \int_0^t g(\pi_h(s), X_{\pi_h(s)})(b_\varepsilon(s, \widetilde{X}_{\pi_h(s)}) - \bar{b}(\widetilde{X}_{\pi_h(s)})) \mathrm{d}s \right|^2 \right)$.

It suffices to show $\mathscr{I}_{21}^{\varepsilon,h} \lesssim h + (\omega(h/\varepsilon))^2$. Set $B_\varepsilon^h(s) := b_\varepsilon(s, \widetilde{X}_{\pi_h(s)}) - \bar{b}(\widetilde{X}_{\pi_h(s)})$. Indeed, letting $M(t) = [t/h]$ and noting that $\pi_h(s) = s$ for $s \in [0,h)$, we have

$$\begin{aligned} \mathscr{I}_{31}^{\varepsilon,h} \lesssim &\mathbb{E}\left(\sup_{t \in [0,h)} \left| \int_0^t g(s, X_s) B_\varepsilon^h(s) \mathrm{d}s \right|^2 \right) + \mathbb{E}\left(\sup_{t \in [h,T]} \left| \int_{M(t)h}^t g(\pi_h(s), X_{\pi_h(s)}) B_\varepsilon^h(s) \mathrm{d}s \right|^2 \right) \\ &+ \mathbb{E}\left(\sup_{t \in [h,T]} \left| \int_h^{M(t)h} g(\pi_h(s), X_{\pi_h(s)}) B_\varepsilon^h(s) \mathrm{d}s \right|^2 \right) =: \mathscr{I}_{311}^{\varepsilon,h} + \mathscr{I}_{312}^{\varepsilon,h} + \mathscr{I}_{313}^{\varepsilon,h}. \end{aligned}$$

It follows from Hölder's inequality and (3.77) that

$$\begin{aligned}\sum_{i=1}^{2}\mathscr{I}_{31i}^{\varepsilon,h} \lesssim & h\|g\|_\infty^2\mathbb{E}\left[\int_0^T |b_\varepsilon(s,\widetilde{X}_s)|^2+|\bar{b}(\widetilde{X}_s)|^2\mathrm{d}s\right]\\ & + h\|g\|_\infty^2\mathbb{E}\left[\int_0^T |b_\varepsilon(s,\widetilde{X}_{\pi_h(s)})|^2+|\bar{b}(\widetilde{X}_{\pi_h(s)})|^2\mathrm{d}s\right]\\ \lesssim & h\|g\|_\infty^2\kappa_0^2.\end{aligned}$$

Thus, we only need to prove $\mathscr{I}_{313}^{\varepsilon,h} \lesssim \big(\omega\,(h/\varepsilon)\,\big)^2$. By the definition of π_h, it is easy to see that

$$\mathscr{I}_{313}^{\varepsilon,h} \leqslant \mathbb{E}\Big(\sup_{2\leqslant m\leqslant M(T)}\Big|\sum_{k=1}^{m-1} g(kh,X_{kh})\int_{kh}^{(k+1)h}(b_\varepsilon(s,\widetilde{X}_{kh})-\bar{b}(\widetilde{X}_{kh}))\mathrm{d}s\Big|^2\Big).$$

Based on the fact that $|\sum_{k=1}^{m-1}a_k|^2\leqslant (m-1)\sum_{k=1}^{m-1}|a_k|^2$, one sees that

$$\mathscr{I}_{313}^{\varepsilon,h}\leqslant M(T)c_g^2\sum_{k=1}^{M(T)-1}\mathbb{E}\Big|\int_{kh}^{(k+1)h}(b_\varepsilon(s,\widetilde{X}_{kh})-\bar{b}(\widetilde{X}_{kh}))\mathrm{d}s\Big|^2.$$

By a change of variables and (6.2), we have

$$\begin{aligned}\mathscr{I}_{313}^{\varepsilon,h} &\lesssim M(T)\sum_{k=1}^{M(T)-1}\mathbb{E}\left|\varepsilon\int_{kh/\varepsilon}^{(k+1)h/\varepsilon}\big(b(s,\widetilde{X}_{kh})-\bar{b}(\widetilde{X}_{kh})\big)\mathrm{d}s\right|^2\\ &\lesssim h\,(\omega(h/\varepsilon))^2\sum_{k=1}^{M(T)-1}\mathbb{E}|H(\widetilde{X}_{kh})|^2.\end{aligned}$$

We note that

$$h\sum_{k=1}^{M(T)-1}\mathbb{E}|H(\widetilde{X}_{kh})|^2=\mathbb{E}\int_h^{M(T)h}|H(\widetilde{X}_{\pi_h(s)})|^2\mathrm{d}s\leqslant\mathbb{E}\int_h^T|H(\widetilde{X}_{\pi_h(s)})|^2\mathrm{d}s.$$

Again by (3.77), we have $\mathscr{I}_{313}^{\varepsilon,h}\lesssim(\omega(h/\varepsilon))^2\sup_\mu\|H(\cdot,\mu)\|_{p_0}^2$ and complete the proof. □

6.3 Weak convergence

In this section, under $(\mathbf{H}^\sigma)$, $(\mathbf{H}_b^1)$ and $(\mathbf{H}_b^2)$, we will derive the convergence rate of $\|\mu^\varepsilon-\mu\|_{var}$. Recall that on the probability space $(\Omega,\mathscr{F},\mathbb{P},(\mathscr{F}_s)_{s\geqslant 0})$ we have a unique strong solution $(X_\cdot^\varepsilon, X_\cdot)$ to the following systems

$$\mathrm{d}X_t^\varepsilon=b(t/\varepsilon,X_t^\varepsilon,\mu_t^\varepsilon)\mathrm{d}t+\sigma(X_t^\varepsilon)\mathrm{d}W_t,\quad X_0^\varepsilon=\xi, \tag{6.11}$$

and

$$\mathrm{d}X_t = \bar{b}(X_t, \mu_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \quad X_0 = \xi, \tag{6.12}$$

where μ_t^ε and μ_t are the distributions of X_t^ε and X_t respectively. For simplicity, in the sequel, let

$$b_\varepsilon(t) := b(t/\varepsilon) \ \ (\varepsilon > 0), \quad \text{and} \quad b_0 := \bar{b}.$$

For any $x \in \mathbb{R}^d$ and $t \geqslant s \geqslant 0$, let $(Y_{s,t}^\varepsilon(x), Y_{s,t}(x))$ be the unique strong solution to the following SDEs

$$\mathrm{d}Y_{s,t}^\varepsilon(x) = b_\varepsilon(t, Y_{s,t}^\varepsilon(x), \mu_t^\varepsilon)\mathrm{d}t + \sigma(Y_{s,t}^\varepsilon(x))\mathrm{d}W_t, \quad Y_{s,s}^\varepsilon(x) = x$$

and

$$\mathrm{d}Y_{s,t}(x) = \bar{b}(Y_{s,t}(x), \mu_t)\mathrm{d}t + \sigma(Y_{s,t}(x))\mathrm{d}W_t, \quad Y_{s,s}(x) = x.$$

Set $Y_t^\varepsilon(x) := Y_{0,t}^\varepsilon(x)$ and $Y_t(x) := Y_{0,t}(x)$ for all $t \geqslant 0$ and $x \in \mathbb{R}^d$. Let $P^{x,\varepsilon}$ and P^x denote the distributions of $Y_\cdot^\varepsilon(x)$ and $Y_\cdot(x)$ in $C([0,T];\mathbb{R}^d)$ respectively. Based on the strong uniqueness of the above SDEs, we have

$$\int_{\mathbb{R}^d} P^{x,\varepsilon} \ \ \mathbb{P}\circ\xi^{-1}(\mathrm{d}x) = \mathbb{P}\circ(X_\cdot^\varepsilon)^{-1} \quad \text{and} \quad \int_{\mathbb{R}^d} P^x \ \ \mathbb{P}\circ\xi^{-1}(\mathrm{d}x) = \mathbb{P}\circ(X_\cdot)^{-1}. \tag{6.13}$$

Therefore, the estimates in Section 3.2.2 hold for X^ε and X, where the constants are independent of ε, since

$$\sup_{\varepsilon\geqslant 0} \sup_{\mu\in\mathcal{P}(\mathbb{R}^d)} \|b_\varepsilon(\cdot,\mu)\|_{\widetilde{\mathbb{L}}_T^{p_0}} < \infty.$$

Moreover, for any $t \in \mathbb{R}_+$ and $\varphi \in C_b^\infty$, consider the following Kolmogorov backward equation

$$\partial_s u^t + a_{ij}\partial_i\partial_j u^t + b_0(\cdot, \mu_s)\cdot\nabla u^t = 0, \tag{6.14}$$

with final condition

$$u^t(t) = \varphi.$$

By Proposition 2.10 and 3.27, there exists a unique solution u^t to (6.14), which is given by

$$u^t(s,x) = \mathbb{E}\varphi(Y_{s,t}(x)).$$

Define $\tilde{u}(s,x) = u^t(t-s,x)$. Then $\tilde{u}$ is the solution to

$$\partial_s\tilde{u} = a_{ij}\partial_i\partial_j\tilde{u} + b_0(\cdot,\mu_{t-s})\cdot\nabla\tilde{u}, \quad \tilde{u}_0 = \varphi.$$

By Lemma 3.34, we have

$$\|\nabla u^t(s)\|_\infty \lesssim (t-s)^{-\frac{1}{2}}\|\varphi\|_\infty \tag{6.15}$$

and

$$\|\nabla u^t(s_1)-\nabla u^t(s_2)\|_\infty \lesssim |s_1-s_2|^{\frac{1}{2}-\frac{d}{2p_0}}(s_1\wedge s_2)^{-1+\frac{d}{2p_0}}\|\varphi\|_\infty. \tag{6.16}$$

Then, for any $t\in[0,T]$, by applying the generalized Itô formula (3.52) to $u^t(s,Y_s^\varepsilon(x))$, one sees that

$$\mathbb{E}\varphi(Y_t^\varepsilon(x))-u^t(0,x)=\mathbb{E}\int_0^t\Big(b_\varepsilon(s,Y_s^\varepsilon(x),\mu_s^\varepsilon)-b_0(Y_s^\varepsilon(x),\mu_s)\Big)\cdot\nabla u^t(s,Y_s^\varepsilon(x))\mathrm{d}s. \tag{6.17}$$

Noting that $u^t(0,x)=\mathbb{E}\varphi(Y_t(x))$, by (6.13), we have

$$\begin{aligned}|\mathbb{E}\varphi(X_t^\varepsilon)-\mathbb{E}\varphi(X_t)| &= \Big|\int_{\mathbb{R}^d}\big(\mathbb{E}\varphi(Y_t^\varepsilon(x))-\mathbb{E}\varphi(Y_t(x))\big)\mathbb{P}\circ\xi^{-1}(\mathrm{d}x)\Big|\\ &\leqslant \sup_x\Big|\mathbb{E}\int_0^t\Big(b_\varepsilon(s,Y_s^\varepsilon(x),\mu_s^\varepsilon)-b_0(Y_s^\varepsilon(x),\mu_s)\Big)\cdot\nabla u^t(s,Y_s^\varepsilon(x))\mathrm{d}s\Big|.\end{aligned} \tag{6.18}$$

Here is the main result of this section:

Theorem 6.6. *Under the conditions* $(\mathbf{H}^\sigma)$ *and* $(\mathbf{H}_b^1)$*–*$(\mathbf{H}_b^2)$*, for any* $T>0$*, there is a constant* $C=C(\kappa_0,\kappa_1,d,T,p_0,\beta)>0$ *such that for all* $\varepsilon>0$ *and* $t\in[0,T]$*,*

$$\|\mu_t^\varepsilon-\mu_t\|_{var}\leqslant C\inf_{h>0}\left(h^{\frac{1}{2}-\frac{d}{2p_0}}+\omega\left(\frac{h}{\varepsilon}\right)\right). \tag{6.19}$$

Proof. For simplicity, in the whole proof, we assume $\|\varphi\|_\infty=1$ and drop the superscript t from u^t. First, let

$$\mathcal{B}^\varepsilon:=\Big|\mathbb{E}\int_0^t\Big(b_\varepsilon(s,Y_s^\varepsilon(x),\mu_s^\varepsilon)-b_0(Y_s^\varepsilon(x),\mu_s)\Big)\cdot\nabla u(s,Y_s^\varepsilon(x))\mathrm{d}s\Big|$$

and

$$\mathcal{E}_h^\varepsilon:=\Big|\mathbb{E}\int_0^t\Big(b_\varepsilon(s,Y_{\pi_h(s)}^\varepsilon(x),\mu_{\pi_h(s)}^\varepsilon)-b_0(Y_{\pi_h(s)}^\varepsilon(x),\mu_{\pi_h(s)}^\varepsilon)\Big)\cdot\nabla u(\pi_h(s),Y_{\pi_h(s)}^\varepsilon(x))\mathrm{d}s\Big|,$$

where $h \in (0,1)$. For any map $f : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ and $h \in (0,1)$, define

$$U^\varepsilon_{1,h}(f) := \Big|\mathbb{E}\int_0^t \big((f\cdot\nabla u)(s, Y^\varepsilon_s(x), \mu^\varepsilon_s) - (f\cdot\nabla u)(s, Y^\varepsilon_{\pi_h(s)}(x), \mu^\varepsilon_s)\big)\mathrm{d}s\Big|$$
$$U^\varepsilon_{2,h}(f) := \Big|\mathbb{E}\int_0^t \Big(\Big[f(s, Y^\varepsilon_{\pi_h(s)}(x), \mu^\varepsilon_s) - f(s, Y^\varepsilon_{\pi_h(s)}(x), \mu^\varepsilon_{\pi_h(s)})\Big]\cdot\nabla u(s, Y^\varepsilon_{\pi_h(s)}(x))\Big)\mathrm{d}s\Big|$$
$$U^\varepsilon_{3,h}(f) := \Big|\mathbb{E}\int_0^t \Big(f(s, Y^\varepsilon_{\pi_h(s)}(x), \mu^\varepsilon_{\pi_h(s)})\cdot\Big[\nabla u(s, Y^\varepsilon_{\pi_h(s)}(x)) - \nabla u(\pi_h(s), Y^\varepsilon_{\pi_h(s)}(x))\Big]\Big)\mathrm{d}s\Big|.$$

Then, we have

$$\begin{aligned}\mathcal{B}^\varepsilon \leqslant& \mathcal{E}^\varepsilon_h + \sum_{i=1}^3 \Big[U^\varepsilon_{i,h}(b_\varepsilon) + U^\varepsilon_{i,h}(b_0)\Big]\\ &+ \Big|\mathbb{E}\int_0^t \Big(b_0(Y^\varepsilon_s(x), \mu^\varepsilon_s) - b_0(Y^\varepsilon_s(x), \mu_s)\Big)\cdot\nabla u(s, Y^\varepsilon_s(x))\mathrm{d}s\Big|.\end{aligned}$$

It follows from (6.15), (3.76) and $\mathbf{(H^1_b)}$ that

$$\begin{aligned}&\Big|\mathbb{E}\int_0^t \Big(b_0(Y^\varepsilon_s(x), \mu^\varepsilon_s) - b_0(Y^\varepsilon_s(x), \mu_s)\Big)\cdot\nabla u(s, Y^\varepsilon_s(x))\mathrm{d}s\Big|\\ &\lesssim \int_0^t s^{-\frac{d}{2p_0}}(t-s)^{-\frac12}\|\mu^\varepsilon_s - \mu_s\|_{var}\mathrm{d}s,\end{aligned}$$

which implies that

$$\mathcal{B}^\varepsilon \lesssim \mathcal{E}^\varepsilon_h + \sum_{i=1}^3 \Big[U^\varepsilon_{i,h}(b_\varepsilon) + U^\varepsilon_{i,h}(b_0)\Big] + \int_0^t s^{-\frac{d}{2p_0}}(t-s)^{-\frac12}\|\mu^\varepsilon_s - \mu_s\|_{var}\mathrm{d}s. \tag{6.20}$$

Now, we divide the rest of the proof into two steps. In Step 1, we estimate $U^\varepsilon_{i,h}(b_\varepsilon) + U^\varepsilon_{i,h}(b_0)$, $i = 1,2,3$, one by one; In Step 2, we calculate $\mathcal{E}^\varepsilon_h$ under the assumption $\mathbf{(H^2_b)}$.

