THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 1 Capitole (UT1 Capitole)*

Présentée et soutenue le *13/12/2024* **par :**

Anh-Dung LE

Well-posedness of McKean-Vlasov Stochastic Differential Equations with Density-dependent Drift

JURY

Francesco Russo	ENSTA Paris	Rapporteur
Elena Issoglio	University of Torino	Rapporteur
Benjamin Jourdain	ENPC Paris	Examinateur
Fabienne Comte	Université Paris Descartes	Examinateur
Stéphane Villeneuve	Toulouse School of Economics	Directeur de Thèse
Sébastien Gadat	Toulouse School of Economics	Co-directeur de Thèse

École doctorale et spécialité :

MITT : Domaine Mathématiques : Mathématiques appliquées

Unité de Recherche :

Toulouse School of Economics – Research (UMR TSE-R)

Président du Jury :

Francesco Russo

Acknowledgment

First and foremost, I would like express my deep gratitude to my supervisors for being with me during my ups and downs. I am grateful to Professor Sébastien Gadat for giving me the opportunity to do my PhD at TSE, as well as for his generosity and empathy in giving me the freedom to pursue this project. I am grateful to Professor Stéphane Villeneuve for his guidance on how to approach a research problem, as well as for his encouragement whenever my research got stuck. Meeting both of you has been my biggest luck, and I remain appreciative of it.

It is my great honor that - Professors Francesco Russo, Elena Issoglio, Benjamin Jourdain and Fabienne Comte - accepted to be members of the jury. I am grateful that they spent their precious time reading this manuscript.

I am grateful to Professor Feng-Yu Wang for patiently addressing all questions about his work [Wan23c], whose techniques are essential to my first paper. I am grateful to Professor Haïm Brezis for writing the gentle textbook [Bre11], which makes it possible for me to start doing research. I am grateful to Thibault Caillet for kindly addressing my questions about the monograph [AGS08].

I am grateful to Professor Jérôme Bolte who helped extend my PhD contract. The additional time and funding allowed me to write this manuscript in peace. I am grateful to Professor Christine Thomas-Agnan for her care during my Master. Her care made my first time in France pleasant. I am grateful to Professor Cuong Le Van and Dr. Quang Vu, whose dedicated support and letters of recommendation were instrumental in securing me Eiffel scholarship. The scholarship helped me have a comfortable life during my Master.

I am grateful to my family - mẹ Hoài, bố Sơn và Nhật - for believing in me and standing by me through all the difficulties in this long journey. You are my everything. I would like to thank my friends - Bảo và Tuấn Anh - for chats and trips full of joy. You two guys have made my life more fulfilling. I would like to thank all kind-hearted people whom I have been fortunate to meet and who have been a great solace to me.

Finally, I am grateful to the financial support from ANR MaSDOL (No. ANR-19-CE23-0017); Air Force Office of Scientific Research, Air Force Material Command, USAF (No. FA9550-19-7026); and ANR Chess (No. ANR-17-EURE-0010).

Abstract

The aim of this thesis is to study a class of McKean-Vlasov stochastic differential equations (*MV-SDEs*) whose drift depends pointwisely on marginal density of the process. In existing literature, the space of probability measures is endowed with total variation metric. In the thesis, we go in a different direction where we use Wasserstein metric for distributions. The structure and main results of the thesis are summarized below.

In Chapter 1, we give motivation for the research problem. After that, we briefly survey related literature.

In Chapter 2, we recall related definitions and results that will be used in subsequent chapters. The first part of the chapter is about gradient flows in Wasserstein space. The second part is about regularity of marginal density for classical SDEs.

In Chapter 3, we carefully apply the framework in [AGS08] to study Langevin dynamic for a function whose gradient is not necessarily locally Lipschitz. First, we prove well-posedness of the associated Fokker-Planck equation. If, in addition, a log-Sobolev inequality is assumed, then we obtain exponential rate of convergence to the stationary distribution.

In Chapter 4, we study well-posedness of MV-SDEs whose drift is density-dependent and locally integrable in space-time. Our approach is by mollifying argument. First, we prove existence of a strong solution. If some more technical conditions are assumed, we obtain strong uniqueness of a solution.

In Chapter 5, we study Euler–Maruyama scheme for a special case of SDE in Chapter 4 where the noise is constant and the drift is bounded. First, we obtain Hölder regularity of the scheme. Second, we derive convergence rate in weighted total variation distance.

The thesis is based on the two preprints:

1. Anh-Dung Le. *Well-posedness of McKean-Vlasov SDEs with density-dependent drift.* 2024. under review.

2. Anh-Dung Le. *Convergence rate of Euler scheme for McKean-Vlasov SDEs with density-dependent drift.* 2024. under review.

Contents

Notations

We adopt the following conventions:

	Generic elements and operations
$\rightharpoonup$	weak convergence on $\mathcal{P}(\mathbb{R}^d)$
$\overset{*}{\rightharpoonup}$	weak-$*$ convergence on $\mathcal{P}(\mathbb{R}^d)$
∞	plus infinity
xy	matrix product of $x \in \mathbb{R}^d \otimes \mathbb{R}^m$ and $y \in \mathbb{R}^m \otimes \mathbb{R}^n$
$\langle x, y \rangle$	Frobenius inner product of $x, y \in \mathbb{R}^d \otimes \mathbb{R}^m$
$\lvert x \rvert$	Frobenius norm of $x \in \mathbb{R}^d \otimes \mathbb{R}^m$
$x \vee y$	maximum between $x, y \in \mathbb{R}$
$x \wedge y$	minimum between $x, y \in \mathbb{R}$
x^+	maximum between $x \in \mathbb{R}$ and 0
x^*	Hölder conjugate of $x \in [1, \infty]$, i.e., $1 = \frac{1}{x} + \frac{1}{x^*}$
$\lceil x \rceil$	the smallest integer greater than or equal to $x \in \mathbb{R}$
$[\![m, n]\!]$	$\{m, m+1, \dots, n-1, n\}$ for integers $m \leq n$
id	identity function on $\mathbb{R}^d$
I_d	identity matrix in $\mathbb{R}^d \otimes \mathbb{R}^d$
$f_\sharp \mu$	push-forward of measure μ through function f
π^i	projection of $\mathbb{R}^d \times \mathbb{R}^d$ onto i-th coordinate
$\mathrm{d}x$	Lebesgue measure on $\mathbb{R}^d$
∇	gradient operator
∇^2	Hessian matrix
Δ	Laplacian operator
div	divergence operator
∂_{x_i}	derivative in spatial x_i direction
∂_t	derivative in time
$x^\top$	transpose of $x \in \mathbb{R}^d \otimes \mathbb{R}^m$
$\operatorname{tr} x$	trace of $x \in \mathbb{R}^d \otimes \mathbb{R}^d$
$M_p(\mu)$	p-th moment of probability measure μ

	Sets and spaces
$\mathbb{R}_+$	set of non-negative real numbers including 0
$\mathbb{N}$	set of natural numbers excluding 0
$\mathbb{R}^d \otimes \mathbb{R}^m$	space of real matrices of size $d \times m$
$\mathcal{B}(\mathbb{R}^d)$	Borel σ-algebra on $\mathbb{R}^d$
$\mathcal{P}(\mathbb{R}^d)$	space of Borel probability measures on $\mathbb{R}^d$
$L^0(\mathbb{R}^d)$	space of real-valued measurable functions on $\mathbb{R}^d$
$L^0_+(\mathbb{R}^d)$	subset of $L^0(\mathbb{R}^d)$ that consists of non-negative functions
$L^0_b(\mathbb{R}^d)$	subset of $L^0(\mathbb{R}^d)$ that consists of bounded functions
$L^p(\mathbb{R}^d)$	Lebesgue space of real-valued p-integrable functions on $\mathbb{R}^d$
$\\|\cdot\\|_{L^p}$	norm of $L^p(\mathbb{R}^d)$
$C_b(\mathbb{R}^d)$	space of real-valued continuous bounded functions on $\mathbb{R}^d$
$C^\alpha_b(\mathbb{R}^d)$	space of real-valued α-Hölder bounded functions on $\mathbb{R}^d$
$C_c(\mathbb{R}^d)$	space of real-valued continuous functions on $\mathbb{R}^d$ with compact support
$C^\infty_c(\mathbb{R}^d)$	subspace $C_c(\mathbb{R}^d)$ that consists of smooth functions

	Abbreviation
w.r.t	with respect to
i.f.f	if and only if
e.g.	for example
a.e.	almost everywhere
a.s.	almost surely
PDE	partial differential equation
SDE	stochastic differential equation
PoC	propagation of chaos
MV-SDE	McKean-Vlasov SDE
MP	martingale problem
MCT	monotone convergence theorem
DCT	dominated convergence theorem
WLOG	without loss of generality
m-BM	m-dimensional Brownian motion
AF	admissible filtration
PS	probability space

Chapter 1

Introduction

1.1 Motivation

1.1.1 Three perspectives on simulated annealing

A fundamental question in optimization is how to minimize a differentiable function $V : \mathbb{R}^d \to \mathbb{R}$ that is bounded from below. This becomes challenging when V is non-convex and possibly has many local minimizers. One approach to find a global minimizer of V is *simulated annealing*, a technique rooted in the seminal work [Met+53]. The authors in [Met+53] attempted to calculate properties of a substance composed of interacting molecules. The guiding principle is that the system of molecules will move to a configuration resulting in lower potential energy. For mathematical studies of simulated annealing, we refer to [HKS89; GM91; Mic92; RRT17; Xu+18; GGZ22], the monographs [Van+87; AK89] and references therein.

In time-continuous formulation, simulated annealing is described by a solution (X_t) of the stochastic differential equation (*SDE*)

$$\mathrm{d}X_t = -\gamma_t \nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t. \tag{1.1}$$

Above, (B_t) is a Brownian motion and $(\gamma_t) \subset \mathbb{R}_+$ is a sequence such that $\gamma_t \to \infty$ as $t \to \infty$. We assume that V has a unique global minimizer $\bar{x}$. Under appropriate assumptions on V and (γ_t), (X_t) converges to $\bar{x}$ in some suitable sense. Let μ_t be the distribution of X_t. By Itô's lemma, (μ_t) satisfies the following Fokker-Planck equation in distributional sense

$$\partial_t \mu_t = \gamma_t \operatorname{div}(\mu_t \nabla V) + \Delta \mu_t. \tag{1.2}$$

We assume that each μ_t admits a density denoted by ρ_t. Let $\mathcal{M}$ be the space of densities on $\mathbb{R}^d$ with finite second moment. We endow $\mathcal{M}$ with the Wasserstein metric W_2. For $\gamma > 0$, we define $\mathcal{V}, \mathcal{H}, \mathcal{U}_\gamma : \mathcal{M} \to (-\infty, \infty]$ by

$$\mathcal{V}(\varrho) := \int_{\mathbb{R}^d} V \varrho \,\mathrm{d}x, \tag{1.3}$$

$$\mathcal{H}(\varrho) := \int_{\mathbb{R}^d} \{\varrho \ln \varrho + 1 - \varrho\}\,\mathrm{d}x, \tag{1.4}$$

$$\mathcal{U}_\gamma(\varrho) := \gamma \mathcal{V}(\varrho) + \mathcal{H}(\varrho).$$

The maps $\mathcal{V}, \mathcal{H}$ are called *potential energy* and *internal energy* respectively. Otto's formalism [Ott96; JKO98; Ott01] allows one to interpret $\mathcal{M}$ as a Riemannian manifold and (1.2) as a gradient flow in $\mathcal{M}$, i.e.,

$$\text{“}\partial_t \rho_t = -\operatorname{grad} \mathcal{U}_{\gamma_t}[\rho_t]\text{”}. \tag{1.5}$$

Each of above three forms provides distinct advantages, depending on the question of study. First, (1.1) provides practitioners with operational algorithms through time discretization, such as

Euler–Maruyama scheme. Second, (1.2) entitles researchers to partial differential equation (*PDE*) techniques, particularly for exploring questions related to the regularity of (μ_t). Third, (1.5) opens a fresh perspective on gradient flows in the space of probability measures. In particular, different forms of $\mathfrak{U}_\gamma$ likely lead to different PDEs.

1.1.2 Otto calculus

Next we follow [FG21, Section 4.2] to motivate Otto's formal interpretation, and there will be no attempt at rigorous justification. Given a curve $(\rho_t)_{t\in[0,1]} \subset \mathcal{M}$ and a vector field $v : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ (we set $v_t(x) := v(t,x)$) such that $v_t \in L^2(\rho_t)$. We say that (ρ_t, v_t) satisfies the *continuity equation*

$$\partial_t \rho_t + \operatorname{div}(v_t \rho_t) = 0 \tag{1.6}$$

in distributional sense if it holds for $(f,g) \in C_c^\infty(0,1) \times C_c^\infty(\mathbb{R}^d)$ that

$$\int_0^1 \int_{\mathbb{R}^d} f'(t) g(x) \rho_t(x) \,\mathrm{d}x \,\mathrm{d}t + \int_0^1 \int_{\mathbb{R}^d} f(t) \langle \nabla g(x), v_t(x) \rangle \rho_t(x) \,\mathrm{d}x \,\mathrm{d}t = 0,$$

or equivalently, if it holds for $g \in C_c^\infty(\mathbb{R}^d)$ that

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} g \rho_t \,\mathrm{d}x = \int_{\mathbb{R}^d} \langle \nabla g, v_t \rangle \rho_t \,\mathrm{d}x \quad \text{for a.e.} \quad t \geq 0.$$

By Benamou–Brenier formula [BB00, Proposition 1.1],

$$\begin{aligned} &W_2^2(\varrho_0, \varrho_1) \\ &= \min \left\{ \int_0^1 \|v_t\|_{L^2(\rho_t)}^2 \,\mathrm{d}t : (\rho_t, v_t) \text{ satisfies (1.6) and } (\rho_0, \rho_1) = (\varrho_0, \varrho_1) \right\} \\ &= \min_{\rho_t} \left\{ \int_0^1 \min_{v_t} \left\{ \int_{\mathbb{R}^d} |v_t|^2 \rho_t \,\mathrm{d}x : (\rho_t, v_t) \text{ satisfies (1.6)} \right\} \mathrm{d}t : (\rho_0, \rho_1) = (\varrho_0, \varrho_1) \right\}. \end{aligned} \tag{1.7}$$

Given a Riemannian manifold (M, g), its Riemannian metric d_M between two points $x, y \in M$ can be defined as

$$\mathrm{d}_M^2(x,y) = \inf \left\{ \int_0^1 \|\dot{\gamma}_t\|_{\gamma_t}^2 \,\mathrm{d}t : (\gamma_t)_{t\in[0,1]} \subset M \text{ such that } (\gamma_0, \gamma_1) = (x,y) \right\}. \tag{1.8}$$

Comparing (1.7) and (1.8), it is natural to define *Wasserstein norm* of the derivative $\partial_t \rho_t$ at ρ_t as

$$\|\partial_t \rho_t\|_{\rho_t}^2 := \min_{v_t} \left\{ \int_{\mathbb{R}^d} |v_t|^2 \rho_t \,\mathrm{d}x : (\rho_t, v_t) \text{ satisfies (1.6)} \right\}. \tag{1.9}$$

It turns out that the minimizer $\bar{v}_t$ of (1.9) is of the form $\bar{v}_t = \nabla \psi_t$ for some $\psi_t : \mathbb{R}^d \to \mathbb{R}$. If the density ρ_t is regular enough (say, positive and smooth), then ψ_t is the unique (up to a constant) solution to $\partial_t \rho_t + \operatorname{div}(\rho_t \nabla \psi_t) = 0$. So we can define

$$\|\partial_t \rho_t\|_{\rho_t}^2 := \int_{\mathbb{R}^d} |\nabla \psi_t|^2 \rho_t \,\mathrm{d}x.$$

If the curve $(\rho_t)_{t\in[0,1]}$ is regular enough, then $\int_{\mathbb{R}^d} \partial_t \rho_t \,\mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} \rho_t \,\mathrm{d}x = 0$. As such, it is natural to define the *Wasserstein tangent space* $T_\varrho \mathcal{M}$ at $\varrho \in \mathcal{M}$ by

$$T_\varrho \mathcal{M} := \left\{ h : \mathbb{R}^d \to \mathbb{R} \,\middle|\, \int_{\mathbb{R}^d} h \,\mathrm{d}x = 0 \right\} / \sim.$$

Above, the equivalence relation $\sim$ is defined via the equation $h + \operatorname{div}(\varrho \nabla \psi) = 0$. Accordingly, the *Wasserstein inner product* on $T_\varrho\mathcal{M}$ is defined as

$$\langle h_1, h_2\rangle_\varrho := \int_{\mathbb{R}^d} \langle \nabla\psi_1, \nabla\psi_2\rangle \varrho \, \mathrm{d}x \quad \text{where} \quad h_i + \mathrm{div}(\varrho\nabla\psi_i) = 0.$$

We have just given a meaning to the LHS of (1.5). Next we do so for the RHS of (1.5). For a functional $\mathcal{U} : \mathcal{M} \to \mathbb{R}$, its *Wasserstein gradient* at $\varrho \in \mathcal{M}$, denoted by $\mathrm{grad}\,\mathcal{U}[\varrho]$, is the unique element in $T_\varrho\mathcal{M}$ (if it exists) such that

$$\left\langle \mathrm{grad}\,\mathcal{U}[\varrho], \frac{\partial\rho_t}{\partial t}\Big|_{t=0} \right\rangle_\varrho = \frac{\partial\mathcal{U}[\rho_t]}{\partial t}\Big|_{t=0}$$

for any smooth curve $\rho : (-t_0, t_0) \to \mathcal{M}$ with $\rho_0 = \varrho$.

To obtain a more explicit formula for $\mathrm{grad}\,\mathcal{U}[\varrho]$, we introduce the *first variation* of $\mathcal{U}$ at $\varrho \in \mathcal{M}$, which is denoted by $\mathrm{var}\,\mathcal{U}[\varrho]$ and defined as the function in $L^2(\mathbb{R}^d)$ such that

$$\frac{\partial\mathcal{U}[\rho_t]}{\partial t}\Big|_{t=0} = \int_{\mathbb{R}^d} \mathrm{var}\,\mathcal{U}[\varrho] \times \frac{\partial\rho_t}{\partial t}\Big|_{t=0} \mathrm{d}x \tag{1.10}$$

for any smooth curve $\rho : (-t_0, t_0) \to \mathcal{M}$ with $\rho_0 = \varrho$. Let ψ be the solution to $\frac{\partial\rho_t}{\partial t}\big|_{t=0} + \mathrm{div}(\varrho\nabla\psi) = 0$. Then

$$\begin{aligned}\left\langle \mathrm{grad}\,\mathcal{U}[\varrho], \frac{\partial\rho_t}{\partial t}\Big|_{t=0} \right\rangle_\varrho &= -\int_{\mathbb{R}^d} \mathrm{var}\,\mathcal{U}[\varrho] \times \mathrm{div}(\varrho\nabla\psi)\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} \langle \nabla(\mathrm{var}\,\mathcal{U}[\varrho]), \nabla\psi\rangle \varrho\,\mathrm{d}x.\end{aligned}$$

By definition of $\langle\cdot,\cdot\rangle_\varrho$, we deduce that

$$\mathrm{grad}\,\mathcal{U}[\varrho] = -\,\mathrm{div}(\varrho\nabla(\mathrm{var}\,\mathcal{U}[\varrho])). \tag{1.11}$$

Example 1.1. Consider a smooth curve $(\rho_t)_{t\in[0,1]} \subset \mathcal{M}$. First,

$$\begin{aligned}\frac{\partial\mathcal{V}[\rho_t]}{\partial t}\Big|_{t=0} &= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \int_{\mathbb{R}^d} V\rho_t\,\mathrm{d}x \quad \text{by (1.3)} \\ &= \int_{\mathbb{R}^d} V \times \frac{\partial\rho_t}{\partial t}\Big|_{t=0} \mathrm{d}x.\end{aligned}$$

Second,

$$\begin{aligned}\frac{\partial\mathcal{H}[\rho_t]}{\partial t}\Big|_{t=0} &= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \int_{\mathbb{R}^d} \{\rho_t \ln\rho_t + 1 - \rho_t\}\,\mathrm{d}x \quad \text{by (1.4)} \\ &= \int_{\mathbb{R}^d} \ln\rho_t \times \frac{\partial\rho_t}{\partial t}\Big|_{t=0} \mathrm{d}x.\end{aligned}$$

By (1.10), $\mathrm{var}\,\mathcal{V}[\varrho] = V$ and $\mathrm{var}\,\mathcal{H}[\varrho] = \ln\varrho$. By (1.11),

$$\begin{aligned}\mathrm{grad}\,\mathcal{U}_\gamma[\varrho] &= -\,\mathrm{div}(\varrho\nabla(\gamma V + \ln\varrho)) \\ &= -\gamma\,\mathrm{div}(\varrho\nabla V) - \Delta\varrho.\end{aligned}$$

Then we recover (1.2) from (1.5).

1.1.3 A new class of density-dependent SDEs

We consider a more general internal energy $\mathcal{H}:\mathcal{M}\to(-\infty,\infty]$ defined as $\mathcal{H}(\varrho):=\int\varphi\circ\varrho\,\mathrm{d}x$ for some twice differentiable function $\varphi:\mathbb{R}_+\to\mathbb{R}_+$. Then $\operatorname{var}\mathcal{H}[\varrho]=\varphi'\circ\varrho$ and thus

$$\operatorname{grad}\mathcal{U}_\gamma[\varrho]=-\operatorname{div}(\varrho\nabla(\gamma V+\varphi'\circ\varrho)).$$

The corresponding version of (1.5) is

$$\partial_t\rho_t=\operatorname{div}(\rho_t\{\gamma_t\nabla V+\nabla(\varphi'\circ\rho_t)\}). \tag{1.12}$$

Next we follow [BMV24, Section 2.2] to find an SDE whose associated Fokker-Planck equation is (1.12). We define a function $\alpha:(0,\infty)\to\mathbb{R}_+$ by

$$\alpha(r):=\frac{1}{r}\int_0^r s\varphi''(s)\,\mathrm{d}s.$$

By formal integration by parts, it holds for $\varrho\in\mathcal{M}$ and $g\in C_c^\infty(\mathbb{R}^d)$ that

$$\begin{aligned}\int_{\mathbb{R}^d}\langle\nabla g,\nabla(\varphi'\circ\varrho)\rangle\varrho\,\mathrm{d}x&=\int_{\mathbb{R}^d}\langle\nabla g,\varrho\varphi''(\varrho)\nabla\varrho\rangle\,\mathrm{d}x\\&=\int_{\mathbb{R}^d}\langle\nabla g,\nabla(\varrho\alpha(\varrho))\rangle\,\mathrm{d}x\\&=-\int_{\mathbb{R}^d}\alpha(\varrho)\{\Delta g\}\varrho\,\mathrm{d}x.\end{aligned}$$

Then (1.12) is equivalent to

$$\partial_t\rho_t=\partial_{x_i}\{\gamma_t(\nabla V)^i\rho_t\}+\partial_{x_i}\partial_{x_i}\{\alpha(\rho_t)\rho_t\}.$$

Above, we adapt Einstein summation convention. Also, $(\nabla V)^i$ is the i-th coordinate of ∇V. This suggests that the SDE for which we are looking is of the form

$$\mathrm{d}X_t=-\gamma_t\nabla V(X_t)\,\mathrm{d}t+\sqrt{2\alpha(\rho_t)}\,\mathrm{d}B_t. \tag{1.13}$$

SDE (1.13) is interesting due to its pointwise dependence on marginal density. Our initial goal is to study the *existence* of a solution to (1.13) when φ is one of the power-like functions $\{\varphi_m:m\in(0,\frac{1}{2})\}$ introduced in [BMV24, Section 3.1]. For $m\in(0,\frac{1}{2})$, the function $\varphi_m:\mathbb{R}_+\to\mathbb{R}_+$ is defined as

$$\varphi_m(r):=\begin{cases}\frac{r^m-1+m(1-r)}{m(m-1)} & \text{if}\quad r\in[0,1],\\ \frac{(r-1)^2}{2} & \text{if}\quad r\in(1,\infty).\end{cases}$$

Then

$$\varphi_m''(r)=\begin{cases}r^{m-2} & \text{if}\quad r\in(0,1],\\ 1 & \text{if}\quad r\in(1,\infty),\end{cases}\quad\text{and thus}\quad\alpha_m(r)=\begin{cases}\frac{r^{m-1}}{m} & \text{if}\quad r\in(0,1],\\ \frac{mr^2+2-m}{2mr} & \text{if}\quad r\in(1,\infty).\end{cases}$$

We define $\beta_m:\mathbb{R}_+\to\mathbb{R}_+$ by

$$\beta_m(r):=r\alpha_m(r)=\begin{cases}\frac{r^m}{m} & \text{if}\quad r\in[0,1],\\ \frac{mr^2+2-m}{2m} & \text{if}\quad r\in(1,\infty).\end{cases}$$

Then α_m and β_m are continuously differentiable on $(0,\infty)$ with

$$\alpha_m'(r) = \begin{cases} \frac{m-1}{m} r^{m-2} & \text{if} \quad r \in (0,1], \\ \frac{1}{2} + \frac{m-2}{2mr^2} & \text{if} \quad r \in (1,\infty), \end{cases} \quad \text{and} \quad \beta_m'(r) = \begin{cases} r^{m-1} & \text{if} \quad r \in [0,1], \\ r & \text{if} \quad r \in (1,\infty). \end{cases}$$

Our focus is on α_m, so we assume that V and ρ_0 are regular enough. By [BR21b, Theorem 4.1], (1.13) has weak uniqueness. Unfortunately, we could not make progress on the existence of a solution. The subjective reason is because papers on (1.13) rely heavily on PDE techniques for which we were unable to get into. Let $\sigma_m \coloneqq \sqrt{\alpha_m}$. One objective reason is the possible blow-up $\alpha_m(\varrho(x)) \to \infty$ as $|x| \to \infty$ for a density ϱ. This makes it hard to establish stability estimates for fixed-point and compactness arguments. In below example, the fact that $\lim_{r\downarrow 0} \alpha(r) < \infty$ is crucial.

Example 1.2. The following reasoning is used in the proof of [BJ19, Lemma 2.4]. For $v \in L^2((0,T); H^1(\mathbb{R}^d))$, it holds for a.e. $t \in (0,T)$ that

$$(u_1,u_2) \mapsto \mathcal{L}_t(u_1,u_2) \coloneqq \frac{1}{2} \int_{\mathbb{R}^d} \alpha(v(t,x)) \langle \nabla u_1(x), \nabla u_2(x) \rangle \, \mathrm{d}x$$

is continuous on $H^1(\mathbb{R}^d) \times H^1(\mathbb{R}^d)$ because

$$\mathcal{L}_t(u_1,u_2) \le \frac{1}{2} \sup\{\alpha(r) : 0 \le r \le \|v\|_\infty\} \|\nabla u_1\|_{L^2(\mathbb{R}^d)} \|\nabla u_2\|_{L^2(\mathbb{R}^d)}.$$

First, σ_m is neither bounded nor Lipschitz. Second, σ_m is not twice differentiable at 1. Let us elaborate on why existing literature does not cover the case of our α_m. The relationship between (α,β,σ) in the the general setting is the same as that between $(\alpha_m,\beta_m,\sigma_m)$. In [JM98], σ is Lipschitz. In [BR20; Gru23], α is bounded. In [BR23a], β is bounded. In [Gru24], α is locally Lipschitz at 0. In [BRR10; BRR11; BCR13; BR18; BR21a], there exists a constant $\eta \ge 1$ such that $\beta(r) \lesssim r^\eta$ and thus $\limsup_{r\downarrow 0} \alpha(r) < \infty$. As such, we turn our attention to a more manageable case where the density-dependence is on the drift. Even in this simpler setting, we only consider the situation where $b(t,x,r)$ is locally integrable in (t,x) uniformly in r.

1.2 Literature review

An SDE is a differential equation that incorporates noise into classical ordinary differential equation (*ODE*), and gives rise to a solution which is itself a stochastic process. The basic form of an SDE is

$$\mathrm{d}X_t = b(t,X_t)\,\mathrm{d}t + \sigma(t,X_t)\,\mathrm{d}B_t, \tag{1.14}$$

where

$$\begin{aligned} b &: \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d, \\ \sigma &: \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d, \end{aligned}$$

are measurable and $(B_t, t \ge 0)$ is a Brownian motion. An extension of (1.14) is to take into account the marginal distribution of the process, i.e.,

$$\begin{cases} \mathrm{d}X_t = b(t,X_t,\mu_t)\,\mathrm{d}t + \sigma(t,X_t,\mu_t)\,\mathrm{d}B_t, \\ \mu_t \text{ is the distribution of } X_t, \end{cases} \tag{1.15}$$

where

$$b : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d,$$
$$\sigma : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^d.$$

SDE (1.15) is called MV-SDE or distribution-dependent SDE. The study of (1.15) started with Kac's seminal paper [Kac56] in which he gave the first rigorous definitions of *chaos* and *propagation of chaos* (*PoC*). A decade later, McKean introduced in [McK66] a class of Markov processes that satisfy this property. Recently, (1.15) has drawn great attention from research community because it appears as the limit of various interacting particle systems in biology [NPT10; MT14; Deg18], mean-field games [CD18; Car19], and data science [MMN18; SS20b; RV22] where the notion of chaos plays a central role. We refer to [CD22a; CD22b] for a more comprehensive review of this aspect. By Itô's lemma, the curve (μ_t) of probability measures in (1.15) satisfies (in distributional sense) the nonlinear PDE

$$\partial_t \mu_t = -\partial_{x_i}\{b^i(t,\cdot,\mu_t)\mu_t\} + \frac{1}{2}\partial_{x_i}\partial_{x_j}\{\sigma^{ik}\sigma^{jk}(t,\cdot,\mu_t)\mu_t\},$$

which makes (1.15) interesting to both probabilists and PDE analysts.