(Step 1) We only estimate $U^\varepsilon_{i,h}(b_\varepsilon)$, for $U^\varepsilon_{i,h}(b_0)$ we can proceed in the same way. First, we estimate $U^\varepsilon_{1,h}(b_\varepsilon)$. By (3.85) and (6.15), we have

$$\begin{aligned}U^\varepsilon_{1,h}(b_\varepsilon) &= \Big|\int_0^t \big(P^{Y^\varepsilon}_s(b_\varepsilon\cdot\nabla u)(s,\cdot,\mu^\varepsilon_s)(x) - P^{Y^\varepsilon}_{\pi_h(s)}(b_\varepsilon\cdot\nabla u)(s,\cdot,\mu^\varepsilon_s)(x)\big)\mathrm{d}s\Big|\\ &\lesssim \int_0^t \Big[(h^\alpha(\pi_h(s))^{-\alpha})\wedge 1\Big](\pi_h(s))^{-\frac{d}{2p_0}}|\!|\!|(b_\varepsilon\cdot\nabla u)(s,\cdot,\mu^\varepsilon_s)|\!|\!|_{p_0}\mathrm{d}s\\ &\lesssim h^\alpha\int_0^t (\pi_h(s))^{-\alpha}(\pi_h(s))^{-\frac{d}{2p_0}}(t-s)^{-\frac12}\mathrm{d}s,\end{aligned}$$

where $\alpha = 1/2 - d/(2p_0)$. Noting that $\pi_h(s) = s$ for $s \leqslant h$, $\pi_h(s) \leqslant s$ for all $s \in [0,T]$ and $1 - \alpha - d/(2p_0) - 1/2 = 0$, one sees that

$$U^\varepsilon_{1,h}(b_\varepsilon) \lesssim h^{\frac{1}{2}-\frac{d}{2p_0}}.$$

For $U^\varepsilon_{2,h}(b_\varepsilon)$, by (3.76), (6.15) and $(\mathbf{H}^1_b)$, we have

$$U^\varepsilon_{2,h}(b_\varepsilon) \lesssim \int_0^t (\pi_h(s))^{-\frac{d}{2p_0}} \|\mu^\varepsilon_s - \mu^\varepsilon_{\pi_h(s)}\|_{var}(t-s)^{-\frac{1}{2}}\mathrm{d}s.$$

It follows from (3.86) that

$$\|\mu^\varepsilon_s - \mu^\varepsilon_{\pi_h(s)}\|_{var} \lesssim [h^{\frac{1}{2}}(\pi_h(s))^{-\frac{1}{2}}] \wedge 1 \lesssim h^{\frac{1}{2}-\frac{d}{2p_0}}(\pi_h(s))^{-\frac{1}{2}+\frac{d}{2p_0}},$$

which implies that

$$U^\varepsilon_{2,h}(b_\varepsilon) \lesssim h^{\frac{1}{2}-\frac{d}{2p_0}} \int_0^t (\pi_h(s))^{-\frac{1}{2}}(t-\pi_h(s))^{-\frac{1}{2}}\mathrm{d}s \lesssim h^{\frac{1}{2}-\frac{d}{2p_0}}.$$

Finally, in view of (3.76) and (6.16), because $p_0 < \infty$, we have

$$U^\varepsilon_{3,h}(b_\varepsilon) \lesssim h^{\frac{1}{2}-\frac{d}{2p_0}} \int_0^t (\pi_h(s))^{-\frac{d}{2p_0}}(t-\pi_h(s))^{-1+\frac{d}{2p_0}}\mathrm{d}s \lesssim h^{\frac{1}{2}-\frac{d}{2p_0}}$$

and obtain that

$$\sum_{i=1}^3 \big(U^\varepsilon_{i,h}(b_\varepsilon) + U^\varepsilon_{i,h}(\bar{b})\big) \lesssim h^{\frac{1}{2}-\frac{d}{2p_0}}. \tag{6.21}$$

(Step 2) Let $M := [t/h]$. Without loss of generality we may assume that $M = t/h \in \mathbb{N}$ and note that

$$\begin{aligned}\mathcal{E}^\varepsilon_h \leqslant & \Big|\mathbb{E}\int_0^h \Big(b_\varepsilon(s, Y^\varepsilon_s(x), \mu^\varepsilon_s) - b_0(Y^\varepsilon_s(x), \mu^\varepsilon_s)\Big)\cdot \nabla u(s, Y^\varepsilon_s(x))\mathrm{d}s\Big| \\ & + \Big|\sum_{k=1}^{M-1}\mathbb{E}\int_{kh}^{(k+1)h}\Big(b_\varepsilon(s, Y^\varepsilon_{kh}(x), \mu^\varepsilon_{kh}) - b_0(Y^\varepsilon_{kh}(x), \mu^\varepsilon_{kh})\Big)\cdot\nabla u(kh, Y^\varepsilon_{kh}(x))\mathrm{d}s\Big| \\ :=& \mathcal{E}_1 + \mathcal{E}_2.\end{aligned}$$

From (3.76) and (6.15),

$$\mathcal{E}_1 \lesssim \int_0^h s^{-\frac{d}{2p_0}}(t-s)^{-\frac{1}{2}}\mathrm{d}s \lesssim h^{\frac{1}{2}-\frac{d}{2p_0}}.$$

By (6.15) and a change of variables, one sees that

$$\begin{aligned}\mathcal{E}_2 &\lesssim \sum_{k=1}^{M-1}(t-kh)^{-\frac{1}{2}}\Big|\mathbb{E}\int_{kh}^{(k+1)h}\Big(b_\varepsilon(s,Y^\varepsilon_{kh}(x),\mu^\varepsilon_{kh})-b_0(Y^\varepsilon_{kh}(x),\mu^\varepsilon_{kh})\Big)\mathrm{d}s\Big| \\ &\lesssim \sum_{k=1}^{M-1}(t-kh)^{-\frac{1}{2}}\Big|\varepsilon\mathbb{E}\int_{kh/\varepsilon}^{(k+1)h/\varepsilon}\Big(b(s,Y^\varepsilon_{kh}(x),\mu^\varepsilon_{kh})-\bar{b}(Y^\varepsilon_{kh}(x),\mu^\varepsilon_{kh})\Big)\mathrm{d}s\Big|.\end{aligned}$$

Based on the assumptions (6.2) and (3.76), we have

$$\begin{aligned}\mathcal{E}_2 &\lesssim h\sum_{k=1}^{M-1}(t-kh)^{-\frac{1}{2}}\omega\left(\frac{h}{\varepsilon}\right)\mathbb{E}H(Y^\varepsilon_{kh}(x),\mu^\varepsilon_{kh}) \\ &\lesssim h\sum_{k=1}^{M-1}(t-kh)^{-\frac{1}{2}}\omega\left(\frac{h}{\varepsilon}\right)(kh)^{-\frac{d}{2p_0}}\sup_{\mu}\|H(\cdot,\mu)\|_{p_0} \\ &\lesssim \omega\left(\frac{h}{\varepsilon}\right)\int_h^t(t-\pi_h(s))^{-\frac{1}{2}}(\pi_h(s))^{-\frac{d}{2p_0}}\mathrm{d}s \lesssim \omega\left(\frac{h}{\varepsilon}\right)\end{aligned}$$

and obtain that

$$|\mathbb{E}\varphi(X^\varepsilon_t)-\mathbb{E}\varphi(X_t)| \lesssim \left(h^{\frac{1}{2}-\frac{d}{2p_0}}+\omega\left(\frac{h}{\varepsilon}\right)\right)+\int_0^t s^{-\frac{d}{2p_0}}(t-s)^{-\frac{1}{2}}\|\mu^\varepsilon_s-\mu_s\|_{var}\mathrm{d}s$$

because of (6.18), (6.20) and (6.21). Finally, taking the supremum over all $\varphi\in C_b^\infty$ with $\|\varphi\|_\infty=1$ and by Lemma A.4, we complete the proof.

□

6.4 Strong convergence

In this section, we consider the process (X^ε, W, X) on the probability space $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geqslant 0},\mathbb{P})$ which satisfies the following system in $\mathbb{R}^d$:

$$X^\varepsilon_t = \xi+\int_0^t b(\frac{s}{\varepsilon},X^\varepsilon_s,\mu^\varepsilon_s)\mathrm{d}s+\int_0^t\sigma(X^\varepsilon_s)\mathrm{d}W_s$$

and

$$X_t=\xi+\int_0^t\bar{b}(X_s,\mu_s)\mathrm{d}s+\int_0^t\sigma(X_s)\mathrm{d}W_s,$$

where W is a standard d-dimensitional Brownian Motion, μ^ε_t and μ_t are the distributions of X^ε_t and X_t respectively and $(b,\bar{b},\sigma)$ satisfies the conditions $(\mathbf{H}^1_b)$-$(\mathbf{H}^2_b)$ and $(\mathbf{H}^\sigma)$. We set

$$X^0:=X,\quad b_\varepsilon(t):=b(t/\varepsilon),\quad\text{and}\quad b_0:=\bar{b}.$$

Proof of Theoerm 6.2. Set

$$B(t,x) := b_0(x,\mu_t)$$

and consider the following backward parabolic PDE

$$\partial_t u + a_{ij}\partial_i\partial_j u - \lambda u + B\cdot\nabla u + B = 0, \quad t\in[0,T], \quad u(T)=0.$$

Since $\|B\|_{\widetilde{\mathbb{L}}^{p_0}(T)} \leqslant \sup_{\mu}\|\bar{b}(\cdot,\mu)\|_{p_0} < \infty$, by Lemma 2.8, for λ large enough there is a unique solution u in the sense of Definition 2.6 satisfying

$$\|\nabla u\|_{\mathbb{L}^\infty_T} \leqslant \frac{1}{2}$$

and for any $2/q + d/p_0 < 1$,

$$\|\nabla^2 u\|_{\widetilde{\mathbb{L}}^{p_0}_q(T)} \leqslant C,$$

which implies that for any $\lambda > 0$

$$\sup_{\varepsilon\geqslant 0}\mathbb{E}\exp\left(\lambda\int_0^T |\nabla^2 u(t,X^\varepsilon_t)|^2\mathrm{d}t\right) < \infty, \tag{6.22}$$

where $X^0 := X$, because of (3.78). Moreover, by (3.93), for all $s,t\in[0,T]$,

$$\|\nabla u(t) - \nabla u(s)\|_\infty \lesssim |t-s|^{1/2-d/(2p_0)}. \tag{6.23}$$

Define

$$\Phi_t(x) := x + u(t,x)$$

and

$$Y^\varepsilon_t := \Phi_t(X^\varepsilon_t), \quad Y_t := \Phi_t(X_t).$$

Then Φ_t is a C^1-diffeomorphism (see Remark 3.33) for any $t\in[0,T]$ with

$$\|\nabla\Phi\|_{\mathbb{L}^\infty_T} + \|\nabla\Phi^{-1}\|_{\mathbb{L}^\infty_T} \leqslant 4. \tag{6.24}$$

By the generalized Itô formula (3.52), we have

$$\mathrm{d}Y_t = \lambda u(t,X_t)\mathrm{d}t + (\sigma^*\nabla\Phi_t)(X_t)\mathrm{d}W_t$$

and

$$\mathrm{d}Y^\varepsilon_t = \lambda u(t,X^\varepsilon_t)\mathrm{d}t + \big(b_\varepsilon(t,X^\varepsilon_t,\mu^\varepsilon_t) - b_0(X^\varepsilon_t,\mu_t)\big)\cdot\nabla\Phi_t\big(X^\varepsilon_t\big)\mathrm{d}t + (\sigma^*\nabla\Phi_t(X^\varepsilon_t))\mathrm{d}W_t,$$

where σ^* is the transpose of σ. It follows from (6.24) that for any $t \in [0, T]$,

$$
\begin{aligned}
|X_t^\varepsilon - X_t|^2 \lesssim |Y_t^\varepsilon - Y_t|^2 \lesssim & \lambda \|\nabla u\|_{\mathbb{L}_T^\infty}^2 \int_0^t |X_s^\varepsilon - X_s|^2 \mathrm{d}s \\
& + \left[\int_0^t \big((\sigma^* \nabla \Phi_s)(X_s^\varepsilon) - (\sigma^* \nabla \Phi_s)(X_s) \big) \mathrm{d}W_s \right]^2 \\
& + \left| \int_0^t \big(b_\varepsilon(s, X_s^\varepsilon, \mu_s^\varepsilon) - b_0(X_s^\varepsilon, \mu_s) \big) \cdot \nabla \Phi_s \big(X_s^\varepsilon \big) \mathrm{d}s \right|^2.
\end{aligned}
$$

Set

$$
\begin{aligned}
A_t^\varepsilon := & \int_0^t \left(\mathcal{M}(\nabla^2 u)(s, X_s) + \mathcal{M}(\nabla^2 u)(s, X_s^\varepsilon) + \|\nabla u\|_{\mathbb{L}_T^\infty} \right)^2 \mathrm{d}s \\
& + \int_0^t \left(\mathcal{M}(\nabla \sigma)(X_s) + \mathcal{M}(\nabla \sigma)(X_s^\varepsilon) + \|\sigma\|_\infty \right)^2 \mathrm{d}s
\end{aligned}
$$

and

$$
\eta_t^\varepsilon := \left| \int_0^t \big(b_\varepsilon(s, X_s^\varepsilon, \mu_s^\varepsilon) - b_0(X_s^\varepsilon, \mu_s) \big) \cdot \nabla \Phi_s \big(X_s^\varepsilon \big) \mathrm{d}s \right|^2.
$$

Then, by (2.16), (6.22) and $(\mathbf{H}^\sigma)$, we have

$$
\sup_\varepsilon \mathbb{E} \exp(A_T^\varepsilon) < \infty. \tag{6.25}
$$

We note that by (2.15)

$$
\begin{aligned}
& \left[\int_0^t \big((\sigma^* \nabla \Phi)(X_s^\varepsilon) - (\sigma^* \nabla \Phi)(X_s) \big) \mathrm{d}W_s \right]^2 \\
& \leqslant \int_0^t |X_s^\varepsilon - X_s|^2 \mathrm{d}A_s^\varepsilon + M_t^\varepsilon,
\end{aligned}
$$

where M^ε is a martingale. Altogether, we have

$$
|X_t^\varepsilon - X_t|^2 \lesssim \int_0^t |X_s^\varepsilon - X_s|^2 \mathrm{d}s + \int_0^t |X_s^\varepsilon - X_s|^2 \mathrm{d}A_s^\varepsilon + M_t^\varepsilon + \eta_t^\varepsilon.
$$

Hence, by (6.25) and the stochastic Gronwall inequality Lemma A.5, one sees that for any $\ell \in (0, 1)$,

$$
\mathbb{E} \Big(\sup_{t \in [0,T]} |X_t^\varepsilon - X_t|^{2\ell} \Big) \lesssim \Big(\mathbb{E} [\sup_{t \in [0,T]} \eta_t^\varepsilon] \Big)^\ell.
$$

Combining (6.23), (6.24) and (6.10), we have

$$\mathbb{E}[\sup_{t\in[0,T]} \eta_t^\varepsilon] \lesssim \mathbb{E}\int_0^T \left|b_0(X_s^\varepsilon,\mu_s^\varepsilon) - b_0(X_s^\varepsilon,\mu_s)\right|^2 \mathrm{d}s + \inf_{h>0}\left(h^{1-\delta} + h^{1-\frac{d}{p_0}} + \omega\left(\frac{h}{\varepsilon}\right)\right). \tag{6.26}$$

Taking $\delta < d/p_0$ in (6.26), from (3.77) and (6.19), for any $2/q + d/p_0 < 1$, one sees that

$$\begin{aligned}\mathbb{E}[\sup_{t\in[0,T]} \eta_t^\varepsilon] &\lesssim \left(\int_0^T \|\mu_s^\varepsilon - \mu_s\|_{var}^q \mathrm{d}s\right)^{2/q} + \inf_{h>0}\left(h^{1-\frac{d}{p_0}} + \left(\omega\left(\frac{h}{\varepsilon}\right)\right)^2\right)\\ &\lesssim \inf_{h>0}\left(h^{1-\frac{d}{p_0}} + \left(\omega\left(\frac{h}{\varepsilon}\right)\right)^2\right)\end{aligned}$$

and this completes the proof.

□

In the rest of this section, we assume that

$$b_\varepsilon(t,x,\mu) = b_\varepsilon(t,x), \quad b_0(x,\mu) = b_0(x)$$

and prove Theorem 6.3. The method is the same as the one of Theorem 6.2, except for using the elliptic equation to construct the Zvonkin's transformation instead of the parabolic.

Proof of Theorem 6.3. Consider the following elliptic PDE

$$a_{ij}\partial_i\partial_j u - \lambda u + b_0 \cdot \nabla u + b_0 = 0. \tag{6.27}$$

Noting that $\|b_0\|_{\widetilde{\mathbb{L}}^{p_0}(T)} \leqslant \|b\|_{L^\infty(\mathbb{R}_+;\widetilde{L}^{p_0})}$ and by (2.24) for λ large enough, we have

$$\|\nabla u\|_\infty \leqslant \frac{1}{2}$$

and

$$\|\nabla^2 u\|_{p_0} \leqslant C\|b_0\|_{\widetilde{L}^{p_0}}, \quad \forall p_0 > d$$

It follows by (3.78) that for any $\lambda > 0$

$$\sup_{\varepsilon\geqslant 0}\mathbb{E}\exp\left(\lambda\int_0^T |\nabla^2 u(X_t^\varepsilon)|^2 \mathrm{d}t\right) < \infty. \tag{6.28}$$

Define

$$\Phi(x) := x + u(x)$$

and

$$Y_t^\varepsilon := \Phi(X_t^\varepsilon), \quad Y_t := \Phi(X_t).$$

Then Φ is a C^1-diffeomorphism. Again by the generalized Itô formula (3.52), we have

$$\mathrm{d}Y_t = \lambda u(X_t)\mathrm{d}t + (\sigma^*\nabla\Phi)(X_t)\mathrm{d}W_t$$

and

$$\mathrm{d}Y_t^\varepsilon = \lambda u(X_t^\varepsilon)\mathrm{d}t + \big(b_\varepsilon(t, X_t^\varepsilon) - b_0(X_t^\varepsilon)\big) \cdot \nabla\Phi\big(X_t^\varepsilon\big)\mathrm{d}t + (\sigma^*\nabla\Phi)(X_t^\varepsilon)\mathrm{d}W_t.$$

Then, we have

$$|X_t^\varepsilon - X_t|^2 \lesssim \int_0^t |X_s^\varepsilon - X_s|^2 \mathrm{d}A_s^\varepsilon + M_t^\varepsilon + \eta_t^\varepsilon,$$

where $(M_t^\varepsilon)_{t\geqslant 0}$ is a martingale,

$$\begin{aligned} A_t^\varepsilon =& t + \int_0^t \big(\mathcal{M}(\nabla^2 u)(X_s) + \mathcal{M}(\nabla^2 u)(X_s^\varepsilon) + \|\sigma\|_\infty\big)^2 \mathrm{d}s \\ &+ \int_0^t (\mathcal{M}(\nabla\sigma)(X_s) + \mathcal{M}(\nabla\sigma)(X_s^\varepsilon) + \|\nabla u\|_\infty)^2 \mathrm{d}s \end{aligned}$$

and

$$\eta_t^\varepsilon = \Big| \int_0^t \big(b(s/\varepsilon, X_s^\varepsilon) - \bar{b}(X_s^\varepsilon)\big) \cdot \nabla\Phi\big(X_s^\varepsilon\big)\mathrm{d}s\Big|^2.$$

Then, in view of (2.16) and (6.28), we have

$$\sup_\varepsilon \mathbb{E}\exp(A_T^\varepsilon) < \infty,$$

which implies that for any $\ell \in (0,1)$,

$$\mathbb{E}\Big(\sup_{t\in[0,T]} |X_t^\varepsilon - X_t|^{2\ell}\Big) \lesssim \Big(\mathbb{E} \sup_{t\in[0,T]} \eta_t^\varepsilon\Big)^\ell$$

because of the Lemma A.5 and we complete the proof by (6.10) with $\alpha = 1$. □

6.5 Examples

Example 6.7. Consider the following DDSDE in $\mathbb{R}^d$

$$\begin{aligned} \mathrm{d}X_t^\varepsilon &= \bigg(\Big[(1+t/\varepsilon)^{-\alpha_1} + 1\Big] \int_{\mathbb{R}^d} \frac{X_t^\varepsilon - y}{|X_t^\varepsilon - y|^{\alpha_2}} \mu_t^\varepsilon(\mathrm{d}y)\bigg)\mathrm{d}t + \mathrm{d}W_t \\ &=: b(t/\varepsilon, X_t^\varepsilon, \mu_t^\varepsilon)\mathrm{d}t + \mathrm{d}W_t, \end{aligned}$$

where $\alpha_1 > 0$, $1 < \alpha_2 < 2 \wedge (1 + \frac{d}{2})$ and μ_t^ε is the distribution of X_t^ε. It is clear that the averaged equation is

$$\begin{aligned} \mathrm{d}X_t &= \left(\int_{\mathbb{R}^d} \frac{X_t - y}{|X_t - y|^{\alpha_2}} \mu_t(\mathrm{d}y) \right) \mathrm{d}t + \mathrm{d}W_t \\ &=: \bar{b}(X_t, \mu_t)\mathrm{d}t + \mathrm{d}W_t, \end{aligned}$$

where μ_t is the distribution of X_t, and

$$\left| \frac{1}{T} \int_t^{t+T} \big(b(s, x, \mu) - \bar{b}(x, \mu) \big) \, \mathrm{d}s \right| \leqslant \omega\,(T)\,(1 - \alpha_1)^{-1} \int_{\mathbb{R}^d} \frac{|x - y|}{|x - y|^{\alpha_2}} \mu(\mathrm{d}y)$$

for all $(T, t, x, \mu) \in \mathbb{R}^2 \times \mathbb{R} \times \mathcal{P}(\mathbb{R}^d)$, where

$$\omega(t) = \begin{cases} t^{-(\alpha_1 \wedge 1)} & \text{for } \alpha_1 \in (0, 1) \cup (1, \infty) \\ t^{-1} \log t & \text{for } \alpha_1 = 1. \end{cases}$$

Then we have for any $\delta > 0$

$$\sup_{t \in [0,T]} \|\mu_t^\varepsilon - \mu_t\|_{var} \leqslant C\varepsilon^{\frac{\alpha(2-\alpha_2)}{2+2\alpha-\alpha_2} - \delta}$$

and

$$\left[\mathbb{E} \left(\sup_{t \in [0,T]} |X_t^\varepsilon - X_t|^{2\ell} \right) \right]^{\frac{1}{\ell}} \leqslant C\varepsilon^{\frac{4\alpha - 2\alpha\alpha_2}{2+2\alpha-\alpha_2} - \delta}$$

for any $0 < \ell < 1$, where $\alpha = \alpha_1 \wedge 1$ for $\alpha_1 \in (0, \infty)$.

Next we give a more general example, where the function $\omega(t) \asymp t^{-1}$, i.e. there exists a constant C such that $C^{-1}t^{-1} \leqslant \omega(t) \leqslant Ct^{-1}$.

Example 6.8. Let $p_0 \in (d \vee 2, \infty)$. Consider the following DDSDE

$$\mathrm{d}X_t^\varepsilon = \left[\int_{\mathbb{R}} F \left(\sin(\xi t/\varepsilon), \int_{\mathbb{R}^d} \phi(X_t^\varepsilon, y) \mu_t^\varepsilon(\mathrm{d}y) \right) \nu(\mathrm{d}\xi) \right] \mathrm{d}t + \mathrm{d}W_t, \tag{6.29}$$

where μ_t^ε is the time marginal law of X_t, $F : [-1, 1] \times \mathbb{R}^m \to \mathbb{R}^d$ is measurable and satisfies for some constant $L_F > 0$

$$|F(t, 0)| \leqslant L_F, \quad |F(t, x) - F(t, y)| \leqslant L_F |x - y| \quad \text{for all } (t, x, y) \in [-1, 1] \times \mathbb{R}^{2m}, \tag{6.30}$$

ν is some finite measure on $\mathbb{R}$ satisfying

$$\int_{\mathbb{R}\setminus\{0\}} \frac{\nu(\mathrm{d}\xi)}{|\xi|} < \infty$$

and $\phi : \mathbb{R}^d \to \mathbb{R}^m$ is measurable and satisfies

$$\sup_y \|\phi(\cdot, y)\|_{p_0} < \infty.$$

Set

$$b(t,x,\mu) := \int_{\mathbb{R}} F\left(\sin(\xi t), \int_{\mathbb{R}^d} \phi(x,y)\mu(\mathrm{d}y)\right)\nu(\mathrm{d}\xi)$$

and

$$\begin{aligned}\bar{b}(x,\mu) :=& \frac{1}{2\pi}\int_0^{2\pi} F\left(\sin\tau, \int_{\mathbb{R}^d} \phi(x,y)\mu(\mathrm{d}y)\right)\nu(\mathbb{R}\setminus\{0\})\mathrm{d}\tau \\ &+ F\left(0, \int_{\mathbb{R}^d} \phi(x,y)\mu(\mathrm{d}y)\right)\nu(\{0\}).\end{aligned}$$

We claim that

$$(b,\bar{b}) \text{ satisfies conditions } (\mathbf{H}_b^1) \text{ and } (\mathbf{H}_b^2) \text{ with } \omega(T) := \frac{4\pi L_F}{T}\int_{\mathbb{R}\setminus\{0\}} \frac{\nu(\mathrm{d}\xi)}{|\xi|}. \tag{6.31}$$

To prove this claim, we need the following lemma.

Lemma 6.9. *Let*

$$h(t) := \int_{\mathbb{R}} B(\sin(\xi t))\nu(\mathrm{d}\xi).$$

and

$$\bar{h} := \Big(\frac{1}{2\pi}\int_0^{2\pi} B(\sin(\xi\tau))\mathrm{d}\tau\Big)\nu(\mathbb{R}\setminus\{0\}) + B(0)\nu(\{0\}),$$

where $B : [-1,1] \to \mathbb{R}^d$ *is measurable and* ν *is a finite measure on* $\mathbb{R}$*. Assume that there is a constant* $C_B > 0$ *such that*

$$|B(u)| \leqslant C_B, \quad \forall u \in [-1,1].$$

Then, for any $t, T \in \mathbb{R}_+$*,*

$$\Big|\frac{1}{T}\int_t^{t+T} \big(h(s) - \bar{h}\big)\mathrm{d}s\Big| \leqslant \frac{4\pi C_B}{T}\int_{\mathbb{R}\setminus\{0\}} \frac{\nu(\mathrm{d}\xi)}{|\xi|}.$$

Proof of Lemma 6.9. If $\int_{\mathbb{R}\setminus\{0\}} \frac{\nu(\mathrm{d}\xi)}{|\xi|} = \infty$, this is trivial. So, we assume that $\int_{\mathbb{R}\setminus\{0\}} \frac{\nu(\mathrm{d}\xi)}{|\xi|} < \infty$. First, one sees that

$$\begin{aligned}\mathscr{I} :=& \left|\frac{1}{T}\int_t^{t+T} \big(h(s) - \bar{h}\big)\mathrm{d}s\right| \\ =& \left|\frac{1}{T}\int_t^{t+T}\int_{\mathbb{R}\setminus\{0\}} B(\sin(\xi s))\nu(\mathrm{d}\xi)\mathrm{d}s - \frac{1}{2\pi}\int_t^{t+2\pi} B(\sin(\tau))\nu(\mathbb{R}\setminus\{0\})\mathrm{d}\tau\right| \\ =& \left|\int_{\mathbb{R}\setminus\{0\}} \Big[\frac{1}{T}\int_t^{t+T} B(\sin(\xi s))\mathrm{d}s - \frac{1}{2\pi}\int_t^{t+2\pi} B(\sin(\tau))\mathrm{d}\tau\Big]\nu(\mathrm{d}\xi)\right|,\end{aligned}$$

by Fubini's theorem. From a change of variable, we have

$$\mathscr{I} = \left| \int_{\mathbb{R}\setminus\{0\}} \Big[\frac{1}{T|\xi|} \int_{t\xi}^{(t+T)\xi} B(\sin s)\mathrm{d}s - \frac{1}{2\pi} \int_t^{t+2\pi} B(\sin \tau)\mathrm{d}\tau \Big] \nu(\mathrm{d}\xi) \right|,$$

where $\int_a^b := -\int_b^a$ if $a > b$. Set

$$G := \int_0^{2\pi} B(\sin s)\mathrm{d}s = \int_t^{t+2\pi} B(\sin s)\mathrm{d}s, \quad \forall t \in \mathbb{R}.$$

Then, noting that $s \to \sin s$ has a period 2π, we have

$$\begin{aligned} \int_{t\xi}^{(t+T)\xi} B(\sin s)\mathrm{d}s &= \Big[\frac{T|\xi|}{2\pi}\Big] G + \int_{\mathrm{sgn}(\xi)[\frac{T|\xi|}{2\pi}]2\pi + t\xi}^{T\xi + t\xi} B(\sin s)\mathrm{d}s \\ &:= \Big[\frac{T|\xi|}{2\pi}\Big] G + H_t(\xi) \end{aligned}$$

where $\mathrm{sgn}(\xi) := \xi/|\xi|$, which implies that

$$\begin{aligned} \mathscr{I} &= \left| \int_{\mathbb{R}\setminus\{0\}} \Big[\frac{1}{T|\xi|} \Big(\Big[\frac{T|\xi|}{2\pi}\Big] G + H_t(\xi) \Big) - \frac{1}{2\pi} G \Big] \nu(\mathrm{d}\xi) \right| \\ &\leqslant \int_{\mathbb{R}\setminus\{0\}} \left| \frac{1}{T|\xi|} \Big[\frac{T|\xi|}{2\pi}\Big] - \frac{1}{2\pi} \right| \nu(\mathrm{d}\xi) G + \int_{\mathbb{R}\setminus\{0\}} \frac{1}{T|\xi|} H_t(\xi) \nu(\mathrm{d}\xi). \end{aligned}$$

We note that

$$\left| \frac{1}{T|\xi|} \Big[\frac{T|\xi|}{2\pi}\Big] - \frac{1}{2\pi} \right| = \frac{1}{T|\xi|} \left| \Big[\frac{T|\xi|}{2\pi}\Big] - \frac{T|\xi|}{2\pi} \right| \leqslant \frac{1}{T|\xi|}$$

and

$$G \vee H_t(\xi) \leqslant \int_0^{2\pi} |B(\sin s)|\mathrm{d}s \leqslant 2\pi C_B.$$

Therefore, we have

$$\mathscr{I} \leqslant \frac{4\pi C_B}{T} \int_{\mathbb{R}\setminus\{0\}} \frac{\nu(\mathrm{d}\xi)}{|\xi|}$$

and complete the proof.

□

Now we can give the

Proof of (6.31). Since $(\mathbf{H}_b^1)$ holds for b obviously, it suffice to show that $(\mathbf{H}_b^2)$ holds. We note that in Example 6.8

$$|F(u,x)| \leqslant |F(u,0)| + |F(u,0) - F(t,x)| \leqslant L_F + L_F|x|$$

because of (6.30), which implies that

$$|F\big(u, \int_{\mathbb{R}^d} \phi(x,y)\mu(\mathrm{d}y)\big)| \leqslant L_F(1 + \int_{\mathbb{R}^d} |\phi(x,y)|\mu(\mathrm{d}y)).$$

Hence, by Lemma 6.9, we see that

$$\left|\frac{1}{T}\int_t^{t+T} \big(b(s,x,\mu) - \bar{b}(x,\mu)\big)\mathrm{d}s\right| \leqslant \frac{4\pi L_F}{T}\int_{\mathbb{R}\backslash\{0\}} \frac{\nu(\mathrm{d}\xi)}{|\xi|} H(x,\mu),$$

where $H(x,\mu) = 1 + \int_{\mathbb{R}^d} |\phi(x,y)|\mu(\mathrm{d}y)$. It is easy to see that

$$\sup_\mu |||H(\cdot,\mu)|||_{p_0} \leqslant \|1\|_\infty + \int_{\mathbb{R}^d} |||\phi(\cdot,y)|||_{p_0}\mu(\mathrm{d}y) \leqslant 1 + \sup_y |||\phi(\cdot,y)|||_{p_0}.$$

This completes the proof. □

Chapter 7

Euler-Maruyama scheme for McKean-Vlasov SDEs of Nemytskii-type with bounded drifts

In this chapter, we consider the following distributional-density dependent SDE (dDSDE):

$$\mathrm{d}X_t = b(t, X_t, \rho_t(X_t))\mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \quad X_0 \overset{(d)}{=} \nu_0, \tag{7.1}$$

where ρ_t is the distributional density of X_t respect to Lebesgue measure, $b : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$ is bounded measurable, ν_0 is a probability measure over $\mathbb{R}^d$ and $\{W_t\}_{t\geqslant 0}$ is a standard d-dimensional Brownian motion on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. As said in the introduction, by Itô's formula, one sees that ρ_t solves the following nonlinear Fokker-Planck equation (FPE) in the distributional sense:

$$\partial_t \rho_t - \Delta \rho_t + \mathrm{div}(b(t, \cdot, \rho_t)\rho_t) = 0, \quad \lim_{t\downarrow 0} \rho_t = \nu_0 \text{ weakly.} \tag{7.2}$$

More precisely, for any $\varphi \in C_0^\infty(\mathbb{R}^d)$,

$$\langle \rho_t, \varphi \rangle = \langle \nu_0, \varphi \rangle + \int_0^t \langle \rho_s, \Delta\varphi \rangle \mathrm{d}s + \int_0^t \langle \rho_s, b(s, \cdot, \rho_s) \cdot \nabla\varphi \rangle \mathrm{d}s, \tag{7.3}$$

where $\langle \rho_t, \varphi \rangle := \int_{\mathbb{R}^d} \varphi(x)\rho_t(x)\mathrm{d}x = \mathbb{E}\varphi(X_t)$. Let us first recall the definition of a weak solution to dDSDE (7.1):

Definition 7.1. *Let ν_0 be a probability measure on $\mathbb{R}^d$. We call a filtered probability space $(\Omega, \mathscr{F}, \mathbf{P}; (\mathscr{F}_t)_{t\geqslant 0})$ together with a pair of processes (X, W) defined on it a weak solution of SDE (7.1) with initial distribution ν_0, if*

(i) $\mathbf{P} \circ X_0^{-1} = \nu_0$, W is a d-dimensional $\mathscr{F}_t$-Brownian motion.

(ii) for each $t > 0$, $\mathbf{P} \circ X_t^{-1}(\mathrm{d}x)/\mathrm{d}x = \rho_t(x)$ and

$$X_t = X_0 + \int_0^t b(s, X_s, \rho_s(X_s))\mathrm{d}s + \sqrt{2}W_t, \quad \mathbf{P} - a.s.$$

We also consider the following Euler scheme to dDSDE (7.1): Let $T > 0$ and $h \in (0, 1)$. For $t \in [0, h]$, we define

$$X_t^h := X_0 + \sqrt{2}W_t,$$

and for $t \in [kh, (k+1)h]$ with $k = 1, 2, .., [\frac{T}{h}]$,

$$X_t^h = X_{kh}^h + \int_{kh}^t b(s, X_{kh}^h, \rho_{kh}^h(X_{kh}^h))\mathrm{d}s + \sqrt{2}(W_t - W_{kh}), \tag{7.4}$$

where ρ_{kh}^h is the distributional density of X_{kh}^h, whose existence is obviously seen from the construction of X_{kh}^h. Here is the main result in this chapter.

Theorem 7.2. *Assume that b is bounded measurable and for any $t > 0$,*

$$\lim_{u \to u_0} |b(t, x, u) - b(t, x, u_0)| = 0. \tag{7.5}$$

(Existence) *For any $T > 0$ and initial distribution ν_0, there are a subsequence h_k with $\lim_{k\to\infty} h_k = 0$ and a weak solution (X, W) to dDSDE (7.1) in the sense of Definition 7.1 so that for any bounded measurable f and $t \in (0, T]$,*

$$\lim_{k\to\infty} \mathbb{E}f(X_t^{h_k}) = \mathbb{E}f(X_t).$$

Moreover, for each $t \in (0, T]$, X_t admits a density ρ_t satisfying the estimate

$$\rho_t(y) \leqslant Ct^{-d/2} \int_{\mathbb{R}^d} \mathrm{e}^{-\frac{|x-y|^2}{\lambda t}} \nu_0(\mathrm{d}x), \ y \in \mathbb{R}^d,$$

where $C, \lambda \geqslant 1$ only depend on $T, d, \|b\|_\infty$, and the following L^1-convergence holds:

$$\lim_{k\to\infty} \int_{\mathbb{R}^d} |\rho_t^{h_k}(y) - \rho_t(y)|\mathrm{d}y = 0. \tag{7.6}$$

(Uniqueness and convergence rate) *Suppose that $\nu_0(\mathrm{d}x) = \rho_0(x)\mathrm{d}x$ with $\rho_0 \in (L^1 \cap L^q)(\mathbb{R}^d)$ for some $q \in (d, \infty]$, and there is a $\kappa > 0$ such that for all t, x, u, u',*

$$|b(t, x, u) - b(t, x, u')| \leqslant \kappa|u - u'|. \tag{7.7}$$

Then weak and strong uniqueness hold for dDSDE (7.1). Moreover, if $q \geqslant 2$, for any $T > 0$, there is a constant $C = C(d, T, \kappa, \|b\|_\infty)$ such that for all $t \in [0, T]$ and $h \in (0, 1)$,

$$\|\rho_t - \rho_t^h\|_1 \leqslant Ch^{\frac{1}{2}}. \tag{7.8}$$

Remark 7.3. It should be noted that this part is based on the results of [44]. Therein, the condition is stronger than (7.5), and we didn't obtain the convergence rate (7.8), which is new in this thesis. Moreover, the related results are improved to α-stable processes cases in [48].

7.1 Euler-Maruyama scheme for SDEs with bounded drift

In this section we show heat kernel estimates and weak convergence rate for Euler's scheme of SDEs with bounded drift. First of all, we recall some basic properties about the Gaussian heat kernel. Recall

$$g(t,x) = g_t(x) = (4\pi t)^{-\frac{d}{2}} \mathrm{e}^{-\frac{|x|^2}{4t}}, \ t > 0, x \in \mathbb{R}^d, \tag{7.9}$$

which is the fundamental solution of Δ, i.e.,

$$\partial_t g(t,x) = \Delta g(t,x).$$

Moreover, we have the following Chapman-Kolmogorov equations:

$$(g(t) * g(s))(x) := \int_{\mathbb{R}^d} g(t, x-z) g(s,z) \mathrm{d}z = g(t+s, x), \ t, s > 0, \tag{7.10}$$

and the following easy facts,

$$g(t, x+y) \leqslant 2^{\frac{d}{2}} g(2t, x) \mathrm{e}^{\frac{|y|^2}{4t}}, \ \ |\nabla g|(t,x) \leqslant \frac{2^{d/2}}{\sqrt{t}} g(2t, x). \tag{7.11}$$

The following lemma is straightforward and elementary. For the readers' convenience, we provide a detail proof.