In the thesis, we consider the following generalization of (1.15), i.e.,

$$\begin{cases} X_t = b(t, X_t, \ell_t(X_t), \mu_t)\,\mathrm{d}t + \sigma(t, X_t, \mu_t)\,\mathrm{d}B_t, \\ \mu_t \text{ is the distribution of } X_t, \text{ and } \ell_t \text{ is the density of } X_t, \end{cases} \tag{1.16}$$

where

$$b : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d,$$
$$\sigma : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^d.$$

The drift in (1.16) depends pointwisely on marginal density of the process. SDE (1.16) is called MV-SDE with density-dependent drift. In particular, $b(t,x,r,\mu)$ is not necessarily continuous in (t,x). Unlike classical MV-SDEs, the map $\mu \mapsto b(t,x,\frac{\mathrm{d}\mu}{\mathrm{d}x}(x),\mu)$ is not even continuous w.r.t the weak topology of $\mathcal{P}(\mathbb{R}^d)$ (see e.g. [SH24]). This discontinuity presents additional challenges in the analysis of these equations. Below are two situations where (1.16) arises.

Example 1.3 (Nonlinear Filtering)**.** Consider the nonlinear filtering problem in continuous time

$$\text{signal process}: \quad \mathrm{d}X_t = M(X_t)\,\mathrm{d}t + \mathrm{d}V_t,$$
$$\text{observation process}: \quad \mathrm{d}Z_t = h(X_t)\,\mathrm{d}t + \mathrm{d}W_t.$$

Above, $M : \mathbb{R}^d \to \mathbb{R}^d$ and $h : \mathbb{R}^d \to \mathbb{R}$ are regular enough; V_t and W_t are independent Brownian motions of dimension d and 1 respectively; and the filtration $\mathcal{Z}_t := \sigma(Z_s, s \le t)$. We fix $T \in (0,\infty)$ and let $\mathbb{T} := [0,T]$. Let $n \ge 1$ be an integer and $\varepsilon := \frac{T}{n}$ be the step size. Let $t_k := k\varepsilon$ for $k = 0,\dots,n$. In practice, we have access to $(Z_{t_k})_{k=0}^n$ instead of $(Z_t, t \in \mathbb{T})$. We define a random function v from $\mathbb{T} \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ to $\mathbb{R}^d$ by

$$v_t(x,\mu) := \frac{1}{\omega_d}\int_{\mathbb{R}^d} \frac{y-x}{|y-x|^d}\{m_t(y) - \bar{m}_t\}\,\mathrm{d}\mu(y),$$
$$m_t(y) := h(y)\frac{Z_{t_{k+1}} - Z_{t_k}}{\varepsilon} - \frac{1}{2}h^2(y) \quad \text{for} \quad t \in [t_k, t_{k+1}),$$
$$\bar{m}_t := \int_{\mathbb{R}^d} m_t(y)\,\mathrm{d}\mu(y),$$

where ω_d is the surface area of the unit ball in $\mathbb{R}^d$. Crisan-Xiong [CX10] proposed to approximate

$(X_t, t \in \mathbb{T})$ by a solution $(Y_t, t \in \mathbb{T})$ to the following SDE (see e.g. [PRS21, Equation (1.5)])

$$\mathrm{d}Y_t = M(Y_t)\,\mathrm{d}t + \mathrm{d}V_t + \frac{v_t(Y_t, \mu_t)}{\ell_t(Y_t)}\,\mathrm{d}t,$$

where μ_t is the conditional distribution of Y_t given $\mathcal{Z}_t$; and ℓ_t is the density of μ_t.

Example 1.4 (Generative Modeling)**.** Assume that an unknown distribution μ_0 of X_0 is perturbed through a mechanism modeled by an SDE

$$\mathrm{d}X_t = f(t, X_t)\,\mathrm{d}t + g(t)\,\mathrm{d}B_t,$$

where $f : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d$ and $g : \mathbb{T} \to \mathbb{R}$ are regular enough. We are given the functions f, g and the distribution μ_T of X_T. We are aimed at sampling from μ_0. We define $Y_t := X_{T-t}$ for $t \in \mathbb{T}$. Anderson [And82] proved that $(Y_t, t \in \mathbb{T})$ satisfies the SDE (see e.g. [Son+21, Equation (6)])

$$\mathrm{d}Y_t = -[\tilde{f}(t, Y_t) - \tilde{g}^2(t)\nabla_x \log \ell_t(Y_t)]\,\mathrm{d}t + \tilde{g}(t)\,\mathrm{d}B_t.$$

Above, ℓ_t is the density of Y_t; $\tilde{f}$ is defined as $\tilde{f}(t, x) := f(T - t, x)$; and $\tilde{g}$ is defined as $\tilde{g}(t) := g(T - t)$.

In the thesis, we study three questions: Langevin dynamics (1.14) for non-convex potentials, well-posedness of (1.16) and Euler-Maruyama scheme for (1.16). The first question is to get ourself into the research problem. The other two questions (with corresponding preprints) will be the original contribution of the thesis. In the following, we give an introduction for each of them. We emphasize that the cited works below do not exhaustively cover the related literature. As such, we also direct the readers to the references therein.

1.2.1 Langevin dynamics

Let V and $\bar{x}$ be as in Section 1.1. For $\gamma > 0$, let $\pi^\gamma \in \mathcal{P}(\mathbb{R}^d)$ such that $\mathrm{d}\pi^\gamma \propto e^{-\gamma V}\,\mathrm{d}x$. Under mild conditions except convexity (see e.g. [Hwa80; Hwa81; AH10; Bra22]), π^γ converges weakly to $\delta_{\bar{x}}$ as $\gamma \to \infty$. Let $\phi^\gamma(\mu)$ be the (non-negative) relative entropy of $\mu \in \mathcal{P}(\mathbb{R}^d)$ w.r.t π^γ. Then $\phi^\gamma(\mu) = 0$ i.f.f $\mu = \pi^\gamma$. Our goal is to construct a curve $(\mu_t^\gamma) \subset \mathcal{P}_2(\mathbb{R}^d)$ that approximates π^γ in the metric W_2. We are motivated by [Cra17] where she established existence of gradient flows in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ for ω-convex functionals. In particular, ∇V is not necessarily locally Lipschitz. In our setting, there exists a non-decreasing continuous function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$, called *Osgood modulus of continuity*, such that

1. $\psi(0) = 0$ and $\int_0^1 \frac{\mathrm{d}s}{\psi(s)} = \infty$.
2. $\psi(s) \geq s$ for $s \in \mathbb{R}_+$.
3. $|\nabla V(x) - \nabla V(y)|^2 \leq \psi(|x - y|^2)$ for $x, y \in \mathbb{R}^d$.

An example of ψ taken from [Cra17] is

$$\psi(s) = \begin{cases} 0 & \text{if} \quad s = 0, \\ s|\ln s|^2 & \text{if} \quad s \in (0, \beta], \\ s - 2\beta\ln(\beta) & \text{if} \quad s \in (\beta, \infty), \end{cases} \qquad \text{with} \quad \beta := e^{-(1+\sqrt{2})}.$$

We refer to [FZ05; LW14; Luo18] for well-posedness results of an SDE that assumed Osgood modulus of continuity. We consider

$$\alpha^\gamma := \inf\{\phi^\gamma(\mu) : \mu \in \mathcal{P}_2(\mathbb{R}^d)\}.$$

Our approach is by gradient flows in the space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$. We obtain

1. **(Existence)** One candidate for such an approximating curve is the unique solution (in distributional sense) to the PDE

$$\partial_t \mu_t^\gamma = \gamma \operatorname{div}(\mu_t^\gamma \nabla V) + \Delta \mu_t^\gamma.$$

2. **(Convergence)** If, in addition, π^γ satisfies a κ-logarithmic Sobolev inequality for some $\kappa > 0$, then we obtain an exponential rate of convergence

$$\phi^\gamma(\mu_t^\gamma) - \alpha \le e^{-2\kappa t}\{\phi^\gamma(\mu_0^\gamma) - \alpha\} \quad \text{for} \quad t \ge 0.$$

To obtain convergence rate, [RRT17; Xu+18; EMS18] assumed that V satisfies $\langle \nabla V(x), x\rangle \ge c_1|x|^2 - c_2$ for some constants $c_1, c_2 > 0$. In [KNS16; BB18; LMS23], V is assumed to satisfies Polyak-Łojasiewicz inequality, i.e., $\frac{1}{2}|\nabla V(x)|^2 \ge c\{V(x) - \min V\}$ for some constant $c > 0$. Optimizing a functional $G : \mathcal{P}(\mathbb{R}^d) \to (-\infty, \infty]$ has attracted considerable research efforts because training an "ideal" two-layer neural network can be cast as such problem.

Example 1.5 (Neural Network)**.** A two-layer neural network is described as a map

$$g^n : \mathbb{R}^d \to \mathbb{R}, \quad x \mapsto \frac{1}{n}\sum_{k=1}^n c_k \sigma(w_k^\top x),$$

for some *parameter* $(w_k, c_k)_{k=1}^n \subset \mathbb{R}^d \times \mathbb{R}$ and *activation function* $\sigma : \mathbb{R} \to \mathbb{R}$. Here n is called the *width* of the hidden layer. Due to law of large numbers and under regularity assumption,

$$g^n(x) \to g_\nu(x) := \int_{\mathbb{R}^d \times \mathbb{R}} c\sigma(w^\top x)\, \mathrm{d}\nu(w,c) \quad \text{as} \quad n \to \infty,$$

where $\nu \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R})$ is the asymptotic distribution of parameters. Let our data be (X, Y). We can estimate ν by solving

$$\inf\{\phi(\mu) : \mu \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R})\} \quad \text{with} \quad \phi(\mu) := \mathbb{E}[\{Y - g_\mu(X)\}^2]. \tag{1.17}$$

We review some related works in this general framework.

1. [MMN18; SS20a] studied an approximation of (1.17) by stochastic gradient descend when g_μ is replaced by g^n; and the population (X, Y) is replaced by its sample $(X_i, Y_i)_{i=1}^m$. They obtained some non-asymptotic convergence rates.

2. Chizat-Bach [CB18] considered $V : \Omega \to \mathcal{F}$ where Ω is the closure of a convex open subset of $\mathbb{R}^d$ and $\mathcal{F}$ is a separable Hilbert space.

3. [Hu+21; Chi22; CRW23] studied the problem $\inf\{\phi(\mu) : \mu \in \mathcal{P}(\mathbb{R}^d)\}$ for a generic map $\phi : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ that admits a *linear functional derivative* (see e.g. [CD18, Volume I - Part II]).

The results in this section are not completely novel, and their purpose is to get ourself familiar with the rigorous machinery developed by Ambrosio-Gigli-Savaré [AGS08]. This approach by gradient flows could be adapted without difficulty to deal with functionals of the form $G^\gamma := F \circ \phi^\gamma$ for some function $F : \mathbb{R}_+ \to \mathbb{R}_+$ regular enough.

1.2.2 Well-posedness of classical MV-SDEs

[McK67; TH81; Tan84; NT85; Léo86; HM86; Szn91; BT96; Mél96; FM97; BT97; Ben+98; BRV98; MS19; MV20; Cha20] studied (1.15) where the coefficients are of integral form, i.e.,

$b(t,x,\mu) = \int_{\mathbb{R}^d} \hat{b}(t,x,y)\,\mathrm{d}\mu(y)$ and $\sigma(t,x,\mu) = \int_{\mathbb{R}^d} \hat{\sigma}(t,x,y)\,\mathrm{d}\mu(y)$ for some $\hat{b} : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ and $\hat{\sigma} : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$. We mention some works where b (or $\hat{b}$) is continuous in x:

1. Dawson [Daw83] and Scheutzow [Sch86] considered specific models where b contains a polynomial component.

2. Funaki [Fun84] assumed that b and σ are continuous. He proved existence of martingale problem (MP) under a Lyapunov-type condition and uniqueness under a Lipschitz condition. The proof is by Euler scheme. We refer to [MS19; HŠS21a; LM22; LM23; LM24] where Lyapunov-type conditions were used.

3. Oelschläger [Oel84] proved well-posedness of MP when b, σ are Lipschitz and σ is bounded. Scheutzow [Sch87] provided some examples of (1.15) having more than one solution when $\sigma = 0$ and b is locally (but not globally) Lipschitz in μ. Méléard-Roelly [MR88] proved well-posedness of MP when b is Lipschitz and σ is of integral form. Gärtner [Gär88] assumed that b, σ are continuous and proved well-posedness of MP under some coercivity and monotonicity conditions, which are weaker than those in [Léo86].

4. Graham-Méléard [Gra92; GM97] considered (1.15) with jumps and proved well-posedness of MP when b, σ are Lipschitz. Erny [Ern22] also considered (1.15) with jumps and proved strong well-posedness when

   a) b, σ satisfy some growth condition and X_0 has exponential moment; and
   b) b is locally Lipschitz in (x, μ) and σ is Lipschitz in (x, μ).

   Graham-Méléard [GM93; GM94a; GM94b] studied interacting systems which are not necessarily exchangeable nor Markovian.

5. [Ben+98; BRV98] studied one-dimensional constant-diffusion models. They proved strong well-posedness when X_0 satisfies some integrability condition and $\hat{b}(t,x,y) = \beta(x-y)$. Here the function $\beta : \mathbb{R} \to \mathbb{R}$ is odd increasing locally Lipschitz and has polynomial growth.

6. Chaudru de Raynal [Cha20] proved strong well-posedness when

   a) $\hat{b}$ is bounded and Hölder in x;
   b) $\hat{b}$ is differentiable in y with bounded derivative;
   c) $\hat{\sigma}$ is uniformly non-degenerate and Lipschitz in (x, y); and
   d) $\hat{\sigma}$ is differentiable in y with derivative being bounded and Hölder in x.

   The proof is by Zvonkin transformation and parametrix expansion of transition density. With differential calculus in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$, Chaudru De Raynal and Frikha [CF22] extended [Cha20] to more general (b, σ). For differentiation in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$, we refer to [HW21a; Wan23a].

7. As a corollary of their path-dependent framework, Hammersley-Šiška-Szpruch [HŠS21b] proved weak existence under two sets of assumptions. First, X_0 satisfies some integrability condition and (b, σ) is bounded continuous in (x, μ) w.r.t the weak topology of $\mathcal{P}(\mathbb{R}^d)$. Second, σ is non-degenerate and b, σ are of integral form with $(\hat{b}, \hat{\sigma})$ being bounded.

8. Kalinin-Meyer-Proske [KMP22] proved strong well-posedness when

   a) (b, σ) is continuous in x and satisfies some growth and local boundedness conditions; and
   b) b is locally monotonic in x and σ is uniformly continuous in x.

The authors also established some exponential stability estimates. We refer to [Wan18; KMP24; HHL24] where monotonicity conditions were used.

9. For state spaces more general than $\mathbb{R}^d$, we refer to [CKS91; DV95; BM19b]. For MV-SDEs driven by noise other than Brownian motion, we refer to [HY21; RJM22; GHM22; Fan+22; RJM23; DH23; GHM23]. For SDEs that involve conditional expectation or conditional distribution, we refer to [Der03; LSZ20; LSZ22].