Lemma 7.4. *For any $T > 0$, $\beta \in (0,1)$ and $j = 0, 1$, there is a constant $C = C(T, \beta, j, d) > 0$ such that for any $0 < t \leqslant T$ and $x_1, x_2 \in \mathbb{R}^d$,*

$$|\nabla^j g(t, x_1) - \nabla^j g(t, x_2)| \leqslant C |x_1 - x_2|^\beta t^{-\frac{j}{2} - \beta} \sum_{i=1,2} g(4t, x_i), \tag{7.12}$$

and for any $0 < t_1 < t_2 \leqslant T$ and $x \in \mathbb{R}^d$,

$$|\nabla^j g(t_1, x) - \nabla^j g(t_2, x)| \leqslant C |t_2 - t_1|^{\frac{\beta}{2}} \sum_{i=1,2} t_i^{-\frac{j+\beta}{2}} g(2t_i, x). \tag{7.13}$$

Proof. (i) By definition (7.9), it is easy to see that for $k = 1, 2, 3$, there is a constant $C > 0$ only depending on k, d such that

$$|\nabla^k g(t,x)| \leqslant C (4\pi t)^{-\frac{d}{2}} t^{\frac{k}{2}} \mathrm{e}^{-\frac{|x|^2}{8t}} = C 2^{\frac{d}{2}} t^{-\frac{k}{2}} g(2t, x). \tag{7.14}$$

Thus, for $j = 0, 1$ and $\beta \in (0,1)$, if $|x_1 - x_2| > \sqrt{t}$, then

$$\begin{aligned} |\nabla^j g(t, x_1) - \nabla^j g(t, x_2)| &\lesssim t^{-\frac{j}{2}} (g(2t, x_1) + g(2t, x_2)) \\ &\lesssim |x_1 - x_2|^\beta t^{-\frac{j}{2} - \beta} (g(2t, x_1) + g(2t, x_2)); \end{aligned}$$

if $|x_1 - x_2| \leqslant \sqrt{t}$, then by the mean-value formula,

$$
\begin{aligned}
|\nabla^j g(t,x_1) - \nabla^j g(t,x_2)| &\leqslant |x_1 - x_2| \int_0^1 |\nabla^{j+1} g(t, x_1 + \theta(x_2 - x_1))| \mathrm{d}\theta \\
&\lesssim |x_1 - x_2| t^{-\frac{j+1}{2}} \int_0^1 g(2t, x_1 + \theta(x_2 - x_1)) \mathrm{d}\theta \\
&\lesssim |x_1 - x_2| t^{-\frac{j+1}{2}} g(4t, x_1) \lesssim |x_1 - x_2|^\beta t^{-j/2-\beta} g(4t, x_1).
\end{aligned}
$$

Combining the above calculations, we get (7.12).

(ii) If $t_2 - t_1 \leqslant t_1$, then by the mean-value formula,

$$
\begin{aligned}
|\nabla^j g(t_1,x) - \nabla^j g(t_2,x)| &\leqslant |t_1 - t_2| \int_0^1 |\nabla^j \partial_t g|(t_1 + \theta(t_2 - t_1), x) \mathrm{d}\theta \\
&= |t_1 - t_2| \int_0^1 |\nabla^j \Delta g|(t_1 + \theta(t_2 - t_1), x) \mathrm{d}\theta \\
&\lesssim |t_1 - t_2| \int_0^1 \frac{g(2(t_1 + \theta(t_2 - t_1)), x)}{(t_1 + \theta(t_2 - t_1))^{1+j/2}} \mathrm{d}\theta \\
&\lesssim |t_1 - t_2| t_1^{-1-\frac{j}{2}} g(2t_2, x) \lesssim |t_1 - t_2|^{\frac{\beta}{2}} t_2^{-\frac{\beta}{2}} g(2t_2, x);
\end{aligned}
$$

if $t_2 - t_1 > t_1$, then $t_2 \leqslant 2(t_2 - t_1)$ and

$$
\begin{aligned}
|\nabla^j g(t_1,x) - \nabla^j g(t_2,x)| &\lesssim t_1^{-\frac{j}{2}} g(2t_1, x) + t_2^{-\frac{j}{2}} g(2t_2, x) \\
&\lesssim |t_1 - t_2|^{\frac{\beta}{2}} \Big(t_1^{-\frac{j+\beta}{2}} g(2t_1, x) + t_2^{-\frac{j+\beta}{2}} g(2t_2, x) \Big).
\end{aligned}
$$

The proof is complete. □

Let $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ be a bound measurable function. Fix $T > 0$ and $x \in \mathbb{R}^d$. For any $h \in (0,1)$, let $X_t^h = X_t^h(x)$ be defined by the following Euler scheme:

$$
X_t^h = x + \int_0^t b(s, X^h_{\pi_h(s)}) \mathrm{d}s + \sqrt{2} W_t, \quad t \in [0,T], \tag{7.15}
$$

where $\pi_h(s) := kh$ for $s \in [kh, (k+1)h)$. We have the following Duhamel formula.

Lemma 7.5. *For each $t \in (0,T]$ and $x \in \mathbb{R}^d$, $X_t^h(x)$ admits a density $p_x^h(t,y)$ which satisfies the following Duhamel formula:*

$$
p_x^h(t,y) = g(t, x - y) + \int_0^t \mathbb{E}\Big[b(s, X^h_{\pi_h(s)}) \cdot \nabla g(t - s, y - X_s^h) \Big] \mathrm{d}s. \tag{7.16}
$$

Proof. Fix $t \in (0, T]$ and $f \in C_c^\infty(\mathbb{R}^d)$. For $s \in [0, t]$, let $u(s, x) := g(t - s, \cdot) * f(x)$. Since $(\partial_s + \Delta)u \equiv 0$ and $u(t, x) = f(x)$, by Itô's formula, we have

$$\mathbb{E}f(X_t^h) = \mathbb{E}u(t, X_t^h) = u(0, x) + \int_0^t \mathbb{E}\Big[b(s, X_{\pi_h(s)}^h) \cdot \nabla u(s, X_s^h)\Big]\mathrm{d}s.$$

From this, we derive the desired Duhamel formula. □

Remark 7.6. For a general initial value $X_0^h = X_0 \in \mathscr{F}_0$ and each $t \in (0, T]$, since for each $x \in \mathbb{R}^d$, $X_t^h(x)$ is independent of X_0, the Euler scheme X_t^h defined by (7.15) with initial value X_0 also has a density $p_{X_0}^h(t, y)$ given by

$$p_{X_0}^h(t, y) = \int_{\mathbb{R}^d} p_x^h(t, y)\mathbb{P} \circ X_0^{-1}(\mathrm{d}x). \tag{7.17}$$

The following Gaussian type estimate for $p_x^h(t, y)$ was proved by Lemaire and Menozzi [71]. Since it is not difficult, for the readers' convenience, we provide a detailed proof here.

Theorem 7.7. *For any $T > 0$, there is a constant $C = C(d, T, \|b\|_\infty)$ such that for all $h \in (0, 1)$, $t \in (0, T]$ and $x, y \in \mathbb{R}^d$,*

$$p_x^h(t, y) \leqslant Cg(4t, x - y). \tag{7.18}$$

Proof. Let $\varepsilon > 0$ be small enough so that

$$\ell_\varepsilon := 2^{d+1}\sqrt{\varepsilon\|b\|_\infty^2}\mathrm{e}^{\varepsilon\|b\|_\infty^2} \leqslant 1/2.$$

Fix $T > 0$. Without loss of generality, we assume

$$Th^{-1} \geqslant (\|b\|_\infty^2 T/(4\log 2)) \vee (T/\varepsilon). \tag{7.19}$$

For simplicity we shall write

$$M := [\varepsilon/h] \in \mathbb{N}.$$

Step 1: In this step we use induction to show that for all $k = 1, \cdots, M \wedge N$,

$$p_x^h(kh, y) \leqslant 2^{d+1}g(4kh, x - y). \tag{7.20}$$

First of all, for $k = 1$, since $X_h^h = x + W_h + \int_0^h b(s, x)\mathrm{d}s$, by (7.11) and (7.19) we have

$$\begin{aligned} p_x^h(h, y) = g(h, y - \int_0^h b(s, x)\mathrm{d}s - x) &\leqslant 2^{d/2}\mathrm{e}^{\|b\|_\infty^2 h/4}g(2h, x - y) \\ &\leqslant 2^d\mathrm{e}^{\|b\|_\infty^2 h/4}g(4h, x - y) \leqslant 2^{d+1}g(4h, x - y). \end{aligned}$$

Suppose now that (7.20) holds for $j = 1, 2, .., k-1$. By Duhamel's formula (7.16), we have

$$\begin{aligned} p_x^h(kh, y) - g(kh, x-y) &= \int_0^{kh} \mathbb{E}\Big[b(s, X^h_{\pi_h(s)}) \cdot \nabla g(kh-s, y - X^h_s)\Big]\mathrm{d}s \\ &= \sum_{j=0}^{k-1} \int_{jh}^{(j+1)h} \mathbb{E}\Big[b(s, X^h_{jh}) \cdot \nabla g(kh-s, y - X^h_s)\Big]\mathrm{d}s. \end{aligned} \tag{7.21}$$

Note that for $s \in (jh, (j+1)h)$,

$$X^h_s = X^h_{jh} + \sqrt{2}(W_s - W_{jh}) + \int_{jh}^s b(r, X^h_{jh})\mathrm{d}r.$$

Since $\sqrt{2}(W_s - W_{jh})$ is independent of X^h_{jh} and has density $g(s-jh, y)$, by the C-K equations (7.10) we have

$$\begin{aligned} \mathscr{I}_j(s) &:= \mathbb{E}\Big[b(s, X^h_{jh}) \cdot \nabla g(kh-s, X^h_s - y)\Big] \\ &= \mathbb{E}\Big[b(s, X^h_{jh}) \cdot \nabla g(kh-s) * g(s-jh)\Big(X^h_{jh} + \int_{jh}^s b(r, X^h_{jh})\mathrm{d}r - y\Big)\Big] \\ &= \mathbb{E}\Big[b(s, X^h_{jh}) \cdot \nabla g\Big(kh - jh, X^h_{jh} + \int_{jh}^s b(r, X^h_{jh})\mathrm{d}r - y\Big)\Big] \\ &\leqslant \|b\|_\infty \int_{\mathbb{R}^d} |\nabla g|\Big(kh-jh, z + \int_{jh}^s b(r,z)\mathrm{d}r - y\Big) p_x^h(jh, z)\mathrm{d}z. \end{aligned}$$

By (7.11) and induction hypothesis, we further have for $s \in (jh, (j+1)h)$,

$$\begin{aligned} \mathscr{I}_j(s) &\leqslant \frac{\|b\|_\infty 2^{d/2}}{\sqrt{kh-jh}} \int_{\mathbb{R}^d} g\Big(2(kh-jh), z + \int_{jh}^s b(r,z)\mathrm{d}r - y\Big) p_x^h(jh,z)\mathrm{d}z \\ &\leqslant \frac{\|b\|_\infty 2^d \mathrm{e}^{(k-j)h\|b\|^2_\infty/4}}{\sqrt{kh-jh}} \int_{\mathbb{R}^d} g(4(kh-jh), z-y) \cdot 2^{d+1} g(4jh, x-z)\mathrm{d}z \\ &\leqslant \frac{\|b\|_\infty 2^{2d+1} \mathrm{e}^{\varepsilon\|b\|^2_\infty/4}}{\sqrt{kh-s}} g(4kh, x-y) = \frac{2^d \ell_\varepsilon/\sqrt{\varepsilon}}{\sqrt{kh-s}} g(4kh, x-y), \end{aligned}$$

where we have used $kh \leqslant Mh \leqslant \varepsilon$. Substituting this into (7.21), we obtain

$$\begin{aligned} |p_x^h(kh, y) - g(kh, x-y)| &\leqslant \frac{2^d \ell_\varepsilon}{\sqrt{\varepsilon}} g(4kh, x-y) \sum_{j=0}^{k-1} \int_{jh}^{(j+1)h} \frac{1}{\sqrt{kh-s}}\mathrm{d}s \\ &\leqslant \frac{2^d \ell_\varepsilon}{\sqrt{\varepsilon}} g(4kh, x-y) \int_0^{kh} \frac{1}{\sqrt{kh-s}}\mathrm{d}s \\ &= \frac{2^d \ell_\varepsilon}{\sqrt{\varepsilon}} g(4kh, x-y) 2\sqrt{kh} \leqslant 2^{d+1}\ell_\varepsilon g(4kh, x-y), \end{aligned}$$

which implies, since $g(t,x) \leqslant 2^d g(4t,x)$ and $2\ell_\varepsilon \leqslant 1$, that

$$p_x^h(kh,y) \leqslant 2^d(1+2\ell_\varepsilon)g(4kh,x-y) \leqslant 2^{d+1}g(4kh,x-y).$$

Step 2: Next we assume $M < T/h$ and consider $k = M+1\cdots,2M$. Note that

$$\begin{aligned} X_{t+Mh}^h &= X_{Mh}^h + W_{t+Mh} - W_{Mh} + \int_{Mh}^{t+Mh} b(s, X_{\pi_h(s)}^h)\mathrm{d}s \\ &= X_{Mh}^h + W_{t+Mh} - W_{Mh} + \int_0^t b(s+Mh, X_{\pi_h(s)+Mh}^h)\mathrm{d}s, \end{aligned}$$

where we have used that $\pi_h(s+Mh) = \pi_h(s) + Mh$. In particular, if we let

$$\bar{X}_t^h := X_{t+Mh}^h, \quad \bar{W}_t := W_{t+Mh} - W_{Mh},$$

then for $t \in [0, Mh]$,

$$\bar{X}_t^h = X_{Mh}^h + \bar{W}_t + \int_0^t b(s+Mh, \bar{X}_{\pi_h(s)}^h)\mathrm{d}s.$$

Let $\bar{p}_x^h(kh,y)$ be the density of $\bar{X}_t^h$ with $\bar{X}_0^N = x$. By Step 1, we have

$$\bar{p}_x^h(kh,y) \leqslant 2^{d+1}g(4kh,x-y), \ \ k=1,\cdots,M.$$

Thus, for $k = 1,\cdots,M$, by (7.17) we have

$$\begin{aligned} p_x^h((k+M)h,y) &= \int_{\mathbb{R}^d} \bar{p}_z^N(kh,y)p_x^h(Mh,z)\mathrm{d}z \\ &\leqslant 4^{d+1}\int_{\mathbb{R}^d} g(4kh,z-y)g(4Mh,x-z)\mathrm{d}z \\ &= 4^{d+1}g(4(k+M)h,x-y). \end{aligned}$$

Repeating the above procedure $[\frac{T}{\varepsilon}]+1$-times, we obtain that for some $C>0$ independent of h,

$$p_x^h(kh,y) \leqslant Cg(kh,x-y), \quad k=1,\cdots,N.$$

Step 3: Note that for $t \in (kh,(k+1)h)$,

$$X_t^h = X_{kh}^h + W_t - W_{kh} + \int_{kh}^t b(s, X_{kh}^h)\mathrm{d}s,$$

where $W_t - W_{kh}$ is independent of X_{kh}^h. Hence,

$$\begin{aligned} p_x^h(t,y) &= \int_{\mathbb{R}^d} g(t-kh, z + \int_{kh}^t b(s,z)\mathrm{d}s - y)p_x^h(kh,z)\mathrm{d}z \\ &\leqslant C\mathrm{e}^{(t-kh)\|b\|_\infty^2/4}\int_{\mathbb{R}^d} g(4(t-kh),y-z)g(4kh,x-z)\mathrm{d}z \\ &\leqslant C\mathrm{e}^{T\|b\|_\infty^2/4}g(4t,x-y). \end{aligned}$$

This completes the proof. $\square$

The following corollary is a combination of Theorem 7.7 and Lemma 7.4.

Corollary 7.8. *Let* $\nu_0(\mathrm{d}y) = \mathbb{P} \circ X_0^{-1}(\mathrm{d}y)$ *be the distribution of* X_0.

(i) For any $T > 0$*, there is a constant* $C = C(d, T, \|b\|_\infty)$ *such that for all* $h \in (0,1)$*,* $t \in (0,T)$ *and* $y \in \mathbb{R}^d$

$$p^h_{X_0}(t,y) \leqslant C \int_{\mathbb{R}^d} g(4t, x-y)\nu_0(\mathrm{d}x). \tag{7.22}$$

(ii) For any $T > 0$ *and* $\beta \in [0,1]$*, there is a constant* $C = C(d,T,\|b\|_\infty, \beta)$ *such that for all* $h \in (0,1)$*,* $t \in (0,T)$ *and* $y_1, y_2 \in \mathbb{R}^d$*,*

$$|p^h_{X_0}(t,y_1) - p^h_{X_0}(t,y_2)| \leqslant C|y_1-y_2|^\beta t^{-\frac{\beta}{2}} \sum_{j=1,2} \int_{\mathbb{R}^d} g(4t, x-y_j)\nu_0(\mathrm{d}x).$$

(iii) For any $T > 0$ *and* $\beta \in [0,1]$*, there is a constant* $C = C(d,T,\|b\|_\infty, \beta)$ *such that for all* $h \in (0,1)$*,* $t_1, t_2 \in (0,T)$ *and* $y \in \mathbb{R}^d$*,*

$$|p^h_{X_0}(t_1,y) - p^h_{X_0}(t_2,y)| \leqslant C|t_1-t_2|^{\beta/2} \sum_{j=1,2} t_j^{-\beta/2} \int_{\mathbb{R}^d} g(4t_j, x-y)\nu_0(\mathrm{d}x).$$

Proof. (i) is a direct consequence of (7.17) and Theorem 7.7. We only show (iii) since (ii) is similar by (7.12). Suppose $t_1 < t_2$. By (7.16), we have

$$\begin{aligned}|p^h_{X_0}(t_1,y) - p^h_{X_0}(t_2,y)| &\leqslant \int_{\mathbb{R}^d} |g(t_1, x-y) - g(t_2,x-y)|\nu_0(\mathrm{d}x) \\ &+ \|b\|_\infty \int_0^{t_1}\int_{\mathbb{R}^d} |\nabla g(t_1-s, y-z) - \nabla g(t_2-s,y-z)| p^h_{X_0}(s,z)\mathrm{d}z\mathrm{d}s \\ &+ \|b\|_\infty \int_{t_1}^{t_2}\int_{\mathbb{R}^d} |\nabla g(t_1-s, y-z)| p^h_{X_0}(s,z)\mathrm{d}z\mathrm{d}s =: I_1 + I_2 + I_3.\end{aligned}$$

For I_1, by (7.13), we have

$$I_1 \lesssim |t_1-t_2|^{\frac{\beta}{2}} \sum_{j=1,2} t_j^{-\frac{\beta}{2}} \int_{\mathbb{R}^d} g(2t_j, x-y)\nu_0(\mathrm{d}x).$$

For I_2, by (i), (7.13) and the C-K equations (7.10), we have

$$\begin{aligned}I_2 &\lesssim \int_0^{t_1} \left[\left[|t_1-t_2|(t_1-s)^{-3/2}\right] \wedge (t_1-s)^{-1/2}\right] \sum_{j=1,2} \int_{\mathbb{R}^d} g(4(t_j-s), z-y) \\ &\qquad\qquad \times \int_{\mathbb{R}^d} g(4s, x-z)\nu_0(\mathrm{d}x)\mathrm{d}z\mathrm{d}s \\ &\lesssim \int_0^{t_1} \left[\left[|t_1-t_2|s^{-1}\right] \wedge 1\right] s^{-1/2}\mathrm{d}s \sum_{j=1,2} \int_{\mathbb{R}^d} g(4t_j, x-y)\nu_0(\mathrm{d}x).\end{aligned}$$

Noting that

$$\int_0^\infty \Big[s^{-1}\wedge 1\Big]s^{-1/2}\mathrm{d}s<\infty,$$

we have

$$I_2\lesssim |t_1-t_2|^{1/2}\sum_{j=1,2}\int_{\mathbb{R}^d} g(4t_j,x-y)\nu_0(\mathrm{d}x)$$

provided by a change of variable. For I_3, by (i), (7.13) and the C-K equations, we have

$$\begin{aligned}I_3 &\lesssim \int_{t_1}^{t_2}\int_{\mathbb{R}^d}(t_1-s)^{-\frac12}g(4(t_1-s),z-y)\int_{\mathbb{R}^d}g(4s,x-z)\nu_0(\mathrm{d}x)\mathrm{d}z\mathrm{d}s\\ &=\int_{t_1}^{t_2}\int_{\mathbb{R}^d}(t_1-s)^{-\frac12}g(4t_1,x-y)\nu_0(\mathrm{d}x)\mathrm{d}s\lesssim |t_2-t_1|^{\frac12}\int_{\mathbb{R}^d}g(4t_1,x-y)\nu_0(\mathrm{d}x).\end{aligned}$$

Combining the above calculations, we obtain the desired estimate. □

In the sequel, we set

$$\rho_t^h:=p_{X_0}^h(t).$$

and give the following result for later use.