In some practical models [JR15; FPZ19; CGL23], the drift b is not continuous in x. We mention some works that tackle the discontinuity of b (or $\hat{b}$) in x:

1. With fixed-point argument, Jourdain [Jou97] proved well-posedness of MP when
   a) b is bounded and Lipschitz in μ; and
   b) $\sigma(t,x,\mu) = \sigma(x)$ is Lipschitz and uniformly elliptic.
2. In their study of Burgers equation, Bossy-Talay [BT96; BT97] considered (1.15) where σ is constant and $\hat{b}(t,x,y) = 1_{\mathbb{R}_+}(x-y)$. We refer to [Jou00a; Jou02; Jou00b; JMW05] where the convolution of $1_{\mathbb{R}_+}$ with measure appears. This special form of distribution-dependence is related to rank-based models [JM08; Shk12; JR13; ASZ19].
3. As a corollary of their path-dependent framework, Li-Min [LM16] proved weak well-posedness when
   a) b is bounded and uniformly continuous in μ; and
   b) $\sigma(t,x,\mu) = \sigma(t,x)$ is uniformly elliptic and Lipschitz in x.

   The proof is by Schauder fixed point theorem and Girsanov's theorem.
4. Bauer-Meyer-Proske [BMP18; BM19a] considered constant-diffusion models. They proved weak existence when b is continuous in μ and has at most linear growth.
5. Inspired by [MV20; GM01], Huang-Wang [HW19] proved weak existence when (b,σ) can be approximated by a sequence $(b^n,\sigma^n)_{n\in\mathbb{N}}$ of functions such that
   a) (b^n,σ^n) is bounded and Lipschitz in (x,μ); and
   b) b^n is locally integrable in (t,x) and σ^n is uniformly elliptic.

   There exists a unique strong solution if we assume, in addition, that
   a) (b,σ) is Lipschitz in μ;
   b) σ is uniformly continuous in x and weakly differentiable in x; and
   c) $|b(\cdot,\cdot,\mu)|^2 + |\nabla\sigma(\cdot,\cdot,\mu)|^2$ is locally integrable for any μ.

   We refer to [WR24] for an exposition of this direction; to [RZ21; Hua21; Zha24] where b is locally integrable in (t,x); and to [HW21b; HW22; Hua23; Ren23; Wan23b] for a generalization where b also contains a Lipschitz component.
6. Inspired by [VK76; Kry08], Mishura-Veretennikov [MV20] proved weak existence when $\hat{b},\hat{\sigma}$ have at most linear growth in x and $\hat{\sigma}$ is non-degenerate. The proof is by Krylov's estimates. Studying constant-diffusion models, Lacker [Lac23] obtained some well-posedness results under more relaxed linear growth assumptions on b. The proof is by relative entropy method. Partial generalization of [Lac23] was obtained by Han [Han22].

1.2.3 Well-posedness of density-dependent MV-SDEs

When replacing distribution-dependence by density-dependence in (1.15), we have

$$\begin{cases} X_t = b(t, X_t, \ell_t(X_t))\,\mathrm{d}t + \sigma(t, X_t, \ell_t(X_t))\,\mathrm{d}B_t, \\ \ell_t \text{ is the density of } X_t, \end{cases} \tag{1.18}$$

where

$$b : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d,$$
$$\sigma : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d \otimes \mathbb{R}^d.$$

Compared to (1.16), (1.18) does not involve distribution but its σ additionally has density-dependence. Let $a := \sigma\sigma^\top$. We use notations $b(t, x, r)$ and $\sigma(t, x, r)$ for coefficients of (1.18). First, we review some works about (1.18).

1. Studying PoC for Burgers equation, [CP83; GK83; OK85; Szn86] considered one-dimensional constant-diffusion models where $b(t, x, r) = r$. Roynette-Vallois [RV95] proved strong well-posedness in a more general setting where $b(t, x, r) = \frac{r^{\alpha-1}}{\alpha}$ for some $\alpha > 1$. Studying porous medium equations, [Ino91; Ben+96] considered driftless models where $\sigma(t, x, r) = r^\alpha I_d$ for some $\alpha > 0$.

2. [Oel85; MR87] considered constant-diffusion models where $b(t, x, r) = b(x, r)$ is bounded Lipschitz and $(x, r) \mapsto rb(x, r)$ is Lipschitz. Jourdain [Jou97] proved well-posedness of MP when

   a) $r \mapsto rb(t, x, r)$ is bounded Lipschitz; and
   b) $\sigma(t, x, r) = \sigma(x)$ is Lipschitz and uniformly elliptic.

   Studying SDEs driven by α-stable process, Simon-Olivera [SO18] considered constant-diffusion models where $b(t, x, r) = b(x, r)$ is bounded Lipschitz.

3. Jourdain-Méléard [JM98] considered models where $b(t, x, r) = b(r)$ and $\sigma(t, x, r) = \sigma(r)$. Bringing together results from [FBL14; LS88], they proved strong well-posedness when

   a) b is C^2 (twice continuously differentiable);
   b) σ is C^3 and Lipschitz;
   c) $r \mapsto (ra(r))'$ is strongly elliptic; and
   d) ℓ_0 belongs to the Hölder space $H^{2+\alpha}(\mathbb{R}^d)$.

4. [Ber+10; BJT11; BJ15; BJ18] considered stochastic Lagrangian models where the dependence is on conditional density.

5. Studying generalized porous media equations, [BRR10; BRR11] considered one-dimensional driftless models where $\sigma(t, x, r) = \sqrt{\frac{\beta(r)}{r}} I_d$. Here the function $\beta : \mathbb{R} \to \mathbb{R}$ is (possibly discontinuous) increasing with $\beta(0) = 0$. The proof is by nonlinear semigroup [Bar10; Bar21]. We refer to [BCR13; BRR17] for partial extensions of those results; to [BBC75; BC79; BC81; Pie82; EM24] for studies of the related PDE $\partial_t u = \frac{1}{2}\partial^2_{xx}\beta(u)$; and to [Var91; Uch00] where the PDE $\partial_t u = \frac{1}{2}\partial^2_{xx}\beta(u)$ is the limit of constant-diffusion interacting particle systems.

6. Bossy-Jabir [BJ19] considered driftless models where $\sigma(t, x, r) = \sigma(r)$. Using comparison principles and energy estimates (see e.g. [Vaz06; Eva22]), they proved strong well-posedness when

a) σ is C^1 and ℓ_0 satisfies some integrability condition; and

b) $r \mapsto (ra(r))'$ is continuous and strongly elliptic.

7. In a series of papers [BR21b; BR20; BR23a; BR18; BR23b; BR24b], Barbu-Röckner considered models where

   a) b and σ are density-dependent; and

   b) $r \mapsto ra(t, x, r)$ satisfies some monotonicity condition.

   They used PDE techniques to study the associated Fokker-Planck equation. Well-posedness of (1.18) then follows from superposition principle [Fig08; Tre16; RXZ20; BRS21]. We refer to [BR24a] for an exposition of this direction; to [Mar23] for a dual variational approach; to [Bog+15; Bar16; Bar18] for studies of Fokker-Planck equations; and to [Gru23; Gru24; RRW22; BSR23; Reh23b; RR23; Reh23a] for partial extensions and related works.

8. [IR23; Iss+24] considered constant-diffusion models where b involves a term belonging to a negative Besov space, besides density-dependence.

The literature on (1.16), which is the object of the thesis, is much scarcer. We use notations $b(t, x, r, \mu)$ and $\sigma(t, x, \mu)$ for coefficients of (1.16).

1. Wang [Wan23c] proved strong well-posedness when

   a) b is locally integrable in (t, x) and Lipschitz in (r, μ);

   b) $\sigma(t, \cdot, \mu)$ is uniformly elliptic and belongs to the Hölder space $C_b^{1+\alpha}(\mathbb{R}^d)$;

   c) $\|\sigma(t, \cdot, \mu) - \sigma(t, \cdot, \tilde{\mu})\|_{C_b^\alpha} \le C\|\mu - \tilde{\mu}\|_\infty$; and

   d) ℓ_0 satisfies some integrability condition.

   He used the space of bounded densities (together with supremum norm) for (absolutely continuous) distributions.

2. Hao-Röckner-Zhang [HRZ24] proved strong well-posedness when

   a) b is locally integrable in (t, x) and Lipschitz in (r, μ);

   b) $\sigma(t, x, \mu) = \sigma(t, x)$ is uniformly elliptic and Hölder in x;

   c) σ is weakly differentiable in x and $\nabla\sigma$ is locally integrable in (t, x); and

   d) ℓ_0 satisfies some integrability condition.

   They used total variation metric for distributions.

3. Zhang [Zha22] considered kinetic models where both b and σ depend on density as well as distribution. Here the distribution-dependence is of integral form.

In the same vein as [Wan23c; HRZ24], we aim at proving some well-posedness results for (1.16). Let us elaborate on our contribution that distinguishes the thesis from [Wan23c; HRZ24]:

1. We use a mollifying argument whereas [HRZ24; Wan23c] employed a Picard-iteration argument. We argue that our approach is more flexible because it can be extended to those situations where reasonable Hölder estimates are available. An interesting case for future study is when b is allowed to grow linearly in space as in [MPZ21].

2. For existence result, the conditions on σ in [Wan23c; HRZ24] are more restrictive than ours. First, [HRZ24] assumed that $\sigma(t, x, \mu) = \sigma(t, x)$. Second, [Wan23c] assumed that σ is Lipschitz in space, that $\nabla\sigma$ is Hölder continuous, and that $\|\sigma(t, \cdot, \mu) - \sigma(t, \cdot, \nu)\|_{C_b^\alpha} \lesssim \|\ell_\mu - \ell_\nu\|_\infty$. Here ℓ_μ, ℓ_ν are the densities of μ, ν respectively.

3. To be more aligned with existing literature of Mckean-Vlasov SDEs, we use Wasserstein metric for assumptions of continuity. This makes estimating supremum norm between marginal densities (as in [HRZ24; Wan23c]) (of two weak solutions) not applicable in our case. However, using Wasserstein metric to estimate the difference between marginal distributions (of two weak solutions) is also not applicable due to the presence of pointwise density $\ell_t(X_t)$ in b. To overcome these difficulties, we will estimate weighted total variation distance between marginal densities.

1.2.4 Euler-Maruyama scheme for density-dependent MV-SDEs

For discretization, we consider a particular case of (1.16), i.e.,

$$X_t = X_0 + \int_0^t b(s, X_s, \ell_s(X_s), \mu_s) \, \mathrm{d}s + \sqrt{2} B_t. \tag{1.19}$$

For $n \in \mathbb{N}$, the step size is defined as $\varepsilon_n := T/n$. The Euler-Maruyama scheme $X^n := (X_t^n, t \in \mathbb{T})$ of (1.19) is constructed by

$$\begin{cases} X_t^n := X_0 + \int_0^t b(s, X^n_{\tau^n_s}, \ell^n_{\tau^n_s}(X^n_{\tau^n_s}), \mu^n_{\tau^n_s}) 1_{(\varepsilon_n, T]}(s) \, \mathrm{d}s + \sqrt{2} B_t, \\ \mu^n_s \text{ is the distribution of } X^n_s, \text{ and } \ell^n_s \text{ is the density of } X^n_s. \end{cases}$$

We review some works about discretization of (1.16).

1. Bencheikh-Jourdain [BJ22] considered constant-diffusion models where $b(t, x, r, \mu) = b(t, x)$ is bounded. They obtained $\sup\{\|\mu_{t_k} - \mu^n_{t_k}\|_{\mathrm{TV}} : k = 1, \ldots, n\} \lesssim \sqrt{\varepsilon_n}$. The proof is by randomization in time.

2. Jourdain-Menozzi [JM24] considered constant-diffusion models where $b(t, x, r, \mu) = b(t, x)$ is locally integrable. They proved that the pointwise difference between transition densities is bounded by $|\varepsilon_n|^\alpha$ (with $\alpha := 1 - (\frac{d}{p} + \frac{2}{q})$) multiplied by some centered Gaussian density. The proof is by randomization in time and cutoff in space. This result was extended by Fitoussi-Jourdain-Menozzi [FJM24] to the case of α-stable noise.

3. Lê-Ling [LL22] considered models where

   a) $b(t, x, r, \mu) = b(t, x)$ is locally integrable;

   b) $\sigma(t, x, \mu) = \sigma(t, x)$ is uniformly elliptic and Hölder in x; and

   c) σ is weakly differentiable in x and $\nabla \sigma$ is locally integrable.

   They obtained strong rate of convergence. The proof is by stochastic sewing lemma.

4. [DGI22; JIP23] considered constant-diffusion models where $b(t, x, r, \mu) = b(t, x)$ belongs to some subspace of Schwartz distributions. They obtained convergence results for $\sup_{t \in \mathbb{T}} \mathbb{E}[|X^n_t - X_t|]$.

5. With Euler approximation, Hao-Röckner-Zhang [HRZ21] proved strong well-posedness when

   a) σ is constant;

   b) $b(t, x, r, \mu) = b(t, x, r)$ is bounded and Lipschitz in r; and

   c) ℓ_0 satisfies some integrability condition.

   Hao [Hao23] obtained $\sup_{t \in \mathbb{T}} \|\ell_t - \ell^n_t\|_1 \lesssim \sqrt{\varepsilon_n}$ for convergence in [HRZ21]. Wu-Hao [WH23] extended [HRZ21] to SDEs driven by α-stable noise. Song-Hao [SH24] obtained $\sup_{t \in \mathbb{T}} \|\ell_t - \ell^n_t\|_1 \lesssim |\varepsilon_n|^{\frac{\alpha-1}{\alpha}}$ for convergence in [WH23].

There appears to be a shortage of results about (1.16). In this direction, we have found only four papers [Wan23c; HRZ24; Zha22; Le24b]. To our best knowledge, there has not been any study on Euler-Maruyama scheme of (1.16). This scarcity is the motivation for our study of (1.19). We assume that $b(t, x, r, \mu)$ is bounded and Lipschitz in (r, μ). The drift $b(t, x, r, \mu)$ is not necessarily continuous in x. Our contribution that distinguishes the thesis from [HRZ21; Hao23; WH23; SH24] is that b also depends on distribution in our case. First, we obtain Hölder continuity of the scheme $(X_t^n, t \in \mathbb{T})_{n\in\mathbb{N}}$. As a consequence of this regularity, we give a direct proof of well-posedness of (1.19). Finally, we derive the rate of convergence

$$\sup_{t\in\mathbb{T}} \int_{\mathbb{R}^d} (1 + |x|)|\ell_t^n(x) - \ell_t(x)| \, \mathrm{d}x \lesssim |\varepsilon_n|^{\frac{\alpha}{2}},$$

where α is the Hölder exponent of ℓ_0.

Chapter 2

Preliminary

2.1 Optimal transport

We have explained at the end of Section 1.2.3 that using Wasserstein metric instead of infinity norm distinguishes the thesis from [Wan23c; HRZ24]. Therefore, we recall definitions and results about Wasserstein spaces from [Vil03; Vil09] and about gradient flows from [AGS08]. Currently, optimal transport enjoys numerous applications in statistics, artificial intelligence, and signal processing, to name just a few. See e.g. the works of Peyre, Cuturi and Bach.