Lemma 7.9. *There is a constant $C=C(d)$ such that for any $f_1,\nabla^k f_2\in L^\infty(\mathbb{R}^d)$ with $k=1,2$, $h\in(0,1)$ and $s>h$,*

$$\Big|\mathbb{E}f_1(X^h_{\pi_h(s)})\left(f_2(X^h_s)-f_2(X^h_{\pi_h(s)})\right)\Big|\leqslant Ch\|f_1\|_\infty\left(\|\nabla f_2\|_\infty\|b\|_\infty+\|\nabla^2 f_2\|_\infty\right). \tag{7.23}$$

Proof. Since $W_s-W_{\pi_h(s)}$ is independent of $X^h_{\pi_h(s)}$, one sees that

$$\begin{aligned}\mathscr{I}_h&:=\mathbb{E}f_1(X^h_{\pi_h(s)})\left(f_2(X^h_s)-f_2(X^h_{\pi_h(s)})\right)\\ &=\int_{\mathbb{R}^{2d}}f_1(x)\cdot\Big(f_2\big(x+\int_{\pi_h(s)}^s b(r,x)\mathrm{d}r+y\big)-f_2(x)\Big)\rho^h_{\pi_h(s)}(x)g(s-\pi_h(s),y)\mathrm{d}x\mathrm{d}y.\end{aligned}$$

Then we have

$$\begin{aligned}|\mathscr{I}_h|\leqslant&\|f_1\|_\infty\int_{\mathbb{R}^{2d}}\Big|f_2\big(x+\int_{\pi_h(s)}^s b(r,x)\mathrm{d}r+y\big)-f_2(x+y)\Big|\rho^h_{\pi_h(s)}(x)g(s-\pi_h(s),y)\mathrm{d}x\mathrm{d}y\\ &+\left|\int_{\mathbb{R}^{2d}}f_1(x)\cdot\Big(f_2(x+y)-f_2(x)\Big)\rho^h_{\pi_h(s)}(x)g(s-\pi_h(s),y)\mathrm{d}x\mathrm{d}y\right|\\ :=&\mathscr{I}_1+\mathscr{I}_2.\end{aligned}$$

For $\mathscr{I}_1$, it is easy to see that

$$\begin{aligned}\mathscr{I}_1 &\leqslant \|f_1\|_\infty\|\nabla f_2\|_\infty \int_{\mathbb{R}^{2d}} \left(\int_{\pi_h(s)}^s |b(r,x)|\mathrm{d}r\right) \rho^h_{\pi_h(s)}(x) g(s-\pi_h(s),y)\mathrm{d}x\mathrm{d}y \\ &\leqslant h\|f_1\|_\infty\|\nabla f_2\|_\infty\|b\|_\infty \int_{\mathbb{R}^{2d}} \rho^h_{\pi_h(s)}(x)g(s-\pi_h(s),y)\mathrm{d}x\mathrm{d}y \leqslant h\|f_1\|_\infty\|\nabla f_2\|_\infty\|b\|_\infty.\end{aligned}$$

For $\mathscr{I}_2$, it follows from the symmetry of $g(t,\cdot)$ that

$$\begin{aligned}\mathscr{I}_2 &= \frac{1}{2}\left|\int_{\mathbb{R}^{2d}} f_1(x)\cdot\Big(f_2(x+y)+f_2(x-y)-2f_2(x)\Big)\rho^h_{\pi_h(s)}(x)g(s-\pi_h(s),y)\mathrm{d}x\mathrm{d}y\right| \\ &\leqslant \|f_1\|_\infty\|\nabla^2 f_2\|_\infty \int_{\mathbb{R}^{2d}} |y|^2\rho^h_{\pi_h(s)}(x)g(s-\pi_h(s),y)\mathrm{d}x\mathrm{d}y \\ &\lesssim h\|f_1\|_\infty\|\nabla^2 f_2\|_\infty.\end{aligned}$$

This completes the proof. □

7.2 Proof of the main result

Let $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geqslant 0},\mathbb{P})$ be a complete filtered probability space, W_t a d-dimensional standard $\mathscr{F}_t$-Brownian motion, and X_0 an $\mathscr{F}_0$-measurable random variable with distribution ν_0. Let $T>0$ and $h\in(0,1)$. Let X^h_t be the Euler approximation of dDSDE (7.1) constructed at the beginning of this chapter. From the construction (7.4), it is easy to see that X^h_t solves the following SDE:

$$X^h_t = X_0 + \int_0^t b^h(s,X^h_{\pi_h(s)})\mathrm{d}s + \sqrt{2}W_t, \tag{7.24}$$

where

$$b^h(s,x) = \mathbf{1}_{\{s\geqslant h\}} b\Big(s,x,\rho^h_{\pi_h(s)}(x)\Big) \tag{7.25}$$

and

$$\pi_h(s) := \sum_{j=0}^\infty jh\mathbf{1}_{[jh,(j+1)h)}(s). \tag{7.26}$$

The following lemma is easy by (7.24) and since $\|b^h\|_\infty \leqslant \|b\|_\infty$.

Lemma 7.10. *For any $T>0$, there is a constant $C>0$ such that for all $s,t\in[0,T]$,*

$$\sup_N \mathbb{E}|X^h_t - X^h_s|^4 \leqslant C|s-t|^2.$$

Let $p^h_x(t, y)$ be the distributional density of the Euler scheme $X^h_t(x)$ of SDE (7.24) starting from x at time 0. Since for each $x \in \mathbb{R}^d$, $X^h_t(x)$ is independent of X_0, the distributional density $\rho^h_t(y)$ of X^h_t with initial distribution ν_0 is given by

$$\rho^h_t(y) = \int_{\mathbb{R}^d} p^h_x(t, y)\nu_0(\mathrm{d}x). \tag{7.27}$$

The following lemma is crucial for the existence of a solution to dDSDE (7.1).

Lemma 7.11. *For fixed $T > 0$, there are a subsequence $(h_k)_{k\in\mathbb{N}}$ and a continuous function $\rho \in C((0, T] \times \mathbb{R}^d)$ such that for any $M \in \mathbb{N}$,*

$$\lim_{k\to\infty} \sup_{|y|\leqslant M} \sup_{1/M\leqslant t\leqslant T} |\rho^{h_k}_t(y) - \rho_t(y)| = 0. \tag{7.28}$$

Proof. First of all, by the upper-bound estimate (7.18) for $p^h_x(t, y)$, we have

$$\sup_{|y|\leqslant M} \sup_{1/M\leqslant t\leqslant T} |\rho^h_t(y)| \leqslant C \int_{\mathbb{R}^d} \sup_{|y|\leqslant M} \sup_{1/M\leqslant t\leqslant T} |g(4t, x - y)|\nu_0(\mathrm{d}x) \leqslant C_M,$$

where C_M is independent of h. Moreover, by Corollary 7.8, we have for any $\beta < 1$, $t_1, t_2 \in [1/M, T]$ and $y_1, y_2 \in \mathbb{R}^d$,

$$\begin{aligned} |\rho^h_{t_1}(y_1) - \rho^h_{t_2}(y_2)| &\leqslant |\rho^h_{t_1}(y_1) - \rho^h_{t_2}(y_1)| + |\rho^h_{t_2}(y_1) - \rho^h_{t_2}(y_2)| \\ &\lesssim |t_2 - t_1|^{\frac{\beta}{2}} \sum_{i=1,2} \int_{\mathbb{R}^d} |g(2t_i, y_1 - x)|\nu_0(\mathrm{d}x) \\ &+ |y_1 - y_2|^{\beta} \sum_{i=1,2} \int_{\mathbb{R}^d} |g(4t_2, y_i - x)|\nu_0(\mathrm{d}x) \\ &\lesssim M^{-(d+1+\beta)/2}\big(|t_2 - t_1|^{\frac{\beta}{2}} + |y_1 - y_2|^{\beta}\big), \end{aligned} \tag{7.29}$$

where the implicit constants in the above $\lesssim$ are independent of h. Thus, by Ascolli-Arzela's theorem, we conclude the proof and have (7.28). □

Now we are in a position to give the

Proof of Theorem 7.2. **(Existence)** Fix $T > 0$. Let $\mathbb{W}$ be the space of all continuous functions from $[0, T]$ to $\mathbb{R}^d$. Let $\mathbb{Q}_h$ be the law of $(X^h_\cdot, W_\cdot)$ in $\mathbb{W} \times \mathbb{W}$. By Lemma 7.10 and Kolmogorov's criterion, $\{\mathbb{Q}_h\}_{N\in\mathbb{N}}$ is tight. Therefore, by Prokhorov's theorem, there are a subsequence $(h_k)_{k\in\mathbb{N}}$ and a probability measure $\mathbb{Q}$ on $\mathbb{W} \times \mathbb{W}$ so that

$$\mathbb{Q}_{h_k} \to \mathbb{Q} \quad \text{weakly.}$$

Without loss of generality, we assume that the subsequence is the same as that in Lemma 7.11. Below, we still denote the above subsequence by $h_k = h$ for simplicity. Now, by

Skorokhod's representation theorem, there are probability space $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{\mathbb{P}})$ and random variables $(\widetilde{X}^h, \widetilde{W}^h)$ and $(\widetilde{X}, \widetilde{W})$ thereon such that

$$(\widetilde{X}^h, \widetilde{W}^h) \to (\widetilde{X}, \widetilde{W}), \quad \widetilde{\mathbb{P}}\text{-a.s.} \tag{7.30}$$

and

$$\widetilde{\mathbb{P}} \circ (\widetilde{X}^h, \widetilde{W}^h)^{-1} = \mathbb{Q}_N = \mathbb{P} \circ (X^h, W)^{-1}, \quad \widetilde{\mathbb{P}} \circ (\widetilde{X}, \widetilde{W})^{-1} = \mathbb{Q}. \tag{7.31}$$

In particular, the distributional density of $\widetilde{X_t^h}$ is ρ_t^h. Moreover, by Lemma 7.11 and (7.30), for any $t \in (0,T)$ and $\varphi \in C_0^\infty(\mathbb{R}^d)$,

$$\mathbb{E}\varphi(\widetilde{X}_t) = \lim_{N\to\infty} \mathbb{E}\varphi(\widetilde{X}_t^h) = \lim_{N\to\infty} \int_{\mathbb{R}^d} \varphi(y)\rho_t^h(y)\mathrm{d}y = \int_{\mathbb{R}^d} \varphi(y)\rho_t(y)\mathrm{d}y.$$

In other words, ρ_t is the density of $\widetilde{X}_t$. Define $\widetilde{\mathscr{F}}_t^h := \sigma(\widetilde{X}^h, \widetilde{W}^h; s \leqslant t)$. We note that

$$\mathbb{P}[W_t - W_s \in \cdot | \mathscr{F}_s] = \mathbb{P}\{W_t - W_s \in \cdot\},$$

hence,

$$\widetilde{\mathbb{P}}[\widetilde{W}_t^h - \widetilde{W}_s^h \in \cdot | \widetilde{\mathscr{F}}_s^N] = \widetilde{\mathbb{P}}\{\widetilde{W}_t^h - \widetilde{W}_s^h \in \cdot\},$$

which means that $\widetilde{W}_t^h$ is an $\widetilde{\mathscr{F}}_t^h$-BM. Thus, by (7.24) and (7.31) we have

$$\widetilde{X}_t^h = \widetilde{X}_0^h + \int_0^t b^h(s, \widetilde{X}_{\pi_h(s)}^h)\mathrm{d}s + \sqrt{2}\widetilde{W}_t^h. \tag{7.32}$$

Let us now show that

$$\int_0^t \mathbf{1}_{s\geqslant h} b\Big(s, \widetilde{X}_{\pi_h(s)}^h, \rho_{\pi_h(s)}^h(\widetilde{X}_{\pi_h(s)}^h)\Big)\mathrm{d}s \to \int_0^t b\Big(s, \widetilde{X}_s, \rho_s(\widetilde{X}_s)\Big)\mathrm{d}s, \tag{7.33}$$

in probability as $h \to 0$.

We note that by (7.18), the dominated convergence theorem, (7.5), (7.28) and (7.29),

$$\begin{aligned}
&\lim_{h\to0} \widetilde{\mathbb{E}} \int_0^t |\mathbf{1}_{s>h} b(s, \widetilde{X}_{\pi_h(s)}^h, \rho_{\pi_h(s)}^h(\widetilde{X}_{\pi_h(s)}^h)) - b(s, \widetilde{X}_{\pi_h(s)}^h, \rho_s(\widetilde{X}_{\pi_h(s)}^h))|\mathrm{d}s \\
&= \lim_{h\to0} \int_0^t \int_{\mathbb{R}^d} |\mathbf{1}_{s>h} b(s, x, \rho_{\pi_h(s)}^h(x)) - b(s, x, \rho_s(x))|\rho_{\pi_h(s)}^h(x)\mathrm{d}x\mathrm{d}s \\
&\leqslant \int_0^t \int_{\mathbb{R}^d} \lim_{h\to0} |\mathbf{1}_{s>h} b(s, x, \rho_{\pi_h(s)}^h(x)) - b(s, x, \rho_s(x))| \int_{\mathbb{R}^d} g(4t, y-x)\nu_0(\mathrm{d}y)\mathrm{d}x\mathrm{d}s = 0.
\end{aligned}$$

For proving (7.33), it remains to show

$$\lim_{h\to0} \widetilde{\mathbb{E}} \int_h^t |b(s, \widetilde{X}_{\pi_h(s)}^h, \rho_s(\widetilde{X}_{\pi_h(s)}^h)) - b(s, \widetilde{X}_s, \rho_s(\widetilde{X}_s))|\mathrm{d}s = 0.$$

Let K_ε be a family of mollifiers in $\mathbb{R}^d$. Define

$$B_\varepsilon(t,x) = b(t,\cdot,\rho_t(\cdot)) * K_\varepsilon(x).$$

Clearly, for fixed $\varepsilon > 0$, by (7.30) we have

$$\lim_{h\to 0} \widetilde{\mathbb{E}} \int_h^t |B_\varepsilon(s, \widetilde{X}^h_{\pi_h(s)}) - B_\varepsilon(s, \widetilde{X}_s)|\mathrm{d}s = 0.$$

Below for notational convenience, we write $\widetilde{X}^\infty_t := \widetilde{X}_t$ and $\pi_0(s) := s$. For $h \in [0,1)$, we have

$$\begin{aligned}
&\widetilde{\mathbb{E}} \int_h^t |B_\varepsilon(s, \widetilde{X}^h_{\pi_h(s)}) - b(s, \widetilde{X}^h_{\pi_h(s)}, \rho_s(\widetilde{X}^h_{\pi_h(s)}))|\mathrm{d}s \\
&\leqslant \widetilde{\mathbb{E}} \int_h^t \mathbf{1}_{|\widetilde{X}^h_{\pi_h(s)}|\leqslant R} \Big| B_\varepsilon(s, \widetilde{X}^h_{\pi_h(s)}) - b(s, \widetilde{X}^h_{\pi_h(s)}, \rho_s(\widetilde{X}^h_{\pi_h(s)})) \Big| \mathrm{d}s \\
&\quad + 2\|b\|_\infty \int_h^t \widetilde{\mathbb{P}}\Big(|\widetilde{X}^h_{\pi_h(s)}| > R \Big) \mathrm{d}s =: I^h_R(\varepsilon) + J^h_R.
\end{aligned}$$

For $I^h_R(\varepsilon)$, by (7.27), (7.18) and Hölder's inequality with $p > 2d$ and $q = \frac{p}{p-1}$, we have

$$\begin{aligned}
I^h_R(\varepsilon) &= \int_h^t \int_{B_R} |B_\varepsilon(s,y) - b(s,y,\rho_s(y))| \rho^h_{\pi_h(s)}(y)\mathrm{d}y\mathrm{d}s \\
&\lesssim \int_h^t \int_{B_R} |B_\varepsilon(s,y) - b(s,y,\rho_s(y))| \int_{\mathbb{R}^d} g(4\pi_h(s), x-y)\nu_0(\mathrm{d}x)\mathrm{d}y\mathrm{d}s \\
&\lesssim \int_h^t \left(\int_{B_R} |B_\varepsilon(s,y) - b(s,y,\rho_s(y))|^p \mathrm{d}y \right)^{\frac{1}{p}} \\
&\qquad \times \left(\int_{B_R} \left| \int_{\mathbb{R}^d} g(4\pi_h(s), x-y)\nu_0(\mathrm{d}x) \right|^q \mathrm{d}y \right)^{\frac{1}{q}} \mathrm{d}s \\
&\lesssim \int_h^t \left(\int_{B_R} |B_\varepsilon(s,y) - b(s,y,\rho_s(y))|^p \mathrm{d}y \right)^{\frac{1}{p}} \pi_h(s)^{-\frac{d}{p}} \mathrm{d}s \\
&\lesssim \left(\int_h^t \left(\int_{B_R} |B_\varepsilon(s,y) - b(s,y,\rho_s(y))|^p \mathrm{d}y \right)^{\frac{2}{p}} \mathrm{d}s \right)^{\frac{1}{2}} \left(\int_h^t \pi_h(s)^{-\frac{2d}{p}} \mathrm{d}s \right)^{\frac{1}{2}} \\
&\lesssim \left(\int_0^t \left(\int_{B_R} |B_\varepsilon(s,y) - b(s,y,\rho_s(y))|^p \mathrm{d}y \right)^{\frac{2}{p}} \mathrm{d}s \right)^{\frac{1}{2}} \left(\int_0^t s^{-\frac{2d}{p}} \mathrm{d}s \right)^{\frac{1}{2}},
\end{aligned}$$

where the implicit constant in the above $\lesssim$ is independent of h, R and ε. Hence, for each $R > 0$, by the dominated convergence theorem, we obtain

$$\lim_{\varepsilon\to 0} \sup_{h\in[0,1)]} I^h_R(\varepsilon) = 0.$$

For J_R^h, by Chebyshev's inequality and (7.24) and since $\|b^h\|_\infty \leqslant \|b\|_\infty$, we have

$$\begin{aligned} J_R^h &= 2\|b\|_\infty \int_h^t \mathbb{P}(|X^h_{\pi_h(s)}| > R)\mathrm{d}s \\ &\leqslant 2\|b\|_\infty \int_0^t \mathbb{P}(|X_0| + s\|b\|_\infty + \sqrt{2}|W_{\pi_h(s)}| > R)\mathrm{d}s \\ &\leqslant 2\|b\|_\infty \Big(\int_0^t \mathbb{P}(|X_0| + s\|b\|_\infty > R/2)\mathrm{d}s + \int_0^t \frac{2\pi_h(s)}{(R/2)^2}\mathrm{d}s \Big), \end{aligned}$$

which converges to zero uniformly in h, as $R \to \infty$. Combining the above calculations, we obtain

$$\lim_{\varepsilon\to 0} \sup_{h\in[0,1)]} \widetilde{\mathbb{E}} \int_h^t |B_\varepsilon(s, \widetilde{X}^h_{\pi_h(s)}) - b(s, \widetilde{X}^h_{\pi_h(s)}, \rho_s(\widetilde{X}^h_{\pi_h(s)}))|\mathrm{d}s = 0.$$

Thus, (7.33) is proven and the existence of a solution to dDSDE (7.1) is obtained.

(Uniqueness) Let X_t and $\bar{X}_t$ be two solutions of dDSDE (7.1) defined on the same probability space and with the same initial value X_0, where X_0 has the distributional density $\rho_0 \in L^q(\mathbb{R}^d)$ with $q \in (d, \infty]$. Let $\rho_t(y)$ and $\bar{\rho}_t(y)$ be the distributional density of X_t and $\bar{X}_t$, respectively. Clearly, these are two solutions of the nonlinear Fokker-Planck equation (7.2) with the same initial value ρ_0. Consider the following linearized SDE:

$$\mathrm{d}X_t = B(t, X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \;\; X_0 = x,$$

where $B(t,x) := b(t, x, \rho_t(x))$. It is well known that $X_t(x)$ admits a density $p_x(t,y)$ with Gaussian type estimate: For some $\lambda, C > 0$, it holds that for all $t \in (0,T]$ and $x, y \in \mathbb{R}^d$,

$$p_x(t,y) \leqslant Cg(\lambda t, x - y).$$

Note that by (7.18) and Hölder's inequality,

$$\|\rho_t\|_q = \| \int_{\mathbb{R}^d} p_x(t,\cdot)\rho_0(x)\mathrm{d}x\|_q \leqslant \|\rho_0\|_q \tag{7.34}$$

and

$$\begin{aligned} \rho_t(y) &= \int_{\mathbb{R}^d} p_x(t,y)\rho_0(x)\mathrm{d}x \lesssim \int_{\mathbb{R}^d} g(\lambda t, x-y)\rho_0(x)\mathrm{d}x \\ &\leqslant \|g(\lambda t, \cdot)\|_{q/(q-1)}\|\rho_0\|_q \lesssim t^{-d/(2q)}\|\rho_0\|_q. \end{aligned} \tag{7.35}$$

Let

$$\Gamma_t := \rho_t - \bar{\rho}_t, \quad B_t := b(t,\cdot,\rho_t)\rho_t - b(t,\cdot,\bar{\rho}_t)\bar{\rho}_t.$$

Then, by (7.7) one sees that

$$\|B_t\|_q \leqslant \kappa\|\Gamma_t\|_q\|\rho_t\|_\infty + \|b\|_\infty\|\Gamma_t\|_q \lesssim t^{-d/(2q)}\|\Gamma_t\|_q. \tag{7.36}$$

Based on Duhamel's formula, we have

$$\Gamma_t = \int_0^t g(t-s) * (\mathrm{div} B_s)\mathrm{d}s = \int_0^t \nabla g(t-s) * B_s \mathrm{d}s.$$

In view of (7.14) and (7.36),

$$\begin{aligned}\|\Gamma_t\|_q &\lesssim \int_0^t \|\nabla g(t-s)\|_1 \|B_s\|_q \mathrm{d}s \\ &\lesssim \int_0^t (t-s)^{-1/2} s^{-d/(2q)} \|\Gamma_s\|_q \mathrm{d}s.\end{aligned}$$

We note that $1/2+d/(2q) < 1$. Thus by Lemma A.4, we get $\Gamma_t \equiv 0$ which implies $\rho_t = \bar{\rho}_t$. Now the pathwise uniqueness of SDE (7.1) follows by the well-known pathwise uniqueness for SDE (7.1) with bounded measurable drift $b(t, x, \rho_t(x))$ (cf. [101]).

(Convergence rate) Assume that $t > h$ in the following and recall

$$b^h(s,x) = \mathbf{1}_{\{s \geqslant h\}} b\Big(s, x, \rho^h_{\pi_h(s)}(x)\Big).$$

To unify the notation, we set $b^0(s,x) := b(s,x,\rho_s(s))$. For any $\varphi \in C_b^\infty$ and $t \in [0,T]$, consider the following backward heat equation

$$\partial_s u^t + \frac{1}{2}\Delta u^t = 0, \quad u(t) = \varphi, \quad s \in (0,t).$$

Then by Lemma 7.5, we know that $u^t(s) = g(t-s) * \varphi$ and by 7.14, for any $k \in \mathbb{N}_0$,

$$\|\nabla^k u^t(s)\|_\infty \lesssim (t-s)^{-\frac{k}{2}} \|\varphi\|_\infty. \tag{7.37}$$

By Itô's formula to $u^t(s, X_s)$ and $u^t(s, X^h_s)$, we have

$$\mathbb{E}\varphi(X_t) = \mathbb{E}u(0,X_0) + \mathbb{E}\int_0^t b^0(s,X_s)\cdot \nabla u^t(s,X_s)\mathrm{d}s$$

and

$$\begin{aligned}\mathbb{E}\varphi(X^h_t) &= \mathbb{E}u(0,X_0) + \mathbb{E}\int_0^t b^h(s, X^h_{\pi_h(s)})\cdot\nabla u^t(s,X^h_s)\mathrm{d}s \\ &= \mathbb{E}\varphi(X_t) + \mathbb{E}\int_0^t b^h(s, X^h_{\pi_h(s)})\cdot\nabla u^t(s,X^h_s)\mathrm{d}s \\ &\quad - \mathbb{E}\int_0^t b^0(s,X_s)\cdot\nabla u^t(s,X_s)\mathrm{d}s,\end{aligned}$$

which implies that

$$\begin{aligned}
&\Big|\mathbb{E}\varphi(X_t^h)-\mathbb{E}\varphi(X_t)\Big|\\
&\lesssim\left|\mathbb{E}\int_0^t\left((b^0\cdot\nabla u^t)(s,X_s^h)-(b^0\cdot\nabla u^t)(s,X_s)\right)\mathrm{d}s\right|\\
&\quad+\left|\mathbb{E}\int_0^t\Big(b(s,X_s^h,\rho_s^h(X_s^h)-b^0(s,X_s^h))\Big)\cdot\nabla u^t(s,X_s^h)\mathrm{d}s\right|\\
&\quad+\left|\mathbb{E}\int_0^t\Big(b(s,X_{\pi_h(s)}^h,\rho_s^h(X_{\pi_h(s)}^h))-b(s,X_s^h,\rho_s^h(X_s^h))\Big)\cdot\nabla u^t(s,X_s^h)\mathrm{d}s\right|\\
&\quad+\left|\mathbb{E}\int_0^t\Big(b^h(s,X_{\pi_h(s)}^h)-b(s,X_{\pi_h(s)}^h,\rho_s^h(X_{\pi_h(s)}^h))\Big)\cdot\nabla u^t(s,X_s^h)\mathrm{d}s\right|\\
&=:\mathscr{I}_1+\mathscr{I}_2+\mathscr{I}_3+\mathscr{I}_4.
\end{aligned}$$

Based on (7.37), one sees that

$$\begin{aligned}
\mathscr{I}_1&\lesssim\int_0^t\|\rho_s^h-\rho_s\|_1\|b^0\cdot\nabla u^t(s)\|_\infty\mathrm{d}s\\
&\lesssim\|\varphi\|_{L^\infty}\|b\|_{L^\infty}\int_0^t(t-s)^{-1/2}\int_0^t\|\rho_s^h-\rho_s\|_1\mathrm{d}s.
\end{aligned}$$

For $\mathscr{I}_2$, by (7.7), (7.37) and (7.35) we have

$$\begin{aligned}
\mathscr{I}_2&\leqslant\kappa\int_0^t\int_{\mathbb{R}^d}|(\rho_s^h-\rho_s)(x)|\nabla u^t(s,x)|\rho_s^h(x)\mathrm{d}x\mathrm{d}s\\
&\lesssim\|\varphi\|_\infty\int_0^t(t-s)^{-1/2}s^{-d/(2q)}\|\rho_s^h-\rho_s\|_1\mathrm{d}s.
\end{aligned}$$

For $\mathscr{I}_3$, we set $\eta^h(t,x):=b(t,x,\rho_t^h(x))$ and note that

$$\begin{aligned}
\mathscr{I}_3&\leqslant\left|\mathbb{E}\int_0^t\left((\eta^h\cdot\nabla u^t)(s,X_{\pi_h(s)}^h)-(\eta^h\cdot\nabla u^t)(s,X_s^h)\right)\mathrm{d}s\right|\\
&\quad+\left|\mathbb{E}\int_0^t\eta^h(s,X_{\pi_h(s)}^h)\cdot\left(\nabla u^t(s,X_s^h)-\nabla u^t(s,X_{\pi_h(s)}^h)\right)\mathrm{d}s\right|\\
&=:\mathscr{I}_{31}+\mathscr{I}_{32}.
\end{aligned}$$

Then,

$$\mathscr{I}_{31}=\left|\int_0^t\int_{\mathbb{R}^d}(\eta^h\cdot\nabla u^t)(s,x)\Big(\rho_{\pi_h(s)}^h(x)-\rho_s^h(x)\Big)\mathrm{d}x\mathrm{d}s\right|.$$

In light of Corollary 7.8-(iii), one sees that for any $p \in [1, q]$ and $0 < s < t \leqslant T$

$$\|\rho_s^h - \rho_t^h\|_p \lesssim |t-s|^{1/2} s^{-1/2} (\|g(4s)\|_1 + \|g(4t)\|_1) \|\rho_0\|_p \lesssim |t-s|^{1/2} s^{-1/2} \|\rho_0\|_p. \quad (7.38)$$

Thus, by (7.37), we have

$$\begin{aligned} \mathscr{I}_{31} &\leqslant \int_0^t \|\eta^h \cdot \nabla u^t(s)\|_\infty \|\rho_{\pi_h(s)}^h - \rho_s^h\|_1 \mathrm{d}x\mathrm{d}s \\ &\lesssim \|b\|_\infty \|\varphi\|_\infty \int_0^t (t-s)^{-1/2} (s - \pi_h(s))^{1/2} (\pi_h(s))^{-1/2} \\ &\lesssim h^{1/2} \|\varphi\|_\infty \int_h^t (t-s)^{-1/2} (s-h)^{-1/2} \mathrm{d}s \lesssim h^{1/2} \|\varphi\|_\infty. \end{aligned}$$

For $\mathscr{I}_{32}$, by (7.23) and (7.37), we have

$$\begin{aligned} \mathscr{I}_{32} &\lesssim \|b\|_\infty \int_{t-h}^t \|\nabla u^t(s)\|_{L^\infty} \mathrm{d}s + h\|b\|_\infty \int_0^{t-h} \left(\|\nabla^2 u^t(s)\|_\infty \|b\|_\infty + \|\nabla^3 u^t(s)\|_\infty\right) \mathrm{d}s \\ &\lesssim \|\varphi\|_\infty \left(\int_{t-h}^t (t-s)^{-1/2} \mathrm{d}s + h \int_0^{t-h} (t-s)^{-3/2} \mathrm{d}s \right) \lesssim h^{1/2} \|\varphi\|_\infty. \end{aligned}$$

For $\mathscr{I}_4$, we note that

$$\begin{aligned} \mathscr{I}_4 = &\left| \mathbb{E} \int_0^h b(s, X_{\pi_h(s)}^h, \rho_s^h(X_{\pi_h(s)}^h)) \cdot \nabla u^t(s, X_s^h) \mathrm{d}s \right| \\ &+ \left| \int_h^t \int_{\mathbb{R}^d} \left(b(s, x, \rho_{\pi_h(s)}^h(x)) - b(s, x, \rho_s^h(x)) \right) \cdot \nabla u^t(s, x) \rho_{\pi_h(s)}^h(x) \mathrm{d}x\mathrm{d}s \right| \\ \lesssim &\|b\|_\infty \|\varphi\|_\infty \int_0^h (t-s)^{-1/2} \mathrm{d}s \\ &+ \|\varphi\|_\infty \int_h^t \int_{\mathbb{R}^d} (t-s)^{-1/2} |\rho_{\pi_h(s)}^h(x) - \rho_s^h(x)| \rho_{\pi_h(s)}^h(x) \mathrm{d}x\mathrm{d}s, \end{aligned}$$

since (7.37) and (7.7). By Hölder's inequality and $q \geqslant 2$, it follows from (7.38) and (7.34) that

$$\begin{aligned} \mathscr{I}_4 &\lesssim h^{1/2} \|\varphi\|_\infty + \|\varphi\|_\infty \int_h^t (t-s)^{-1/2} (s - \pi_h(s))^{1/2} (\pi_h(s))^{-1/2} \mathrm{d}s \|\rho_0\|_p \|\rho_0\|_q \\ &\lesssim h^{1/2} \|\varphi\|_\infty, \end{aligned}$$

where $p = q/(q-1) \leqslant q$.

In summary, by taking supermum of φ we have

$$\|\rho_t - \rho_t^h\|_1 \lesssim \int_0^t (t-s)^{-1/2} s^{-d/(2q)} \|\rho_s - \rho_s^h\|_1 \mathrm{d}s + h^{1/2}.$$

By Lemma A.4, we complete the proof.

□

Appendix A

A.1 Technical lemmas

Lemma A.1. *[73, Lemma 3.4] Let* $(E, \|\cdot\|)$ *be a normed vector space,* $\tau, \eta \in [0,1]$ *with* $\tau > \eta$ *and* $X : (0,1] \to E$ *be a function satisfying*

$$\|X_t - X_s\| \leqslant C s^{-\eta}(t-s)^{\tau}, \quad \forall 0 < s \leqslant t \leqslant 1$$

for some constant C*. Then,*

$$\|X_t - X_s\| \leqslant C(1 - 2^{\eta-\tau})^{-1}(t-s)^{\tau-\eta}, \quad \forall 0 < s \leqslant t \leqslant 1.$$

Lemma A.2. *Let* $\alpha \in (0,1]$ *and* $\beta > 0$*.*

$$B(\alpha, \beta) := \int_0^1 s^{\beta-1}(1-s)^{\alpha-1} \mathrm{d}s \leqslant \frac{2}{\alpha}(\beta^{-1} + \beta^{-\alpha}).$$

Proof. When $\beta \leqslant 1$,

$$\begin{aligned} B(\alpha, \beta) &\leqslant (1/2)^{\alpha-1} \int_0^{1/2} s^{\beta-1} \mathrm{d}s + (1/2)^{\beta-1} \int_{1/2}^1 (1-s)^{\alpha-1} \mathrm{d}s \\ &\leqslant \frac{2}{\beta} + \frac{2}{\alpha} \leqslant \frac{2}{\alpha}(\beta^{-1} + 1) \leqslant \frac{2}{\alpha}(\beta^{-1} + \beta^{-\alpha}). \end{aligned}$$

When $\beta > 1$, one sees that $\beta^{-1} < 1$ and

$$\begin{aligned} B(\alpha, \beta) &\leqslant \beta^{1-\alpha} \int_0^{1-\beta^{-1}} s^{\beta-1} \mathrm{d}s + \int_{1-\beta^{-1}}^1 (1-s)^{\alpha-1} \mathrm{d}s \\ &\leqslant \frac{\beta^{1-\alpha}}{\beta} + \frac{\beta^{-1}}{\alpha} \leqslant \frac{2}{\alpha}\beta^{-\alpha}. \end{aligned}$$

This completes the proof. □

Lemma A.3. *Let $\beta > 1 > \alpha \geqslant 0$. There is a constant $C = C(\alpha, \beta)$ such that for all $t, r > 0$,*

$$\int_0^t (t+r-s)^{-\beta} s^{-\alpha} \mathrm{d}s \leqslant C r^{1-\beta} t^{-\alpha}.$$

Proof. When $r \geqslant t/2$, one sees that

$$\mathscr{I} := \int_0^t (t+r-s)^{-\beta} s^{-\alpha} \mathrm{d}s \leqslant r^{-\beta} \int_0^t s^{-\alpha} \mathrm{d}s \lesssim r^{-\beta} t^{1-\alpha} \lesssim r^{1-\beta} t^{-\alpha}.$$

We only consider the case $r < t/2$ in the following. We first make a decomposition

$$\begin{aligned} \mathscr{I} &= \left(\int_0^{t/2} + \int_{t/2}^{t-r} + \int_{t-r}^t \right) (t+r-s)^{-\beta} s^{-\alpha} \mathrm{d}s \\ &=: \mathscr{I}_1 + \mathscr{I}_2 + \mathscr{I}_3. \end{aligned}$$

For $\mathscr{I}_1$, one sees that

$$\mathscr{I}_1 \lesssim (t/2 + r)^{-\beta} \int_0^{t/2} s^{-\alpha} \mathrm{d}s \lesssim t^{-\beta+1-\alpha} \lesssim r^{1-\beta} t^{-\alpha}.$$

For $\mathscr{I}_2$, we have

$$\mathscr{I}_2 \lesssim t^{-\alpha} \int_{t/2}^{t-r} (t+r-s)^{-\beta} \mathrm{d}s \lesssim t^{-\alpha} \int_{t/2}^{t-r} (t-s)^{-\beta} \mathrm{d}s \lesssim t^{-\alpha} r^{1-\beta}.$$

Finialy, for $\mathscr{I}_3$, we have

$$\mathscr{I}_3 \lesssim t^{-\alpha} \int_{t-r}^t (t+r-s)^{-\beta} \mathrm{d}s \lesssim t^{-\alpha} r^{-\beta} \int_{t-r}^t \mathrm{d}s \lesssim t^{-\alpha} r^{1-\beta}$$

and complete the proof. □

A.2 Two types of Gronwall inequalities

Lemma A.4 (Gronwall's inequality of Volterra's type)**.** *Let $T > 0$ and $\alpha, \beta \in [0, 1)$ with $\alpha + \beta < 1$. Assume that $f, g : [0, T] \to \mathbb{R}_+$ be two measurable functions satisfying that for almost all $t \in (0, T]$,*

$$f(t) \leqslant g(t) + c_0 \int_0^t (t-s)^{-\alpha} s^{-\beta} f(s) \mathrm{d}s$$

with some constant c_0. Then there is a constant $C = C(T, \alpha, \beta, c_0) > 0$ such that for almost all $t \in (0, T]$,

$$f(t) \leqslant g(t) + C \int_0^t (t-s)^{-\alpha} s^{-\beta} g(s) \mathrm{d}s.$$