The genesis of optimal transport can be traced back to Gaspard Monge in 1781 [Mon81]. In modern terminology, Monge problem is formulated as follows: given two probability measures μ, ν defined on measurable spaces X, Y respectively, find a measurable map $\bar{T} : X \to Y$ with $\nu = \bar{T}_\sharp \mu$ such that $\bar{T}$ minimizes the transport cost, i.e.,

$$\int_X c(x, \bar{T}(x) \,\mathrm{d}\mu(x) = \min \left\{ \int_X c(x, T(x) \,\mathrm{d}\mu(x) : \nu = T_\sharp \mu \right\}, \tag{2.1}$$

where $c : X \times Y \to \mathbb{R}_+$ is a given cost function. When T satisfies $\nu = T_\sharp \mu$, it is called a *transport map*, and if T also minimizes the cost, we call it *optimal transport map*. Even existence of a solution to (2.1) turns out to be very challenging (see e.g. [Mag23, Sections 1.1 and 1.2]).

A major advance is due to Kantorovich [Kan42; Kan48], who suggested to look for *transport plans* instead of transport maps, i.e., probability measures on $X \times Y$ whose first and second marginal are μ and ν respectively. With this relaxation, existence of an *optimal transport plan* can be established easily by direct method from calculus of variations.

The next advance is due to Brenier [Bre87], who proved well-posedness of optimal transport map when the cost is the squared Euclidean metric. The result was then extended to compact Riemannian manifolds by McCann [McC01] and led to connections with PDEs, fluid mechanics, differential geometry, probability theory and functional analysis. See e.g. the works of Otto, Villani, Figalli and Ambrosio.

2.1.1 Wasserstein spaces

We fix $p \in [1, \infty)$. Let $M_p(\mu) \coloneqq \int_{\mathbb{R}^d} |x|^p \,\mathrm{d}\mu(x)$ be the p-th moment of $\mu \in \mathcal{P}(\mathbb{R}^d)$. Let $\mathcal{P}_p(\mathbb{R}^d)$ be the set of those measures in $\mathcal{P}(\mathbb{R}^d)$ with finite p-th moment. For $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, the set of *transport plans* (or *couplings*) between them is defined as

$$\Gamma(\mu, \nu) \coloneqq \{\varrho \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \mu = \pi^1_\sharp \varrho \text{ and } \nu = \pi^2_\sharp \varrho\},$$

where π^i is the projection of $\mathbb{R}^d \times \mathbb{R}^d$ onto its i-th coordinate, and $\pi^i_\sharp \varrho$ is the push-forward of ϱ through π^i. For $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, we define

$$W_p(\mu, \nu) \coloneqq \inf \left\{ \int_{\mathbb{R}^d} |x - y|^p \,\mathrm{d}\varrho(x, y) : \varrho \in \Gamma(\mu, \nu) \right\}^{1/p}. \tag{2.2}$$

By [Vil09, Theorem 6.18], $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ is a Polish space. By [Vil09, Theorem 6.9], it holds for $\mu_n, \mu \in \mathcal{P}_p(\mathbb{R}^d)$ that $W_p(\mu_n, \mu) \to 0$ i.f.f $\mu_n \rightharpoonup \mu$ and $M_p(\mu_n) \to M_p(\mu)$. Let Φ_p be the collection of all $(\varphi, \psi) \in C_b(\mathbb{R}^d) \times C_b(\mathbb{R}^d)$ such that $\varphi(x) + \psi(y) \le |x-y|^p$ for $x, y \in \mathbb{R}^d$. For brevity, we denote $W_p^p(\mu, \nu) := (W_p(\mu, \nu))^p$. We denote by $|\mu - \nu|$ the variation of the signed measure $\mu - \nu$ as in [Rud87, Section 6.1]. By [Rud87, Theorems 6.2 and 6.4], $|\mu - \nu|$ is a non-negative finite measure. We recall properties needed for Chapter 4:

Lemma 2.1. 1. *[Vil03, Theorem 1.3] It holds for $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ that*

$$W_p^p(\mu, \nu) = \sup \left\{ \int_{\mathbb{R}^d} \varphi \, \mathrm{d}\mu + \int_{\mathbb{R}^d} \psi \, \mathrm{d}\nu : (\varphi, \psi) \in \Phi_p \right\}.$$

2. *[Vil03, Theorem 1.14] It holds for $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ that*

$$W_1(\mu, \nu) = \sup \left\{ \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu - \nu) : f \in L^1(|\mu - \nu|) \text{ with } [f]_1 \le 1 \right\}.$$

3. *[Vil03, Remark 7.1.2] It holds for $1 \le p \le q < \infty$ and $\mu, \nu \in \mathcal{P}_q(\mathbb{R}^d)$ that $W_p(\mu, \nu) \le W_q(\mu, \nu)$.*

Above, the first claim is called Kantorovich duality while the second one is called Kantorovich-Rubinstein theorem. For more information about optimal transport, we refer to [FG21; Mag23; Vil09; Vil03; AG13; San15; ABS21; Tho23]. The next result states that W_p^p is controlled by the weighted L^1-metric.

Lemma 2.2. *Let $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ be absolutely continuous with corresponding densities ℓ_μ, ℓ_ν. Then*

$$W_p^p(\mu, \nu) \le (1 \vee 2^{p-1}) \int_{\mathbb{R}^d} |x|^p |\ell_\mu - \ell_\nu|(x) \, \mathrm{d}x.$$

Proof. For $B \in \mathcal{B}(\mathbb{R}^d)$, we denote by $\Pi(B)$ the collection of all finite measurable partitions of B. This means $(B_1, \ldots, B_n) \in \Pi(B)$ i.f.f $\{B_1, \ldots, B_n\} \subset \mathcal{B}(\mathbb{R}^d)$ are pairwise disjoint and $B = \bigcup_{k=1}^n B_k$. We have

$$\begin{aligned} |\mu - \nu|(B) &= \sup \left\{ \sum_{k=1}^n |(\mu - \nu)(B_k)| : (B_1, \ldots, B_n) \in \Pi(B) \right\} \\ &= \sup \left\{ \sum_{k=1}^n \Big| \int_{B_k} (\ell_\mu - \ell_\nu)(x) \, \mathrm{d}x \Big| : (B_1, \ldots, B_n) \in \Pi(B) \right\} \\ &\le \sup \left\{ \sum_{k=1}^n \int_{B_k} |\ell_\mu - \ell_\nu|(x) \, \mathrm{d}x : (B_1, \ldots, B_n) \in \Pi(B) \right\} \\ &= \int_B |\ell_\mu - \ell_\nu|(x) \, \mathrm{d}x. \end{aligned}$$

On the other hand, we have from [Vil03, Proposition 7.10] that

$$W_p^p(\mu, \nu) \le (1 \vee 2^{p-1}) \int_{\mathbb{R}^d} |x|^p \, \mathrm{d}|\mu - \nu|(x).$$

The claim then follows. □

2.1.2 Geodesics in Wasserstein spaces

In this section, we work with $(\mathcal{P}_2(\mathbb{R}^d), W_2)$. Let $\Gamma_o(\mu, \nu)$ be the set of minimizers in (2.2). By [Vil03, Theorem 1.3], $\Gamma_o(\mu, \nu)$ is non-empty. We call $\varrho \in \Gamma_o(\mu, \nu)$ an *optimal transport plan* between μ and ν. A *curve* $\gamma : [0,1] \to \mathcal{P}_2(\mathbb{R}^d)$ from μ to ν is a continuous map with $\gamma_0 = \mu$ and $\gamma_1 = \nu$. The *length* of γ is defined as

$$\mathrm{Len}(\gamma) := \sup \left\{ \sum_{k=0}^{n-1} W_2(\gamma_{t_k}, \gamma_{t_{k+1}}) : n \in \mathbb{N} \text{ and } 0 = t_0 < t_1 < \cdots < t_n = 1 \right\}.$$

We define $\pi_t : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ by $\pi_t := (1-t)\pi^1 + t\pi^2$ for $t \in [0,1]$. The *(interpolating) curve induced by* $\varrho \in \Gamma(\mu,\nu)$ is a curve γ from μ to ν defined by $\gamma_t := (\pi_t)_\sharp \varrho$ for $t \in [0,1]$. A *geodesic* from μ to ν is a curve with shortest length among those curves from μ to ν. A *constant-speed* geodesic $\gamma : [0,1] \to \mathcal{P}_2(\mathbb{R}^d)$ from μ to ν is a geodesic from μ to ν such that $W_2(\gamma_s, \gamma_t) = |s-t| W_2(\mu,\nu)$ for $s,t \in [0,1]$. By [AGS08, Theorem 7.2.2], γ is a constant-speed geodesic from μ to ν i.f.f γ is induced by some $\varrho \in \Gamma_o(\mu,\nu)$. From now on, all geodesics are understood as constant-speed geodesics.

Let $\mathcal{P}_2^a(\mathbb{R}^d)$ be the subset of $\mathcal{P}_2(\mathbb{R}^d)$ that contains measures admitting a density. Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, \infty]$. The *proper domain* $D(\phi)$ of ϕ is defined as $\{\mu \in \mathcal{P}_2(\mathbb{R}^d) : \phi(\mu) < \infty\}$. We say that ϕ is *proper* if $D(\phi)$ is non-empty, and that ϕ is *lower semi-continuous* if $\phi(\mu) \le \liminf_n \phi(\mu_n)$ whenever $\mu_n \rightharpoonup \mu$. In the remaining of this section, we assume $D(\phi) \subset \mathcal{P}_2^a(\mathbb{R}^d)$.

Definition 2.3.

1. [AGS08, Definition 9.2.2] For $\mu_1, \mu_2, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, a *generalized geodesic* from μ_1 to μ_2 with base ν is a curve induced by $\pi^{1,2}_\sharp \varrho \in \Gamma(\mu_1,\mu_2)$ for some $\varrho \in \Gamma(\mu_1,\mu_2,\nu)$ with $\pi^{1,3}_\sharp \varrho \in \Gamma_o(\mu_1,\nu)$ and $\pi^{2,3}_\sharp \varrho \in \Gamma_o(\mu_2,\nu)$. Here $\pi^{i,j}$ is the projection of $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$ onto its i-th and j-th coordinates. If ν coincides with either μ_1 or μ_2, then a generalized geodesic is also a geodesic.

2. [Cra17, Definition 2.17] A map $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ is a *modulus of continuity* if it is continuous non-decreasing and vanishes only at $s=0$. If, in addition, there exists another modulus of continuity $\bar\omega$ such that $\int_0^1 \frac{\mathrm{d}s}{\bar\omega(s)} = \infty$ and $|\omega(s) - \omega(t)| \le \bar\omega(|s-t|)$ for $s,t \ge 0$, then ω is called an *Osgood modulus of convexity*. In this case, $\bar\omega$ is called an *Osgood modulus of continuity* of ω.

3. [Cra17, Definition 2.4] For a modulus ω of convexity and a curve γ induced by $\varrho \in \Gamma(\mu,\nu)$, the map ϕ is *ω-convex along γ (with constant $\lambda \in \mathbb{R}$)* if it holds for $t \in [0,1]$ that

$$\begin{aligned}\phi(\gamma_t) \le{}& (1-t)\phi(\gamma_0) + t\phi(\gamma_1) \\ &- \frac{\lambda}{2}\{(1-t)\omega(t^2\mathbb{K}_2^2(\varrho)) + t\omega((1-t)^2\mathbb{K}_2^2(\varrho))\},\end{aligned}$$

where $\mathbb{K}_2(\varrho) := \|\pi^2 - \pi^1\|_{L^2(\varrho)}$. In case $\omega = \mathrm{id}$, we simply say ϕ is *λ-convex along* γ.

The map ϕ is *ω-convex along geodesics* if for $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$ there exists a geodesic γ from μ to ν such that ϕ is ω-convex along γ. The map ϕ is *ω-convex along generalized geodesics* if for $\mu_0, \mu_1, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ there exists a generalized geodesic γ from μ_0 to μ_1 with base ν such that ϕ is ω-convex along γ. Clearly, ω-convexity along generalized geodesics implies that along geodesics. In case $\omega = \mathrm{id}$, we simply say ϕ is *λ-convex along geodesics/generalized geodesics*, which is consistent with [AGS08, Definitions 9.1.1 and 9.2.2]. In case $\lambda = 0$, we simply say ϕ is *convex along geodesics/generalized geodesics*.

Lemma 2.4. *[Cra17, Propositions 2.8 and 4.7] Let V and ψ be given by Assumption 3.1. We define $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ by $\omega(s) := \sqrt{s\psi(s)}$. We define $\mathcal{V} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ by $\mathcal{V}(\mu) := \int_{\mathbb{R}^d} V \,\mathrm{d}\mu$. We fix $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\varrho \in \Gamma(\mu,\nu)$. Let γ be the curve induced by ϱ. Then $t \mapsto \mathcal{V}(\gamma_t)$ is differentiable with*

$$\mathcal{V}(\gamma_1) - \mathcal{V}(\gamma_0) - \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{V}(\gamma_t)\Big|_{t=0} \ge 2C\omega(\mathbb{K}_2^2(\varrho)). \tag{2.3}$$

As a consequence, $\mathcal{V}$ is ω-convex along any interpolating curve with constant $\lambda_\omega = -4C < 0$.

Proof. First, we are going to prove (2.3). By Assumption 3.1(2) and Leibniz integral rule,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{V}(\gamma_t) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \nabla V \circ \pi_t, \pi^2 - \pi^1\rangle \,\mathrm{d}\varrho.$$

By Taylor formula,

$$\mathcal{V}(\gamma_1) - \mathcal{V}(\gamma_0) - \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{V}(\gamma_t)\Big|_{t=0} = \int_0^1 \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \nabla V \circ \pi_t - \nabla V\circ \pi^1, \pi^2 - \pi^1\rangle \,\mathrm{d}\varrho \,\mathrm{d}t.$$

Notice that $\pi_t - \pi^1 = t(\pi^2 - \pi^1)$. We have

$$\begin{aligned}
&\int_{\mathbb{R}^d\times\mathbb{R}^d} |\langle \nabla V \circ \pi_t - \nabla V\circ \pi^1, \pi^2 - \pi^1\rangle| \,\mathrm{d}\varrho \\
&\le \|\nabla V \circ \pi_t - \nabla V\circ \pi^1\|_{L^2(\varrho)} \|\pi^2 - \pi^1\|_{L^2(\varrho)} \quad \text{by Cauchy–Schwarz inequality} \\
&\le 2C \Big\|\sqrt{\psi(|\pi^2 - \pi^1|^2)}\Big\|_{L^2(\varrho)} \|\pi^2 - \pi^1\|_{L^2(\varrho)} \quad \text{by Assumption 3.1(3)} \\
&\le 2C\sqrt{\|\pi^2 - \pi^1\|^2_{L^2(\varrho)} \psi(\|\pi^2 - \pi^1\|^2_{L^2(\varrho)})} \quad \text{because } \sqrt{\psi} \text{ is concave} \\
&= 2C\omega(\mathbb{K}_2^2(\varrho)).
\end{aligned}$$

Second, we are going to prove the ω-convexity of $\mathcal{V}$. We fix $t \in (0,1)$. We need to prove

$$\begin{aligned}
\mathcal{V}(\gamma_t) \le{}& (1-t)\mathcal{V}(\gamma_0) + t\mathcal{V}(\gamma_1) \\
&- \frac{\lambda_\omega}{2}\{(1-t)\omega(t^2\mathbb{K}_2^2(\varrho)) + t\omega((1-t)^2\mathbb{K}_2^2(\varrho))\}.
\end{aligned} \tag{2.4}$$

Let $\hat\varrho := (\pi_t, \pi^2)_\sharp \varrho \in \Gamma(\gamma_t, \gamma_1)$ and $\check\varrho := (\pi_t, \pi^1)_\sharp \varrho \in \Gamma(\gamma_t, \gamma_0)$. Let $\hat\gamma, \check\gamma$ be the curves induced by $\hat\varrho, \check\varrho$ respectively. By (2.3),

$$\mathcal{V}(\hat\gamma_1) - \mathcal{V}(\hat\gamma_0) - \frac{\mathrm{d}}{\mathrm{d}s}\mathcal{V}(\hat\gamma_s)\Big|_{s=0} \ge 2C\omega(\mathbb{K}_2^2(\hat\varrho)), \tag{2.5}$$

$$\mathcal{V}(\check\gamma_1) - \mathcal{V}(\check\gamma_0) - \frac{\mathrm{d}}{\mathrm{d}s}\mathcal{V}(\check\gamma_s)\Big|_{s=0} \ge 2C\omega(\mathbb{K}_2^2(\check\varrho)). \tag{2.6}$$

Notice that $\hat\gamma_s = \gamma_{s(1-t)+t}$ and $\check\gamma_s = \gamma_{t(1-s)}$ for $s \in [0,1]$. By chain rule,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{V}(\hat\gamma_s)\Big|_{s=0} &= (1-t)\frac{\mathrm{d}}{\mathrm{d}r}\mathcal{V}(\gamma_r)\Big|_{r=t}, \\
\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{V}(\check\gamma_s)\Big|_{s=0} &= -t\frac{\mathrm{d}}{\mathrm{d}r}\mathcal{V}(\gamma_r)\Big|_{r=t}.
\end{aligned}$$

Notice that $\mathbb{K}_2(\hat\varrho) = (1-t)\mathbb{K}_2(\varrho)$ and $\mathbb{K}_2(\check\varrho) = t\mathbb{K}_2(\varrho)$. We multiply (2.5) with t, multiply (2.6) with $(1-t)$, and add the resulting inequalities. Then (2.4) follows. This completes the proof. □

2.1.3 Gradient flows in Wasserstein spaces

We define the *tangent space* at $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ as

$$T_\mu \mathcal{P}_2(\mathbb{R}^d) := \overline{\{\nabla\varphi : \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu)}.$$

Definition 2.5. 1. [AGS08, Definition 1.1.1] A map $\mu : (0,\infty) \to \mathcal{P}_2(\mathbb{R}^d)$ belongs to $\mathrm{AC}^p_{\mathrm{loc}}((0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ with $p \in [1,\infty)$ if there exists $f \in L^p_{\mathrm{loc}}((0,\infty);\mathbb{R}_+)$ such that $W_2(\mu_s,\mu_t) \le \int_s^t f(r)\,\mathrm{d}r$ for $0 < s < t < \infty$. In case $p = 1$, we say that (μ_t) is *locally* *absolutely continuous*. By [AGS08, Theorem 1.1.2], its *metric derivative*

$$|\mu'|(t) := \lim_{s\to 0} \frac{W_2(\mu_{t+s}, \mu_t)}{|s|}$$

exists for a.e. $t > 0$ and $W_2(\mu_s, \mu_t) \le \int_s^t |\mu'|(r)\,\mathrm{d}r$ for $0 < s < t < \infty$.

2. [AGS08, Definition 10.1.1] For $\mu \in D(\phi)$, a vector field $f \in L^2(\mu)$ belongs to the *Fréchet subdifferential* $\partial\phi(\mu)$ if

$$\phi(\nu) - \phi(\mu) \ge \int_{\mathbb{R}^d} \langle f, \mathbf{t}_\mu^\nu - \mathrm{id}\rangle\,\mathrm{d}\mu + \mathcal{O}(W_2(\mu,\nu)),$$

or equivalently

$$\liminf_{\nu\to\mu} \frac{\phi(\nu) - \phi(\mu) - \int_{\mathbb{R}^d}\langle f, \mathbf{t}_\mu^\nu - \mathrm{id}\rangle\,\mathrm{d}\mu}{W_2(\mu,\nu)} \ge 0,$$

where $\mathbf{t}_\mu^\nu$ is the optimal map that transports μ to ν. If $f \in \partial\phi(\mu)$ also satisfies

$$\phi(\mathbf{t}_\sharp\nu) - \phi(\mu) \ge \int_{\mathbb{R}^d} \langle f, \mathbf{t} - \mathrm{id}\rangle\,\mathrm{d}\mu + \mathcal{O}(\|\mathbf{t} - \mathrm{id}\,\|_{L^2(\mu)}),$$

then f is called a *strong subdifferential*.

3. [AGS08, Definition 11.1.1] A curve $(\mu_t) \in \mathrm{AC}^2_{\mathrm{loc}}((0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ is a *gradient flow* for ϕ if there exists a time-dependent Borel vector filed v such that

   a) $v_t \in T_{\mu_t}\mathcal{P}_2(\mathbb{R}^d)$ and $v_t \in -\partial\phi(\mu_t)$ for a.e. $t > 0$.

   b) $t \mapsto \|v_t\|_{L^2(\mu_t)}$ belongs to $L^2_{\mathrm{loc}}(0,\infty)$.

   c) $\partial_t\mu_t + \mathrm{div}(\mu_t v_t) = 0$ holds in distributional sense.

2.1.4 Auxiliary lemmas

Definition 2.6. 1. [AGS08, Definition 1.2.1] A map $g : \mathcal{P}_2(\mathbb{R}^d) \to [0,\infty]$ is a *strong upper gradient* for ϕ if, for every locally absolutely continuous curve (μ_t), the map $t \mapsto g(\mu_t)$ is Borel measurable and

$$|\phi(\mu_t) - \phi(\mu_s)| \le \int_s^t g(\mu_r)|\mu'|(r)\,\mathrm{d}r \quad \text{for} \quad 0 < s < t < \infty, \tag{2.7}$$

with the convention $|\phi(\mu_t) - \phi(\mu_s)| = \infty$ whenever $\phi(\mu_t) + \phi(\mu_s) = \infty$.

2. [AGS08, Definition 1.2.4] The *local* and *global slopes* of ϕ at $\mu \in D(\phi)$ are defined as

$$|\partial\phi|(\mu) := \limsup_{\nu\to\mu} \frac{(\phi(\mu) - \phi(\nu))^+}{W_2(\mu,\nu)} \quad \text{and} \quad \mathfrak{l}_\phi(\mu) := \sup_{\nu\neq\mu} \frac{(\phi(\mu) - \phi(\nu))^+}{W_2(\mu,\nu)}.$$

3. [AGS08, Definition 1.3.2] For $p \in (1,\infty)$, a locally absolutely continuous curve (μ_t) is a *p-curve of maximal slope* for ϕ w.r.t $|\partial\phi|$ if $t \mapsto \phi(\mu_t)$ is equal a.e. to a non-increasing map f and

$$\frac{|\mu'|^p(t)}{p} + \frac{|\partial\phi|^q(\mu_t)}{q} \le f'(t) \quad \text{for a.e.} \quad t > 0, \tag{2.8}$$

where q is the Hölder conjugate of p, i.e., $\frac{1}{p} + \frac{1}{q} = 1$.

4. [AGS08, Definition 10.1.4] The map ϕ is *regular* if it satisfies Assumption 2.7(1) and

$$\begin{cases} \text{the strong differential } f_n \in \partial\phi(\mu_n), \quad \varphi_n = \phi(\mu_n), \\ \mu_n \to \mu \text{ in } \mathcal{P}_2(\mathbb{R}^d), \quad \varphi_n \to \varphi, \quad \sup_n \|f_n\|_{L^2(\mu_n)} < \infty, \\ f_n \to f \in L^2(\mu) \text{ weakly in the sense of [AGS08, Definition 5.4.3]}, \end{cases} \implies \begin{cases} f \in \partial\phi(\mu), \\ \varphi = \phi(\mu). \end{cases}$$

5. [Cra17, Definition 2.12] For $t \geq 0$, the *proximal map* $J_t\phi$ of ϕ is defined as $J_0\phi[\mu] \coloneqq \mu$ and

$$J_t\phi[\mu] \coloneqq \arg\min \left\{ \phi(\mu) + \frac{W_2^2(\mu,\nu)}{2t} : \nu \in \mathcal{P}_2(\mathbb{R}^d) \right\} \quad \text{for} \quad t > 0.$$

First, we have a sufficient condition for being regular:

Assumption 2.7. 1. ϕ is proper and lower semi-continuous with $D(\phi) \subset \mathcal{P}_2^{\mathrm{a}}(\mathbb{R}^d)$.

2. There exists $\tau \in (0,\infty]$ such that $J_t\phi[\mu] \neq \emptyset$ for $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $t \in (0,\tau)$.

Lemma 2.8. *Assume that ϕ satisfies Assumption 2.7(1) and is ω-convex along geodesics with $\omega(s) = \mathcal{O}(\sqrt{s})$ as $s \downarrow 0$. Then ϕ is regular.*

Proof. The argument in the proof of [AGS08, Lemma 10.1.3] (for λ-convex functionals) also works in our case of ω-convex functionals. This is because that argument is due to the variational characterization of the Fréchet subdifferential, whose counterpart in our case (which is Lemma 2.9) exists. □

Second, we have a characterization of Fréchet subdifferential:

Lemma 2.9. *Let ϕ be ω-convex along geodesics with $\omega(s) = \mathcal{O}(\sqrt{s})$ as $s \downarrow 0$. We fix $\mu \in D(\phi)$ and $f \in L^2(\mu)$. Then $f \in \partial\phi(\mu)$ i.f.f it holds for $\nu \in D(\phi)$ that*

$$\int_{\mathbb{R}^d} \langle f, \mathbf{t}_\mu^\nu - \mathrm{id} \rangle \, \mathrm{d}\mu + \frac{\lambda_\omega}{2}\omega(W_2^2(\mu,\nu)) \leq \phi(\nu) - \phi(\mu).$$

Proof. The implication $\Longleftarrow$ is obvious due to $\omega(W_2^2(\mu,\nu)) = \mathcal{O}(W_2(\mu,\nu))$ as $\nu \to \mu$. Let's prove the other implication. We define $\gamma_t \coloneqq ((1-t)\,\mathrm{id} + t\mathbf{t}_\mu^\nu)_\sharp\mu$ for $t \in [0,1]$. Then γ is a geodesic from μ to ν. Then

$$\begin{aligned}\phi(\gamma_t) &\leq (1-t)\phi(\gamma_0) + t\phi(\gamma_1) \\ &\quad - \frac{\lambda_\omega}{2}\{(1-t)\omega(t^2 W_2^2(\gamma_0,\gamma_1)) + t\omega((1-t)^2 W_2^2(\gamma_0,\gamma_1))\}.\end{aligned}$$

Thus

$$\begin{aligned}\frac{\phi(\gamma_t) - \phi(\gamma_0)}{t} &\leq \phi(\gamma_1) - \phi(\gamma_0) \\ &\quad - \frac{\lambda_\omega}{2}\{(1-t)\frac{\omega(t^2 W_2^2(\gamma_0,\gamma_1))}{t} + \omega((1-t)^2 W_2^2(\gamma_0,\gamma_1))\}.\end{aligned}$$

Because $\lim_{t\downarrow 0} \frac{\omega(t^2 W_2^2(\gamma_0,\gamma_1))}{t} = 0$,

$$\liminf_{t\downarrow 0} \frac{\phi(\gamma_t) - \phi(\gamma_0)}{t} \leq \phi(\gamma_1) - \phi(\gamma_0) - \frac{\lambda_\omega}{2}\omega(W_2^2(\gamma_0,\gamma_1)).$$

Notice that $W_2(\gamma_t, \gamma_0) = tW_2(\gamma_1, \gamma_0)$ and $\mathbf{t}_\mu^{\gamma_t} = (1-t)\,\mathrm{id} + t\mathbf{t}_\mu^\nu$. Fréchet sub-differentiability implies

$$\begin{aligned}\liminf_{t\downarrow 0} \frac{\phi(\gamma_t) - \phi(\gamma_0)}{t} &\geq \liminf_{t\downarrow 0} \frac{1}{t}\int_{\mathbb{R}^d} \langle f, \mathbf{t}_\mu^{\gamma_t} - \mathrm{id}\rangle \,\mathrm{d}\mu \\ &= \int_{\mathbb{R}^d} \langle f, \mathbf{t}_\mu^\nu - \mathrm{id}\rangle \,\mathrm{d}\mu.\end{aligned}$$

The claim then follows. □

Finally, being regular implies several nice properties:

Lemma 2.10. *Let ϕ be regular and satisfy Assumption 2.7.*

1. *[AGS08, Lemma 10.1.5] We have $\mu \in D(|\partial\phi|)$ i.f.f $\partial\phi(\mu) \neq \emptyset$ and*

$$|\partial\phi|(\mu) = \inf_{f\in\partial\phi(\mu)} \|f\|_{L^2(\mu)}. \tag{2.9}$$

   In this case, there exists a unique vector $\bar{f} \in \partial\phi(\mu)$ which attains the minimum in (2.9): we denote it by $\partial^o\phi(\mu)$.

2. *[AGS08, Theorem 11.1.3] We have (μ_t) is a p-curve of maximal slope for ϕ w.r.t $|\partial\phi|$ i.f.f (μ_t) is a gradient flow for ϕ and $(\phi(\mu_t))$ is a.e. equal to a function of bounded variation. In this case,*

$$v_t = -\partial^o\phi(\mu_t) \quad \textit{for a.e.} \quad t > 0.$$

3. *[Cra17, Theorem 3.12 and Remark 2.11] Assume, in addition, that ϕ is ω-convex along generalized geodesics for an Osgood modulus ω of convexity with $\bar{\omega}(s) = o(\sqrt{s})$ as $s \downarrow 0$. For $\mu_0 \in D(\phi)$, there exists a unique gradient flow (μ_t) for ϕ such that $\mu_t \to \mu_0$ as $t \downarrow 0$.*

2.2 Stochastic differential equations

In the rest of the manuscript, (B_t) is a given m-dimensional Brownian motion (m-BM) and $(\mathcal{F}_t)$ is a given admissible filtration (AF) on a given probability space (PS) $(\Omega, \mathcal{A}, \mathbb{P})$. This means for $0 \leq s \leq t$ that B_t is adapted to $\mathcal{F}_t$ and that $B_t - B_s$ is independent of $\mathcal{F}_s$. We assume that $(\Omega, \mathcal{A}, (\mathcal{F}_t), \mathbb{P})$ satisfies the usual conditions.