Proof. We define

$$r_1(t,s) := c_0(t-s)^{-\alpha}s^{-\beta}, \quad r_{n+1}(t,s) := c_0\int_s^t (t-u)^{-\alpha}u^{-\beta}r_n(u,s)\mathrm{d}u,$$

for any $0 \leqslant s < t \leqslant T$. By [112, Lemma 2.1 and 2.2], we only need to show that

$$\Big|\sum_{n=1}^{\infty} r_n(t,s)\Big| \leqslant C(t-s)^{-\alpha}s^{-\beta}$$

with some constant C. To this end, we give the following estimate by induction:

$$|r_n(t,s)| \leqslant c_0^{n+1}\big(\Pi_{k=0}^n a_k\big)(t-s)^{-\alpha}s^{-\beta}(t-s)^{n(1-\alpha-\beta)}, \quad n \geqslant 0 \tag{A.2.1}$$

where $a_0 := 1$ and for $k \geqslant 1$,

$$a_k := \frac{2}{1-\alpha}\left([k(1-\alpha-\beta)]^{-1} + [k(1-\alpha-\beta)]^{\alpha-1}\right).$$

We assume (A.2.1) holds for n. Then, by a change of variable one sees that

$$\begin{aligned}
|r_{n+1}(t,s)| &\leqslant c_0^{n+1}\big(\Pi_{k=0}^n a_k\big)s^{-\beta}\int_s^t (t-u)^{-\alpha}u^{-\beta}(u-s)^{-\alpha+n(1-\alpha-\beta)}\mathrm{d}s \\
&\leqslant c_0^{n+1}\big(\Pi_{k=0}^n a_k\big)s^{-\beta}\int_s^t (t-u)^{-\alpha}(u-s)^{-\alpha-\beta+n(1-\alpha-\beta)}\mathrm{d}s \\
&= c_0^{n+1}\big(\Pi_{k=0}^n a_k\big)s^{-\beta}(t-s)^{-\alpha+(n+1)(1-\alpha-\beta)}\int_0^1 (1-u)^{-\alpha}u^{-\alpha-\beta+n(1-\alpha-\beta)}\mathrm{d}s,
\end{aligned}$$

and by Lemma A.2 we have

$$\int_0^1 (1-u)^{-\alpha}u^{-\alpha-\beta+n(1-\alpha-\beta)}\mathrm{d}s \leqslant a_{n+1},$$

which implies (A.2.1) holds for $n = n+1$. By induction, we obtain (A.2.1). We note that

$$a_k \leqslant C_{\alpha,\beta}k^{\alpha-1},$$

with some constant $C_{\alpha,\beta}$ which only depends on α, β. Hence, we have

$$\begin{aligned}
\Big|\sum_{n=1}^{\infty} r_n(t,s)\Big| &\leqslant c_0\sum_{n=1}^{\infty}\frac{\big(c_0C_{\alpha,\beta}(t-s)^{1-\alpha-\beta}\big)^n}{(n!)^{1-\alpha}}(t-s)^{-\alpha}s^{-\beta} \\
&\leqslant c_0\sum_{n=1}^{\infty}\frac{\big(c_0C_{\alpha,\beta}T^{1-\alpha-\beta}\big)^n}{(n!)^{1-\alpha}}(t-s)^{-\alpha}s^{-\beta} \leqslant C_{\alpha,\beta,c_0,T}(t-s)^{-\alpha}s^{-\beta},
\end{aligned}$$

and complete the proof. ∎

Lemma A.5 (Stochastic Gronwall's inequality)**.** *[116, Lemma 2.8][91] Let $\xi(t)$ and $\eta(t)$ be two nonnegative continuous $\mathscr{F}_t$-adapted processes, A_t a continuous nondecreasing $\mathscr{F}_t$-adapted process with $A_0 = 0$, M_t a local martingale with $M_0 = 0$. Suppose that*

$$\xi(t) \leqslant \eta(t) + \int_0^t \xi(s)\mathrm{d}A_s + M_t, \quad a.s. \text{ for all } t \geqslant 0.$$

Then for any $0 < q < p < 1$ and $T > 0$, we have

$$\left[\mathbb{E}\left(\sup_{t\in[0,T]} \xi(t)\right)^q\right]^{1/q} \leqslant \left(\frac{p}{p-q}\right)\left(\mathbb{E}e^{\frac{pA_T}{1-p}}\right)^{\frac{1-p}{p}} \mathbb{E}\left(\sup_{t\in[0,T]} \eta(t)\right).$$

A.3 Schauder estimates for parabolic equations

In this part, we consider the following parabolic equation

$$\partial_t u = a_{ij}\partial_i\partial_j u + f, \quad u_0 = 0, \tag{A.3.1}$$

where $a = (a_{ij})$:$\mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ is a symmetric matrix-valued measurable function satisfying $(\mathbf{H}_a)$ with the Hölder parameter $\theta \in (0, 1)$.

The main result in this part is the following Schauder estimates.

Theorem A.6 (Schauder's estimate)**.** *Assume $(\mathbf{H}_a)$ holds with the Hölder parameter $\theta \in (0, 1)$. Let $T > 0$, $\alpha \in [0, 1)$, $f \in C((0, T]; \mathcal{C}^\theta)$ and u be a classical solution of (A.3.1). Then there is a constant $C = C(d, \theta, \|a\|_{\mathcal{C}^\theta}, T)$ independent of u and f such that*

$$\sup_{t\in[0,T]} (t^\alpha \|u(t)\|_{\mathcal{C}^{2+\theta}}) \leqslant C \sup_{t\in[0,T]} (t^\alpha \|f(t)\|_{\mathcal{C}^\theta}). \tag{A.3.2}$$

First of all, for any $z \in \mathbb{R}^d$, recall $P_t^z f := p_t^z * f$ with

$$p_t^z(x) := \frac{\exp(-\frac{\langle a(z)^{-1}x, x\rangle}{4t})}{\sqrt{(4\pi t)^d \det(a(z))}}.$$

In particular, we recall $P_t f = g_t * f$ with

$$g_t(x) = (4\pi t)^{-d/2} e^{-|x|^2/(4t)}.$$

Then we introduce the following characterization of Hölder space by using the heat semigroup P_t. We note that these types of characterizations are standard in harmonic analysis, see [96, Page 142; Section 4.2], for instance, for the characterization of Hölder space by using Poisson kernel. We adopt here a Gaussian type heat semigroup to replace the Poisson one since it appears naturally in Duhamel's formulal. Gaussian type heat semigroup is also used to characterize a Hölder-Dini continuous function in [106, Section 4.3].

Lemma A.7. *Let $\theta \in (0,1)$. There is a constant $C = C(d,\theta)$ such that for any $f \in \mathcal{C}^\theta$ and $t \in (0,1)$,*

$$\|\nabla^k P_t f\|_\infty \leqslant C t^{-\frac{k-\theta}{2}} \|f\|_{\mathcal{C}^\theta}, \quad k = 1,2. \tag{A.3.3}$$

On the contrary, if

$$c_f := \sum_{k=1}^{2} \sup_{t\in(0,1)} \left(t^{\frac{k-\theta}{2}} \|\nabla^k P_t f\|_\infty \right) < \infty,$$

then there is a constant $C = C(d,\theta)$ such that for any $x \in \mathbb{R}^d$ and $|y| < 1$

$$|f(x+y) - f(x)| \leqslant C c_f |y|^\theta. \tag{A.3.4}$$

Proof. For (A.3.3), by the fact

$$\int_{\mathbb{R}^d} \nabla^k g_t(x) \mathrm{d}x = 0, \quad k = 1,2,$$

and the scaling $g_t(x) = t^{-d/2} g_1(t^{-1/2}x)$, we have

$$\begin{aligned} |\nabla^k P_t f(x)| &= \left| \int_{\mathbb{R}^d} \nabla^k g_t(y) \left(f(x-y) - f(x) \right) \mathrm{d}y \right| \\ &\leqslant \int_{\mathbb{R}^d} |\nabla^k g_t(y)| |y|^\theta \mathrm{d}y \|f\|_{\mathcal{C}^\theta} \\ &\leqslant t^{-\frac{k-\theta}{2}} \int_{\mathbb{R}^d} |\nabla^k g_1(y)| |y|^\theta \mathrm{d}y \|f\|_{\mathcal{C}^\theta} \end{aligned}$$

and obtain (A.3.3). For (A.3.4), given any $t \in (0,1)$, we make the following decomposition

$$\begin{aligned} |f(x+y) - f(x)| &\leqslant |P_t f(x+y) - P_t f(x)| + |P_t f(x+y) - f(x+y)| + |P_t f(x) - f(x)| \\ &\leqslant |y| \|\nabla P_t f\|_\infty + \left| \int_0^t \partial_s P_s f(x+y) \mathrm{d}s \right| + \left| \int_0^t \partial_s P_s f(x) \mathrm{d}s \right|. \end{aligned}$$

We note that $\partial_s P_s f = \Delta P_s f$ and have

$$|f(x+y) - f(x)| \leqslant c_f \left(|y| t^{-\frac{1-\theta}{2}} + 2 \int_0^t s^{-\frac{2-\theta}{2}} \mathrm{d}s \right) \leqslant c_f \left(|y| t^{-\frac{1-\theta}{2}} + 2 t^{\frac{\theta}{2}} \right).$$

By taking $t = |y|^2$, we complete the proof. □

Proof of Theorem A.6. We set

$$u^z(t,x) := u(t,x+z), \quad a^z(x) := a(x+z) \quad \text{and} \quad f^z(t,x) := f(t,x+z).$$

Then we have

$$\partial_t u^z = a_{ij}(z)\partial_i\partial_j u^z + (a^z_{ij} - a_{ij}(z))\partial_i\partial_j u^z + f^z,$$

which by Duhamel's formula implies that

$$u^z(t) = \int_0^t P^z_{t-s}\left((a^z_{ij} - a_{ij}(z))\partial_i\partial_j u^z(s)\right)\mathrm{d}s + \int_0^t P^z_{t-s} f^z(s)\mathrm{d}s.$$

For $r \in (0,1)$ and $k = 1,2$, taking operator $\nabla^k P_r \nabla^2$ on the both sides we have

$$\begin{aligned}\nabla^k P_r \nabla^2 u^z(t,x) = &\int_0^t \nabla^k P_r \nabla^2 P^z_{t-s}\left((a^z_{ij} - a_{ij}(z))\partial_i\partial_j u^z(s)\right)(x)\mathrm{d}s\\ &+ \int_0^t \nabla^k P_r \nabla^2 P^z_{t-s} f^z(s)(x)\mathrm{d}s.\end{aligned}$$

Letting $x = 0$, one sees that

$$\begin{aligned}&\nabla^k P_r \nabla^2 u(t,z) = \nabla^k P_r \nabla^2 u^z(t,0)\\ =&\int_0^t \nabla^k P_r \nabla^2 P^z_{t-s}\left((a^z_{ij} - a_{ij}(z))\partial_i\partial_j u^z(s)\right)(0)\mathrm{d}s\\ &+ \int_0^t \nabla^k P_r \nabla^2 P^z_{t-s} f^z(s)(0)\mathrm{d}s\\ =&\int_0^t \int_{\mathbb{R}^d} (\nabla^k g_r * \nabla^2 p^z_{t-s})(x)\left((a_{ij}(x+z) - a_{ij}(z))\partial_i\partial_j u(s,x+z)\right)\mathrm{d}x\mathrm{d}s\\ &+ \int_0^t \int_{\mathbb{R}^d} (\nabla^k g_r * \nabla^2 p^z_{t-s})(x)(f(s,x+z) - f(s,z))\mathrm{d}x\mathrm{d}s,\end{aligned}$$

where we used the fact that

$$\int_{\mathbb{R}^d} (\nabla^k g_r * \nabla^2 p^z_{t-s})(x) f(s,z)\mathrm{d}x = 0.$$

Hence, we have

$$|\nabla^k P_r \nabla^2 u(t,z)| \lesssim \int_0^t \int_{\mathbb{R}^d} |\nabla^k g_r * \nabla^2 p^z_{t-s}|(x)|x|^\theta s^{-\alpha}\mathrm{d}x\mathrm{d}s(\mathscr{U} + \mathscr{F}),$$

where

$$\mathscr{U} := \sup_{t\in[0,T]}\left(t^\alpha \|\nabla^2 u(t)\|_\infty\right), \quad \mathscr{F} := \sup_{t\in[0,T]}\left(t^\alpha \|f(t)\|_{\mathcal{C}^\theta}\right).$$

We note that

$$
\begin{aligned}
&|\nabla^k g_r * \nabla^2 p^z_{t-s}|(x)\\
&\leqslant \left(\int_{\mathbb{R}^d} |\nabla^{k+2} g_r(x-y)| p^z_{t-s}(y)\mathrm{d}y\right) \wedge \left(\int_{\mathbb{R}^d} g_r(x-y)|\nabla^{k+2} p^z_{t-s}(y)|\mathrm{d}y\right)\\
&\lesssim \left(r^{-\frac{k+2}{2}} \wedge (t-s)^{-\frac{k+2}{2}}\right) \int_{\mathbb{R}^d} g_{cr}(x-y) g_{c(t-s)}(y)\mathrm{d}y\\
&\lesssim \left(r^{-\frac{k+2}{2}} \wedge (t-s)^{-\frac{k+2}{2}}\right) g_{c(t+r-s)}(x) \lesssim (t+r-s)^{-\frac{k+2}{2}} g_{c(t+r-s)}(x)
\end{aligned}
$$

with some constant $c>1$, by the elementary estimate. Thus, we have

$$
\begin{aligned}
|\nabla^k P_r \nabla^2 u(t,z)| &\lesssim \int_0^t (t+r-s)^{-\frac{k+2}{2}} s^{-\alpha} \int_{\mathbb{R}^d} |x|^\theta g_{c(t+r-s)}(x)\mathrm{d}x\mathrm{d}s(\mathscr{U}+\mathscr{F})\\
&\lesssim \int_0^t (t+r-s)^{-\frac{k+2-\theta}{2}} s^{-\alpha}\mathrm{d}s(\mathscr{U}+\mathscr{F})\\
&\lesssim r^{-\frac{k-\theta}{2}} t^{-\alpha}(\mathscr{U}+\mathscr{F}),
\end{aligned}
$$

which is from Lemma A.3. In view of (A.3.4), we have

$$
\sup_{t\in[0,T],|y|\leqslant 1}\left(t^\alpha \frac{|\nabla^2 u(t,x+y)-\nabla^2 u(t,x)|}{|y|^\theta}\right) \lesssim (\mathscr{U}+\mathscr{F}).
$$

Based on the following maximal principle

$$
\|u(t)\|_\infty \leqslant \int_0^t \|f(s)\|_\infty \mathrm{d}s \lesssim \mathscr{F}, \tag{A.3.5}
$$

we have for some constant $C_0 = C_0(d,\theta,T)$,

$$
\sup_{t\in[0,T]} (t^\alpha \|u(t)\|_{\mathcal{C}^{2+\theta}}) \leqslant C_0(\mathscr{U}+\mathscr{F}).
$$

By the following interpolation inequality (see [66, Theorem 3.2.1] for instance)

$$
\|\nabla^2 f\|_\infty \leqslant \|f\|_{\mathcal{C}^{2+\theta}}^{\frac{2}{2+\theta}} \|f\|_\infty^{\frac{\theta}{2+\theta}} \leqslant \frac{1}{2C_0}\|f\|_{\mathcal{C}^{2+\theta}} + (2C_0)^{\frac{2(2+\theta)}{\theta^2}}\|f\|_\infty, \quad \forall f\in\mathcal{C}^{2+\theta},
$$

where we used Young's inequality in the second step, and maximal principle (A.3.5), we have

$$
\sup_{t\in[0,T]} (t^\alpha \|u(t)\|_{\mathcal{C}^{2+\theta}}) \leqslant 2C_0\mathscr{F} + C \sup_{t\in[0,T]} (t^\alpha \|u(t)\|_\infty) \leqslant C\mathscr{F}
$$

with some constant $C>0$ and complete the proof. □

It should be noted that this method of proof is also used in [49, 23, 47, 46]. But therein they used Littlewood-Paley decomposition and Besov space. In the present proof, we only use semigroup P_t to characterize the Hölder space.

Bibliography

[1] Bakhtin V. and Kifer Y.: Diffusion approximation for slow motion in fully coupled averaging. *Probab. Theory Related Fields* **129** (2004), 157–181.

[2] Bao J. and Huang X.: Approximations of Mckean-Vlasov Stochastic Differential Equations with irregular coefficients. *J. Theoret. Probab.* (2021), 1-29.

[3] Barbu V. and Röckner M.: Probabilistic representation for solutions to nonlinear Fokker-Planck equations. *SIAM J.Math.Anal.* **50** (2018), 4246-4260.

[4] Barbu V. and Röckner M.: Uniqueness for nonlinear Fokker-Planck equations and weak uniqueness for McKean-Vlasov SDEs. *Stoch. PDE Anal. Comput.* **9** (2019), 702-713.

[5] Barbu V. and Röckner M.: From nonlinear Fokker-Planck equations to solutions of distribution dependent SDE. *Ann. Probab.* **48** (2020), 1902-1920.

[6] Barbu V. and Röckner M.: Solutions for nonlinear Fokker-Planck equations with measures as initial data and McKean-Vlasov equations. *J. Functional Anal.* **280** (2021), 1-35.

[7] Barbu V. and Röckner M.: Uniqueness for nonlinear Fokker-Planck equations and for McKean-Vlasov SDEs: The degenerate case. arXiv:2203.00122 (2022).

[8] Barbu V. and Röckner M.: Correction to: Uniqueness for nonlinear Fokker-Planck equations and weak uniqueness for McKean-Vlasov SDEs. *Stoch. Partial Differ. Equ. Anal. Comput.* **11** (2023), 426-431.

[9] Bencheikh O. and Jourdain B.: Convergence in total variation of the Euler-Maruyama scheme applied to diffusion processes with measurable drift coefficient and additive noise. *SIAM J. Numer. Anal.* **60** (2022), 1701-1740.

[10] Benedek A. and Panzone R.: The spaces L^p, with mixed norm. *Duke Math. J.* **28** (1961), 301-324.

[11] Bogoliubov N. N. and Mitropolsky Y. A.: *Asymptotic Methods in the Theory of Non-linear Oscillations.* Translated from the second revised Russian edition International Monographs on Advanced Mathematics and Physics Hindustan Publishing Corp., Delhi, Gordon and Breach Science Publishers, Inc., New York 1961 x+537 pp.