We fix $T \in (0, \infty)$ and let $\mathbb{T}$ be the interval $[0, T]$. We consider the SDE

$$\begin{cases} \mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t, \\ \nu \text{ is the distribution of } X_0. \end{cases} \tag{2.10}$$

Above, the coefficients

$$\begin{aligned} b &: \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d, \\ \sigma &: \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^m, \end{aligned}$$

are measurable. The study of (2.10) started with theory of stochastic integration created by Itô in 1940's [Itô44; Itô46]. The theory was further developed by Doob [Doo53], Meyer [Mey62; Mey67a; Mey67b; Mey67c; Mey67d], Courrège [Cou63], Kunita-Watanabe [KW67], and Doléans-Dade-Meyer [DM70], among others. We refer to [JP04; Mey09] for a brief history of stochastic calculus and to [Din00; Bic02; Kuo06; Med07; Sok12; Sok13; Pro13; CW13; MP14; MR14] for expositions of stochastic integration. SDEs have many practical applications such as modeling stock prices and random growth phenomena (see e.g. [KP92, Chapter 7]).

Definition 2.11. 1. A *strong solution* to (2.10) is a continuous $\mathbb{R}^d$-valued process (X_t) on $(\Omega, \mathcal{A}, \mathbb{P})$ such that for $t \in \mathbb{T}$: X_t is $\mathcal{F}_t$-adapted and

$$X_t = X_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s \quad \mathbb{P}\text{-a.s.},$$
$$\int_0^t \mathbb{E}[|b(s, X_s)| + |\sigma(s, X_s)|^2]\,\mathrm{d}s < \infty.$$

2. A *weak solution* to (2.10) is a continuous $\mathbb{R}^d$-valued process (X_t) on some PS $(\Omega, \mathcal{A}, \mathbb{P})$ where there exist some m-BM (B_t) and some AF $(\mathcal{F}_t)$ such that the conditions in (1.) are satisfied.

3. SDE (2.10) has *strong uniqueness* if, whenever the PS, the AF and the m-BM are fixed, two strong solutions (X_t) and (Y_t) such that $X_0 = Y_0$ coincide $\mathbb{P}$-a.s. on the path space $C(\mathbb{T}; \mathbb{R}^d)$. SDE (2.10) has *weak uniqueness* if two weak solutions with the same initial distribution induce the same distribution on $C(\mathbb{T}; \mathbb{R}^d)$.

4. SDE (2.10) is *strongly well-posed* if it has strong solution and strong uniqueness. SDE (2.10) is *weakly well-posed* if it has weak solution and weak uniqueness. SDE (2.10) is *well-posed* if it is both strongly and weakly well-posed.

Let $a := \sigma\sigma^\top$. We denote $b_t := b(t, \cdot), \sigma_t := \sigma(t, \cdot)$ and $a_t := a(t, \cdot)$.

2.2.1 Heat kernel estimates of transition density

In this section, we mainly recall heat kernel estimates from [MPZ21] and heat semigroup estimates from [Wan23c]. They are essential tools to obtain regularity of marginal density.

Assumption 2.12. 1. a_t is invertible for $t \in \mathbb{T}$.

2. There exist constants $C > 0$ and $\beta \in (0, 1)$ such that for $t \in \mathbb{T}$ and $x, y \in \mathbb{R}^d$:

$$|b_t(0)| + \|a_t\|_\infty + \|a_t^{-1}\|_\infty \le C,$$
$$|b_t(x) - b_t(y)| \le C(1 \vee |x - y|),$$
$$|\sigma_t(x) - \sigma_t(y)| \le C|x - y|^\beta.$$

We gather parameters in Assumption 2.12:

$$\Theta_2 := (d, T, \beta, C).$$

In Section 2.2.1, we assume that (b, σ) satisfies Assumption 2.12. For $(s, x) \in [0, T) \times \mathbb{R}^d$, the SDE

$$\mathrm{d}X^x_{s,t} = b(t, X^x_{s,t})\,\mathrm{d}t + \sigma(t, X^x_{s,t})\,\mathrm{d}B_t, \quad t \in [s, T], X^x_{s,s} = x, \tag{2.11}$$

is weakly well-posed by [MPZ21, Theorem 1.2] and has transition density denoted by $(p^{b,\sigma}_{s,t})_{0 \le s < t \le T}$. This means $p^{b,\sigma}_{s,t}(x, \cdot)$ is the density of $X^x_{s,t}$. The semigroup $(P^{b,\sigma}_{s,t})_{0 \le s < t \le T}$ is defined for $x \in \mathbb{R}^d$ and $f \in L^0_+(\mathbb{R}^d) \cup L^0_b(\mathbb{R}^d)$ by

$$P^{b,\sigma}_{s,t} f(x) := \mathbb{E}[f(X^x_{s,t})] = \int_{\mathbb{R}^d} p^{b,\sigma}_{s,t}(x, y) f(y)\,\mathrm{d}y. \tag{2.12}$$

The associated time-dependent second-order differential operator $(L^{b,\sigma}_t)_{t \in \mathbb{T}}$ is defined for $f \in C^2(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ by

$$L^{b,\sigma}_t f(x) := \langle b_t(x), \nabla f(x) \rangle + \frac{1}{2} \operatorname{tr}(a_t(x) \nabla^2 f(x)).$$

The backward Kolmogorov equation holds, i.e., for $f \in C_b^2(\mathbb{R}^d), x \in \mathbb{R}^d$ and $0 \le s < t \le T$:

$$\partial_s P_{s,t}^{b,\sigma} f(x) + L_s^{b,\sigma} P_{s,t}^{b,\sigma} f(x) = 0. \tag{2.13}$$

As in [MPZ21, Section 1.2], we will construct a family $(\psi_{s,t})_{s,t\in\mathbb{T}}$ of C^∞-diffeomorphisms on $\mathbb{R}^d$. Let $\rho : \mathbb{R}^d \to \mathbb{R}$ be a smooth symmetric density whose support is contained in the unit ball of $\mathbb{R}^d$. We define $\underline{b} : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d$ by $\underline{b}(t, \cdot) := b(t, \cdot) * \rho$ where $*$ is the convolution operator. By [MPZ21, Inequalities (1.9) and (1.10)], it holds for $n \in \mathbb{N}$ that

$$\sup\{\|\nabla^n \underline{b}(t, \cdot)\|_\infty + \|b(t, \cdot) - \underline{b}(t, \cdot)\|_\infty : t \in \mathbb{T}\} < \infty. \tag{2.14}$$

For $(s, x) \in \mathbb{T} \times \mathbb{R}^d$, we consider the ODE

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\psi_{s,t}(x) & = \underline{b}(t, \psi_{s,t}(x)) \quad \text{for} \quad t \in \mathbb{T}, \\ \psi_{s,s}(x) & = x. \end{cases}$$

For $\kappa > 0$, we consider the Gaussian heat kernel defined for $t > 0$ and $x \in \mathbb{R}^d$ by

$$p_t^\kappa(x) := \frac{1}{(\kappa\pi t)^{\frac{d}{2}}} \exp\left(-\frac{|x|^2}{\kappa t}\right).$$

The following results give density and gradient estimates for (2.11):

Theorem 2.13. *Let Assumption 2.12 hold.*

1. *[MPZ21, Theorem 1.2] There exist constants $c, \kappa > 0$ (depending on Θ_2) such that for $i \in \{0, 1, 2\}, 0 \le s < t \le T$ and $x, y \in \mathbb{R}^d$:*

$$|\nabla_x^i p_{s,t}^{b,\sigma}(x, y)| \le c(t - s)^{-\frac{i}{2}} p_{t-s}^\kappa(\psi_{s,t}(x) - y).$$

2. *[MPZ21, Lemma A.1] For $\alpha \in (0, \beta)$, there exist constants $c, \kappa > 0$ (depending on Θ_2, α) such that for $0 \le s < t \le T$ and $x, y, y' \in \mathbb{R}^d$:*

$$\begin{aligned} |\nabla_x p_{s,t}^{b,\sigma}(x, y) - \nabla_x p_{s,t}^{b,\sigma}(x, y')| \le c|y - y'|^\alpha (t - s)^{-\frac{1+\alpha}{2}} \\ \times \{p_{t-s}^\kappa(\psi_{s,t}(x) - y) + p_{t-s}^\kappa(\psi_{s,t}(x) - y')\}. \end{aligned}$$

We define for $\kappa > 0, f \in L_+^0(\mathbb{R}^d) \cup L_b^0(\mathbb{R}^d), x \in \mathbb{R}^d$ and $0 \le s < t \le T$:

$$\begin{aligned} P_t^\kappa f(x) &:= \int_{\mathbb{R}^d} p_t^\kappa(x - y) f(y) \, \mathrm{d}y, \\ \hat{P}_{s,t}^\kappa f(x) &:= \int_{\mathbb{R}^d} p_{t-s}^\kappa(\psi_{s,t}(x) - y) f(y) \, \mathrm{d}y, \\ \tilde{P}_{s,t}^\kappa f(x) &:= \int_{\mathbb{R}^d} p_{t-s}^\kappa(\psi_{s,t}(y) - x) f(y) \, \mathrm{d}y. \end{aligned} \tag{2.15}$$

For brevity, we denote $\frac{q-p}{pq} := \frac{1}{p} - \frac{1}{q}$ for $p, q \in [1, \infty]$. By Young's inequality for convolution, there exists a constant $c > 0$ (depending on d, κ) such that for $t > 0$ and $1 \le p \le \bar{p} \le \infty$:

$$\|P_t^\kappa\|_{L^p \to L^{\bar{p}}} := \sup_{\|f\|_{L^p} \le 1} \|P_t^\kappa f\|_{L^{\bar{p}}} \le ct^{-\frac{d(\bar{p}-p)}{2p\bar{p}}}. \tag{2.16}$$

We recall a class of functions locally integrable in space-time. Let $p, q \in [1, \infty]$. The localized version $\tilde{L}^p(\mathbb{R}^d)$ of $L^p(\mathbb{R}^d)$ is defined by the norm

$$\|f\|_{\tilde{L}^p} := \sup_{x\in\mathbb{R}^d} \|1_{B(x,1)} f\|_{L^p},$$

Above, $B(x,r)$ is the open ball centered at x with radius r. For $0 \le t_0 < t_1 \le T$, we define the Bochner space

$$L^p_q(t_0,t_1) := L^q([t_0,t_1]; L^p(\mathbb{R}^d)).$$

The localized version $\tilde{L}^p_q(t_0,t_1)$ of $L^p_q(t_0,t_1)$ is defined by the norm

$$\|g\|_{\tilde{L}^p_q(t_0,t_1)} := \sup_{x\in\mathbb{R}^d} \|1_{B(x,1)} g\|_{L^p_q(t_0,t_1)}.$$

Then $\|g\|_{\tilde{L}^\infty_\infty(t_0,t_1)} = \|g\|_{L^\infty_\infty(t_0,t_1)}$. It holds for $p,q \in [1,\infty)$ that

$$\|g\|_{L^p_q(t_0,t_1)} = \left(\int_{t_0}^{t_1}\left(\int_{\mathbb{R}^d} |g(s,y)|^p \,\mathrm{d}y\right)^{\frac{q}{p}} \mathrm{d}s\right)^{\frac{1}{q}},$$

$$\|g\|_{\tilde{L}^p_q(t_0,t_1)} = \sup_{x\in\mathbb{R}^d}\left(\int_{t_0}^{t_1}\left(\int_{B(x,1)} |g(s,y)|^p \,\mathrm{d}y\right)^{\frac{q}{p}} \mathrm{d}s\right)^{\frac{1}{q}}.$$

For brevity, we denote

$$L^p_q(t) := L^p_q(0,t), \quad \tilde{L}^p_q(t) := \tilde{L}^p_q(0,t), \quad L^p_q := L^p_q(0,T), \quad \tilde{L}^p_q := \tilde{L}^p_q(0,T).$$

We recall generalizations of (2.16) for dealing with unbounded drift:

Lemma 2.14. *[Wan23c, Lemma 3.2] Let Assumption 2.12 hold.*

1. *There exists a constant $c > 0$ (depending on Θ_2) such that for $0 \le s < t \le T$ and $1 \le p \le \bar{p} \le \infty$:*

$$\|P^\kappa_{t-s}\|_{\tilde{L}^p\to\tilde{L}^{\bar{p}}} + \|\hat{P}^\kappa_{s,t}\|_{\tilde{L}^p\to\tilde{L}^{\bar{p}}} + \|\tilde{P}^\kappa_{s,t}\|_{\tilde{L}^p\to\tilde{L}^{\bar{p}}} \le c(t-s)^{-\frac{d(\bar{p}-p)}{2p\bar{p}}}.$$

2. *There exists a constant $c > 0$ (depending on Θ_2) such that for $0 \le s < t \le T, 1 \le p \le \bar{p} \le \infty$ and $q \in [1,\infty]$:*

$$\|\hat{P}^\kappa_{\cdot,t} f\|_{\tilde{L}^{\bar{p}}_q(t)} + \|\tilde{P}^\kappa_{\cdot,t} f\|_{\tilde{L}^{\bar{p}}_q(t)} \le c\|(t-\cdot)^{-\frac{d(\bar{p}-p)}{2p\bar{p}}} f\|_{\tilde{L}^p_q(t)} \quad \text{for} \quad f \in L^0_+([0,t]\times\mathbb{R}^d),$$

$$\sup_{z\in\mathbb{R}^d} \|g\hat{P}^\kappa_{s,t}\{1_{B(z,1)} f\}\|_{L^1} \le c(t-s)^{-\frac{d(\bar{p}-p)}{2p\bar{p}}} \|g\|_{\tilde{L}^{\bar{p}^*}} \|f\|_{\tilde{L}^p} \quad \text{for} \quad f,g \in L^0_+(\mathbb{R}^d).$$

Proof. We only include the proof of (1.). Let $z \in \mathbb{R}^d, f \in L^0_+(\mathbb{R}^d)$ and $\mathbf{B}_n := \{v = (v_1,\dots,v_d) \in \mathbb{Z}^d : \sum_{i=1}^d |v_i| = n\}$ for $n \in \mathbb{N}$. We write $M_1 \lesssim M_2$ if there exists a constant $c > 0$ (depending on Θ_2) such that $M_1 \le cM_2$. By [Wan23c, Inequality (3.11)], there exists a constant $c_1 > 0$ (depending on Θ_2) such that

$$\|1_{B(z,1)}\hat{P}^\kappa_{s,t} f\|_{L^{\bar{p}}} \lesssim (t-s)^{-\frac{d(\bar{p}-p)}{2p\bar{p}}} \sum_{n=0}^\infty \sum_{v\in\mathbf{B}_n} e^{-\frac{n^2}{c_1(t-s)}} \|1_{B(z+v,d)} f\|_{L^p}.$$

We have $\mathrm{card}(\mathbf{B}_n) \le (2n+1)^d$, so

$$\sum_{n=0}^\infty \sum_{v\in\mathbf{B}_n} e^{-\frac{n^2}{c_1(t-s)}} \le \sum_{n=0}^\infty (2n+1)^d e^{-\frac{n^2}{c_1 T}} \lesssim 1.$$