[12] Bolley F. and Villani C.: Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Annales de la Faculté des sciences de Toulouse: Mathématiques* **14** (2005), 331-352.

[13] Bossy M. and Talay D.: Convergence rate for the approximation of the limit law of weakly interacting particles: application to the Burgers equation. *Ann. Appl. Probab.* **6** (1996), 818-861.

[14] Budhiraja A. and Dupuis P.: Analysis and approximation of rare events. Representations and weak convergence methods. Probability Theory and Stochastic Modelling, **94**. *Springer, New York*, 2019. xix+574 pp.

[15] Carmona R. and Delarue F.: Probabilistic theory of mean field games with applications. I. Mean field FBSDEs, control, and games. Probability Theory and Stochastic Modeling, **83**. *Springer, Cham,* 2018. xxv+713 pp.

[16] Carrillo J.A., Choi Y.-P. and Hauray M.: The derivation of swarming models: mean-field limit and Wasserstein distances. (English summary) *Collective dynamics from bacteria to crowds*, 1-46, CISM Courses and Lect., 553, Springer, Vienna, 2014.

[17] Carrillo J. A., Choi Y.-P. and Salem S.: Propagation of chaos for the Vlasov-Poisson-Fokker-Planck equation with a polynomial cut-off. *Commun. Contemp. Math.* **21** (2019), 1850039.

[18] Cerrai S.: A Khasminskii type averaging principle for stochastic reaction-diffusion equations. *Ann. Appl. Probab.* **19** (2009), 899-948.

[19] Cerrai S. and Freidlin M.: Averaging principle for a class of stochastic reaction-diffusion equations. *Probab. Theory Related Fields* **144** (2009), 137-177.

[20] Cerrai S. and Lunardi A.: Averaging principle for nonautonomous slow-fast systems of stochastic reaction-diffusion equations: the almost periodic case. *SIAM J. Math. Anal.* **49** (2017), 2843-2884.

[21] Chaintron L.P. and Diez A.: Propagation of chaos: a review of models, methods and applications. II. Applications. *arXiv:2106.14812* (2021).

[22] Chaintron L.P. and Diez A.: Propagation of chaos: a review of models, methods and applications. I. Models and methods. *arXiv:2203.00446* (2022).

[23] Chen Z.-Q., Hao Z. and Zhang X.: Hölder regularity and gradient estimates for SDEs driven by cylindrical α-stable processes. *Electron. J. Probab.* **25** (2020), 1-23.

[24] Chen Z.-Q., Hu E., Xie L. and Zhang X.: Heat kernels for non-symmetric diffusion operators with jumps. *J. Differential Equations* **263** (2017), 6576-6634.

[25] Cheng M., Hao Z. and Röckner M.: Strong and weak convergence for averaging principle of DDSDE with singular drift. *arXiv:2207.12108* (2022).

[26] Cheng M. and Liu Z.: The second Bogolyubov theorem and global averaging principle for SPDEs with monotone coefficients. *arXiv:2109.00371* (2021).

[27] Choi Y.-P. and Salem S.: Propagation of chaos for aggregation equations with no-flux boundary conditions and sharp sensing zones. *Math. Models Methods Appl. Sci.* **28** (2018), 223-258.

[28] Dareiotis K. and Gerencsér M.: On the regularisation of the noise for the Euler-Maruyama scheme with irregular drift. *Electron. J. Probab.* **25** (2020), 1-18.

[29] Del Moral P. and Miclo L.: Genealogies and increasing propagation of chaos for Feynman-Kac and genetic models. *Ann. Appl. Probab.* **11** (2001), 1166-1198.

[30] Duan J. and Wang W.: Effective dynamics of stochastic partial differential equations. Elsevier Insights. *Elsevier, Amsterdam*, 2014. xii+270 pp.

[31] Dupuis P. and Ellis R. S.: A weak convergence approach to the theory of large deviations. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. *John Wiley & Sons, Inc., New York,* 1997. xviii+479 pp.

[32] Figalli A.: Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. *J. Funct. Anal.* **254** (2008), 109-153.

[33] Flandoli F., Leimbach M. and Olivera C.: Uniform convergence of proliferating particles to the FKPP equation.*J. Math. Anal. Appl.* **473** (2019), 27-52.

[34] Flandoli F. and Leocata M.: A particle system approach to aggregation phenomena. *J. Appl. Probab.* **56** (2019), 282-306.

[35] Flandoli F., Olivera C. and Simon M.: Uniform approximation of 2 dimensional Navier-Stokes equation by stochastic interacting particle systems. *SIAM J. Math. Anal.* **52** (2020), 5339-5362.

[36] Freidlin M. I. and Wentzell A. D.: Random Perturbations of Dynamical Systems. Translated from the Russian by Joseph Szücs. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 260. *Springer-Verlag, New York*, 1984. viii+326 pp.

[37] Fontbona J. and Martinez M.: Paths clustering and an existence result for stochastic vortex systems. *J. Stat. Phys.* **128** (2007), 699-719.

[38] Fournier, N., Hauray, M. and Mischler, S.: Propagation of chaos for the 2D viscous vortex model. *J. Eur. Math. Soc.* **16** (2014), 1423-1466.

[39] Glose F.: On the dynamics of large particle systems in the mean field limit. (English summary) *Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity*, 1-144, Lect. Notes Appl. Math. Mech., 3, Springer, [Cham], 2016.

[40] Graham, C., Kurtz, T., Méléard, S., Protter, P., Pulvirenti, M. and Talay, D.: Probabilistic models for nonlinear partial differential equations. Lectures given at the 1st Session and Summer School held in Montecatini Terme, May 22-30, 1995. Edited by D. Talay and L. Tubaro. Lecture Notes in Mathematics, 1627. Fondazione CIME/CIME Foundation Subseries. *Springer-Verlag, Berlin; Centro Internazionale Matematico Estivo (C.I.M.E.), Florence,* 1996. x+301 pp.

[41] Hairer M. and Li X.-M.: Averaging dynamics driven by fractional Brownian motion. *Ann. Probab.* **48** (2020), 1826-1860.

[42] Han Y.: Solving McKean-Vlasov SDEs via relative entropy. *arXiv:2204.05709* (2022).

[43] Hao Z., Peng X. and Zhang X.: Hörmander's hypoelliptic theorem for nonlocal operators. *J. Theoret. Probab.* **34** (2021), 1870-1916.

[44] Hao Z., Röckner, M. and Zhang X.: Euler scheme for density dependent stochastic differential equations. *J. Differential Equations* **274** (2021), 996-1014.

[45] Hao Z., Röckner M., and Zhang X.: Strong convergence of propagation of chaos for McKean-Vlasov SDEs with singular interactions. *arXiv:2204.07952* (2022).

[46] Hao Z., Röckner M., and Zhang X.: Second order fractional mean-field SDEs with singular kernels and measure initial datas. *arXiv:2302.04392* (2023).

[47] Hao Z., Wang Z. and Wu M.: Schauder's estimates for nonlocal equations with singular Lévy measures. *arXiv:2002.09887* (2020).

[48] Hao Z. and Wu M.: Well-posedness of density dependent SDE driven by α-stable process with Hölder drifts. *arXiv:2112.06757* (2021).

[49] Hao Z., Wu M. and Zhang X.: Schauder's estimate for nonlocal kinetic equations and applications. *J. Math. Pures Appl.* **140** (2020), 139-184.

[50] Hao Z., Zhang X., Zhu R. and Zhu X.: Singular kinetic equations and applications. *arXiv:2108.05042* (2021).

[51] Hauray M. and Mischler, S.: On Kac's chaos and related problems. *J. Funct. Anal.* **266** (2014), 6055-6157.

[52] Ho K.: Strong maximal operator on mixed-norm spaces. *Ann. Univ. Ferrara.* **62** (2016), 275-291.

[53] Hoeksema J., Holding T., Maurelli M. and Tse O.: Large deviations for singularly interacting diffusions. *arXiv: 2002.01295* (2020).

[54] Hong W., Li S. and Liu W.: Strong convergence rates in averaging principle for slow-fast McKean-Vlasov SPDEs. *J. Differential Equations* **316** (2022), 94-135.

[55] Jabin P.-E.: A review of mean field limits for Vlasov equations. *Kinet. Relat. Models* **7** (2014), 661-711.

[56] Jabin P.-E. and Wang Z.: Mean field limit and propagation of chaos for Vlasov system with bounded forces. *J. Func. Anal.* **271** (2016), 3588-3627.

[57] Jabin P.-E. and Wang Z.: Mean field limit for stochastic particle systems. *Active particles. Vol. 1. Advances in theory, models, and applications*, 379-402, Model. Simul. Sci. Eng. Technol., Birkhäuser/Springer, Cham, 2017.

[58] Jabin P.-E. and Wang Z.: Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels. *Invent. Math.* **214** (2018), 523-591.

[59] Jabir J.-F., Talay D. and Tomašević M.: Mean-field limit of a particle approximation of the one-dimensional parabolic-parabolic Keller-Segel model without smoothing. *Electron. Commun. Probab.* **23** (2018), Paper No. 84, 14 pp.

[60] Jourdain B.: Diffusions with a nonlinear irregular drift coefficient and probabilistic interpretation of generalized Burgers' equations. *ESAIM Probab. Statist.* **1** (1995/97), 339-355.

[61] Jourdain B. and Méléard S.: Propagation of chaos and fluctuations for a moderate model with smooth initial data. *Ann. Inst. H. Poincaré Probab. Statist.* **34** (1998), 727-766.

[62] Kac M.: Foundations of kinetic theory. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954-1955, vol. III, pp. 171-197. University of California Press, Berkeley-Los Angeles, Calif.*, 1956.

[63] Khasminshii R. Z.: On the principle of averaging the Itô's stochastic differential equations. (Russian) *Kybernetika (Prague)* **4** (1968), 260-279.

[64] Kifer Y.: Some recent advances in averaging. (English summary) *Modern dynamical systems and applications, 385-403, Cambridge Univ. Press, Cambridge*, 2004.

[65] Krylov N. and Bogolyubov N.: Introduction to Non-Linear Mechanics. Annals of Mathematics Studies, No. 11 *Princeton University Press, Princeton, N. J.*, 1943. iii+105 pp.

[66] Krylov N. V.: Lectures on elliptic and parabolic equations in Hölder spaces. Graduate Studies in Mathematics, 12. *American Mathematical Society, Providence, RI*, 1996. xii+164 pp.

[67] Krylov N. V. and Röckner M.: Strong solutions of stochastic equations with singular time dependent drift. *Probab. Theory Relat. Fields* **131** (2005), 154-196.

[68] Kurtz T.G.: Martingale problems for conditional distributions of Markov processes. *Electron. J. Probab.* **3** (1998), 1-29.

[69] Lacker D.: On a strong form of propagation of chaos for McKean-Vlasov equations. *Electron. Commun. Probab.* **23** (2018), paper no. 45, 11 pp.

[70] Lacker D.: Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions. *arXiv:2105.02983* (2021).

[71] Lemaire V. and Menozzi S.: On some Non Asymptotic Bounds for the Euler Scheme. *Electron. J. Probab.* **15** (2010), 1645-1681.

[72] Li J. and Min H.: Weak solutions of mean-field stochastic differential equations and application to zero-sum stochastic differential games. *SIAM J. Control Optim.* **54** (2016), 1826-1858.

[73] Lê K. and Ling C.: Taming singular stochastic differential equations: A numerical method. *arXiv:2110.01343* (2021).

[74] Ling C. and Xie L.: Strong Solutions of Stochastic Differential Equations with Coefficients in Mixed-Norm Spaces. *Potential Anal.* **57** (2022), 227-241.

[75] Maslowski B., Seidler J. and Vrkoč I.: An averaging principle for stochastic evolution equations. II. *Math. Bohem.* **116** (1991), 191-224.

[76] McKean H. P.: A class of Markov processes associated with nonlinear parabolic equations. *Proc Nat. Acad Sci USA*, **56**(1966), 1907-1911.

[77] McKean H. P. Jr.: Propagation of chaos for a class of non-linear parabolic equations. 1967 Stochastic Differential Equations (*Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967*) pp. 41-57 *Air Force Office Sci. Res., Arlington, Va.*

[78] Menozzi S. and Zhang X.: Heat kernel of supercritical nonlocal operators with unbounded drifts. *J. Éc. polytech. Math.* **9** (2022), 537-579.

[79] Mischler S. and Mouhot C.: Kac's program in kinetic theory. *Invent. Math.* **193** (2013), 1-147.

[80] Mishura Y. S. and Veretennikov A. Yu.: Existence and uniqueness theorems for solutions of McKean-Vlasov stochastic equations. Theory Probab. Math. Statist. **103** (2020), 59-101.

[81] Oelschläger K.: A law of large numbers for moderately interacting diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **69** (1985), 279-322.

[82] Olivera C., Richard A and Tomašević M.: Quantitative particle approximation of nonlinear Fokker-Planck equations with singular kernel. To appear in *Ann. Sc. Norm. Super. Pisa Cl. Sci* (2021).

[83] Osada H.: A stochastic differential equation arising from the vortex problem. *Proc. Japan Acad. Ser. A Math. Sci.* **61** (1985), 333-336.

[84] Osada H.: Propagation of chaos for the two-dimensional Navier-Stokes equation. *Probabilistic methods in mathematical physics (Katata/Kyoto, 1985)*, 303-334, *Academic Press, Boston, MA*, 1987.

[85] Pei B., Inahama Y. and Xu Y.: Averaging principle for fast-slow system driven by mixed fractional Brownian rough path. *J. Differential Equations* **301** (2021), 202-235.

[86] Revuz D., and Yor, M.: Continuous martingales and Brownian motion. Third edition. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 293. *Springer-Verlag, Berlin*, 1999. xiv+602 pp.

[87] Röckner M., Sun X., and Xie L.: Strong and weak convergence in the averaging principle for SDEs with Hölder coefficients. *arXiv:1907.09256* (2019).

[88] Röckner M., Sun X., and Xie Y.: Strong convergence order for slow-fast McKean-Vlasov stochastic differential equations. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** (2021), 547-576.

[89] Röckner M. and Zhang X.: Well-posedness of distribution dependent SDEs with singular drifts. *Bernoulli* **27** (2021), 1131-1158.

[90] Sanders J. A. and Verhulst F.: Averaging methods in nonlinear dynamical systems. Second edition. Applied Mathematical Sciences, 59. *Springer, New York*, 2007. xxii+431 pp.

[91] Scheutzow M.: A stochastic Gronwall lemma. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **16** (2013), 1350019, 4 pp.

[92] Shen G., Xiang J. and Wu J.: Averaging principle for distribution dependent stochastic differential equations driven by fractional Brownian motion and standard Brownian motion. *J. Differential Equations* **321** (2022), 381-414.

[93] Shkolnikov M.: Large systems of diffusions interacting through their ranks. *Stochastic Process. Appl.* **122** (2012), 1730-1747.

[94] Simon M. and Olivera C.: Non-local conservation law from stochastic particle systems. *J. Dynam. Differential Equations* **30** (2018), 1661-1682.

[95] Spohn H.: Large scales dynamics of interacting particles. *Springer Verlag, New York*, 1991.

[96] Stein E. M.: Singular integrals and differentiability property of functions. Princeton Mathematical Series, No. 30 *Princeton University Press, Princeton, N.J.* 1970 xiv+290 pp.

[97] Stroock D.W. and Varadhan S. R.: Multidimensional diffusion processes. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 233. *Springer-Verlag, Berlin-New York*, 1979. xii+338 pp.

[98] Sznitman A.S.: Topics in propagation of chaos. *École d'Été de Prob. de Saint-Flour XIX-1989*, 165-251, Lecture Notes in Math., 1464, *Springer, Berlin*, 1991.

[99] Tomašević M.: Propagation of chaos for stochastic particle systems with singular mean-field interaction of L^q-L^p type. *hal-03086253* (2020).

[100] Trevisan D.: Well-posedness of multidimensional diffusion process with weakly differentiable coefficients. *Electron. J. Probab.* **21** (2016).

[101] Veretennikov A. Yu.: Strong solutions of stochastic differential equations. (Russian. English summary) *Teor. Veroyatnost. i Primenen.* **24** (1979), 348-360.

[102] Veretennikov A. Yu.: On large deviations in the averaging principle for SDEs with a "full dependence". *Ann. Probab.* **27** (1999), 284-296.

[103] Vlasov A. A.: The vibrational properties of an electron gas. *Soviet Physics Uspekhi* **10** (1968), 721.

[104] Wang F.-Y.: Distribution dependent SDEs for Landau type equations. *Stochastic Process. Appl.* **128** (2018), 595-621.

[105] Wang F.-Y.: Singular density dependent stochastic differential equations. *J. Differential Equations* **361** (2023), 562-589.

[106] Wang F.-Y. and Zhang X.: Degenerate SDE with Hölder-Dini drift and non-Lipschitz noise coefficient. *SIAM J. Math. Anal.* **48** (2016), 2189-2222.

[107] Wang W. and Roberts A. J.: Average and deviation for slow-fast stochastic partial differential equations. *J. Differential Equations* **253** (2012), 1265-1286.

[108] Wang Z.: Mean Field Limit for Stochastic Particle Systems with Singular Forces. Thesis (Ph.D.)-University of Maryland, College Park. 2017. 154 pp.

[109] Xia P., Xie L., Zhang X. and Zhao G.: $L^q(L^p)$-theory of stochastic differential equations. *Stochastic Processes Appl.* **130** (2020), 5188-5211.

[110] Xie L. and Zhang X.: Ergodicity of stochastic differential equations with jumps and singular coefficients. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** (2020), 175-229.

[111] Xue J.: Noncollision singularities in a planar two-center-two-body problem. Thesis (Ph.D.)-University of Maryland, College Park. 2013. 133 pp.

[112] Zhang X.: Stochastic Volterra equations in Banach spaces and stochastic partial differential equation. *J. Funct. Anal.* **258** (2010), 1361-1425.

[113] Zhang, X.: Stochastic homeomorphism flows of SDEs with singular drifts and Sobolev diffusion coefficients. *Electron. J. Probab.* **16** (2011), 1096-1116.

[114] Zhang X.: Stochastic differential equations with Sobolev coefficients and applications.*Ann. Appl. Probab.* **26** (2016), 2697-2732.

[115] Zhang X.: A discretized version of Krylov's estimate and its applications. *Electron. J. Probab.*, **24** (2019), Paper No. 131, 17 pp.

[116] Zhang X. and Zhao G.: Singular Brownian Diffusion Processes. *Commun. Math. Stat.* **6** (2018), 533-581.

[117] Zhao G.: On Distribution depend SDEs with singular drifts. *arXiv:2003.04829v3* (2020).

[118] Zvonkin, A.K.: A transformation of the phase space of a diffusion process that removes the drift.(Russian) *Mat. Sb.* **93 (135)** (1974), 129-149.