On the other hand, $\|1_{B(z+v,d)} f\|_{L^p} \lesssim \|f\|_{\tilde{L}^p}$. Then

$$\|\hat{P}^\kappa_{s,t} f\|_{\tilde{L}^{\bar{p}}} \lesssim (t-s)^{-\frac{d(\bar{p}-p)}{2p\bar{p}}} \|f\|_{\tilde{L}^p}.$$

Notice that P^κ_{t-s} is a special case of $\hat{P}^\kappa_{s,t}$ where $\psi_{s,t} = \mathrm{id}$. Then

$$\|P^\kappa_{t-s} f\|_{\tilde{L}^{\bar{p}}} \lesssim (t-s)^{-\frac{d(\bar{p}-p)}{2p\bar{p}}} \|f\|_{\tilde{L}^p}.$$

It remains to prove for $\tilde{P}^\kappa_{s,t}$. By [Wan23c, Inequality (3.2)], there exists a constant $c_2 \ge 1$ (depending on Θ_2) such that

$$\sup\{\|\nabla \psi_{s,t}\|_\infty + \|\nabla \psi^{-1}_{s,t}\|_\infty : 0 \le s \le t \le T\} \le c_2. \tag{2.17}$$

We have

$$\begin{aligned}
\tilde{P}^\kappa_{s,t} f(x) &= \int_{\mathbb{R}^d} p^\kappa_{t-s}(\psi_{s,t}(y) - x) f(y) \,\mathrm{d}y \quad \text{by (2.15)} \\
&\le \int_{\mathbb{R}^d} p^\kappa_{t-s}\Big(\frac{\psi^{-1}_{s,t}(x) - y}{c_2}\Big) f(y) \,\mathrm{d}y \quad \text{by (2.17)} \\
&\lesssim \int_{\mathbb{R}^d} p^{\bar{\kappa}}_{t-s}(\psi^{-1}_{s,t}(x) - y) f(y) \,\mathrm{d}y \quad \text{where } \bar{\kappa} := \kappa |c_2|^2 \\
&= (P^{\bar{\kappa}}_{t-s} f) \circ \psi^{-1}_{s,t}(x). \qquad (2.18)
\end{aligned}$$

It suffices to consider $\bar{p} < \infty$. We have

$$\begin{aligned}
\|1_{B(z,1)} \tilde{P}^\kappa_{s,t} f\|^{\bar{p}}_{L^{\bar{p}}} &\lesssim \int_{B(z,1)} |(P^{\bar{\kappa}}_{t-s} f) \circ \psi^{-1}_{s,t}(x)|^{\bar{p}} \,\mathrm{d}x \quad \text{by (2.18)} \\
&= \int_{\psi^{-1}_{s,t}(B(z,1))} |(P^{\bar{\kappa}}_{t-s} f)(x)|^{\bar{p}} |\det \nabla \psi_{s,t}(x)| \,\mathrm{d}x \qquad (2.19) \\
&\lesssim \int_{\psi^{-1}_{s,t}(B(z,1))} |(P^{\bar{\kappa}}_{t-s} f)(x)|^{\bar{p}} \,\mathrm{d}x \qquad (2.20) \\
&= \|1_{\psi^{-1}_{s,t}(B(z,1))} P^{\bar{\kappa}}_{t-s} f\|^{\bar{p}}_{L^{\bar{p}}} \\
&\lesssim \|P^{\bar{\kappa}}_{t-s} f\|^{\bar{p}}_{\tilde{L}^{\bar{p}}} \quad \text{by (2.17)}.
\end{aligned}$$

Above, (2.19) is due to change of variables formula and (2.20) due to Hadamard's inequality for determinants. Then $\|1_{B(z,1)} \tilde{P}^\kappa_{s,t} f\|_{L^{\bar{p}}} \lesssim \|P^{\bar{\kappa}}_{t-s} f\|_{\tilde{L}^{\bar{p}}}$. This completes the proof. □

2.2.2 Stability estimates of marginal density

The inspiration for Chapter 4 comes from [Wan23c, Lemma 5.1], which is about *uniform* Hölder regularity in space of marginal density. We intended to establish its counterpart for time regularity, so this section was intended as a preparation of auxiliary results leading to the counterpart. Unfortunately, Professor Francesco Russo has pointed out that Wang's proof contains a gap, which is the symmetry $p^\gamma_{s,t}(x,y) = p^\gamma_{s,t}(y,x)$. For the sake of our understanding of related techniques, we decide to keep this section.

In our study, b is not necessarily continuous but only locally integrable in (t,x). To measure the local integrability in space-time, we introduce the class $\mathcal{K}$ of exponent parameters:

$$\mathcal{K} := \left\{ (p,q) \in (2,\infty]^2 : \frac{d}{p} + \frac{2}{q} < 1 \right\}.$$

Assumption 2.15. 1. a_t is invertible for $t \in \mathbb{T}$.

2. There exist measurable maps $b^{(0)} : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d$ and $b^{(1)} : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d$ such that $b_t(x) = b_t^{(0)}(x) + b_t^{(1)}(x)$ for $t \in \mathbb{T}$ and $x \in \mathbb{R}^d$.

3. There exists $1 \le f_0 \in \tilde{L}_{q_0}^{p_0}$ with $(p_0, q_0) \in \mathcal{K}$ such that $|b_t^{(0)}(x)| \le f_0(t,x)$ for $t \in \mathbb{T}$ and $x \in \mathbb{R}^d$.

4. There exist constants $\beta \in (0,1), C > 0$ such that for $t \in \mathbb{T}$ and $x, y \in \mathbb{R}^d$:

$$\begin{aligned} |b_t^{(1)}(x) - b_t^{(1)}(y)| &\le C|x-y|, \\ |b_t^{(1)}(0)| + \|\sigma_t\|_\infty + \|a_t^{-1}\|_\infty &\le C, \\ |\sigma_t(x) - \sigma_t(y)| &\le C|x-y|^\beta. \end{aligned}$$

By (2.14), if b satisfies Assumption 2.12 then it also satisfies Assumption 2.15. We gather parameters in Assumption 2.15:

$$\Theta_3 := (d, T, \beta, C, p_0, q_0, f_0).$$

Remark 2.16. If $f \in \tilde{L}_q^p$ for some $(p,q) \in \mathcal{K}$, then there exists $(\bar{p}, \bar{q}) \in \bar{\mathcal{K}}$ such that $|f|^2 \in \tilde{L}_{\bar{q}}^{\bar{p}}$.

Assumption 2.15 is appealing because it is a general but sufficient condition to obtain Krylov's and Khasminskii's estimates. The class $\bar{\mathcal{K}}$ of exponent parameter is defined by

$$\bar{\mathcal{K}} := \left\{ (p,q) \in (1,\infty)^2 : \frac{d}{p} + \frac{2}{q} < 2 \right\}.$$

Proposition 2.17. *Let (b, σ) satisfy Assumption 2.15 and $(X_t, t \in \mathbb{T})$ be the solution to (2.10). We fix $(p,q) \in \bar{K}$.*

1. (Khasminskii's estimate) *There exist constants $c > 0, k > 1$ (depending on Θ_3, p, q) such that for $0 \le t_0 < t_1 \le T$ and $g \in \tilde{L}_q^p(t_0, t_1)$:*

$$\mathbb{E}\Big[\exp\Big(\int_{t_0}^{t_1} |g(s, X_s)| \, \mathrm{d}s \Big) \Big| \mathcal{F}_{t_0} \Big] \le \exp(c(1 + \|g\|^k_{\tilde{L}_q^p(t_0,t_1)})). \tag{2.21}$$

2. (Krylov's estimate) *For $j \in [1, \infty)$, there exists a constant $c > 0$ (depending on Θ_3, p, q, j) such that for $0 \le t_0 < t_1 \le T$ and $g \in \tilde{L}_q^p(t_0, t_1)$:*

$$\mathbb{E}\Big[\Big(\int_{t_0}^{t_1} |g(s, X_s)| \, \mathrm{d}s \Big)^j \Big| \mathcal{F}_{t_0} \Big] \le c \|g\|^j_{\tilde{L}_q^p(t_0,t_1)}. \tag{2.22}$$

Proof. 1. There exists $\bar{q} \in (1, q)$ such that $(p, \bar{q}) \in \bar{K}$. By [ZY21, Theorem 3.1], there exists a constant $c_1 > 0$ (depending on $\Theta_3, p, \bar{q}$) such that for $0 \le t_0 < t_1 \le T$, stopping time τ and $g \in \tilde{L}_{\bar{q}}^p(t_0, t_1)$:

$$\mathbb{E}\Big[\int_{t_0 \wedge \tau}^{t_1 \wedge \tau} |g(s, X_s)| \, \mathrm{d}s \Big| \mathcal{F}_{t_0} \Big] \le c_1 \|g\|_{\tilde{L}_{\bar{q}}^p(t_0,t_1)}. \tag{2.23}$$

Let $\delta := \frac{1}{\bar{q}} - \frac{1}{q} \in (0,1)$. By Hölder's inequality, it holds for $0 \le t_0 < t_1 \le T$ and $g \in \tilde{L}_q^p(t_0, t_1)$ that

$$\|g\|_{\tilde{L}_{\bar{q}}^p(t_0,t_1)} \le (t_1 - t_0)^\delta \|g\|_{\tilde{L}_q^p(t_0,t_1)}. \tag{2.24}$$

We denote by I_j^n the open interval $(\frac{(j-1)(t_1-t_0)}{n}, \frac{j(t_1-t_0)}{n})$ for $j = 1, \ldots, n$. We fix $g \in \tilde{L}_q^p(t_0,t_1) \subset \tilde{L}_{\bar{q}}^p(t_0,t_1)$. Let $n \geq 2$ be the smallest integer such that

$$\|g\|_{\tilde{L}_{\bar{q}}^p(I_j^n)} \leq \frac{1}{2c_1} \quad \text{for} \quad j = 1, \ldots, n. \tag{2.25}$$

By (2.23) and [XZ20, Lemma 3.5],

$$\mathbb{E}\Big[\exp\Big(\int_{t_0}^{t_1} |g(s,X_s)| \,\mathrm{d}s\Big)\Big|\mathcal{F}_{t_0}\Big] \leq 2^n.$$

By (2.25), there exists $\bar{j} \in \{1, \ldots, n-1\}$ such that

$$\|g\|_{\tilde{L}_{\bar{q}}^p(I_{\bar{j}}^{n-1})} > \frac{1}{2c_1}. \tag{2.26}$$

By (2.24) and (2.26),

$$\Big(\frac{t_1-t_0}{n-1}\Big)^\delta \|g\|_{\tilde{L}_q^p(I_{\bar{j}}^{n-1})} > \frac{1}{2c_1}.$$

Then

$$n < 1 + T(2c_1)^{-\frac{1}{\delta}} \|g\|_{\tilde{L}_q^p(t_0,t_1)}^{1/\delta}.$$

The estimate (2.21) then follows with $k := \frac{1}{\delta}$.

2. We follow an elegant idea from [HW22, Lemma 2.3]. Let $C_j := e^{j-1}$. We define $h : \mathbb{R}_+ \to \mathbb{R}_+$ by $h(r) := |\ln(C_j + r)|^j$. Then h is concave. We have

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\int_{t_0}^{t_1} |g(s,X_s)| \,\mathrm{d}s\Big)^j \Big|\mathcal{F}_{t_0}\Big] \\
&\leq \mathbb{E}\Big[\Big\{\ln\Big(C_j + \exp\Big(\int_{t_0}^{t_1} |g(s,X_s)| \,\mathrm{d}s\Big)\Big)\Big\}^j \Big|\mathcal{F}_{t_0}\Big] \\
&\leq \Big\{\ln\Big(C_j + \mathbb{E}\Big[\exp\Big(\int_{t_0}^{t_1} |g(s,X_s)| \,\mathrm{d}s\Big)\Big|\mathcal{F}_{t_0}\Big]\Big)\Big\}^j \quad \text{by Jensen's inequality} \\
&\leq \{\ln[C_j + \exp(c(1 + \|g\|_{\tilde{L}_q^p(t_0,t_1)}^k))]\}^j.
\end{aligned}$$

Above, the constants $c, k > 0$ are given by (2.21). As a result, there exists a constant $\bar{C}_j > 0$ (depending on c, j) such that

$$\mathbb{E}\Big[\Big(\int_{t_0}^{t_1} |g(s,X_s)| \,\mathrm{d}s\Big)^j \Big|\mathcal{F}_{t_0}\Big] \leq \bar{C}_j(1 + \|g\|_{\tilde{L}_q^p(t_0,t_1)}^k)^j.$$

Replacing g with $\frac{g}{\|g\|_{\tilde{L}_q^p(t_0,t_1)}}$ in above inequality, we obtain

$$\mathbb{E}\Big[\Big(\int_{t_0}^{t_1} |g(s,X_s)| \,\mathrm{d}s\Big)^j \Big|\mathcal{F}_{t_0}\Big] \leq \bar{C}_j 2^j \|g\|_{\tilde{L}_q^p(t_0,t_1)}^j.$$

The estimate (2.22) then follows. This completes the proof.

□

For another proof of Proposition 2.17, see e.g. [Xia+20, Lemma 4.1]. We recall the following result about Lipschitz continuity w.r.t initial data. This result is interesting because the drift b

in Assumption 2.15 is not necessarily continuous. The key ingredient in its proof is Zvonkin's transform that allows one to extract Lipschitz regularity from a locally integrable drift.

Lemma 2.18. *[HW22, Theorem 1.1(2)] Let Assumption 2.15 hold and $p \in [1, \infty)$. Assume, in addition, that*

1. *σ_t is weakly differentiable for $t \in \mathbb{T}$.*
2. *For $i \in \{1, 2, \ldots, l\}$, there exists $1 \le f_i \in \tilde{L}^{p_i}_{q_i}$ with $(p_i, q_i) \in \mathcal{K}$ such that $|\nabla \sigma_t(x)| \le \sum_{i=1}^{l} f_i(t, x)$ for $t \in \mathbb{T}$ and $x \in \mathbb{R}^d$.*

For $i \in \{1, 2\}$, let $(X^i_t, t \in \mathbb{T})$ satisfy (2.10) and ν_i be the distribution of X^i_0. There exists a constant $c > 0$ (depending on $\Theta_3, p, (p_i, q_i, f_i)_{i=1}^{l}$) such that

$$\mathbb{E}[\sup_{t\in\mathbb{T}} |X^1_t - X^2_t|^p] \le c\mathbb{E}[|X^1_0 - X^2_0|^p].$$

Proof. By Assumption 2.15(2), there exist measurable maps $b^{(0)} : \mathbb{T}\times\mathbb{R}^d \to \mathbb{R}^d$ and $b^{(1)} : \mathbb{T}\times\mathbb{R}^d \to \mathbb{R}^d$ such that $b_t(x) = b^{(0)}_t(x) + b^{(1)}_t(x)$. By Assumption 2.15(3), there exists $1 \le f_0 \in \tilde{L}^{p_0}_{q_0}$ with $(p_0, q_0) \in \mathcal{K}$ such that $|b^{(0)}_t(x)| \le f_0(t, x)$. We write $M_1 \lesssim M_2$ if there exists a constant $c > 0$ (depending on Θ_3, p) such that $M_1 \le cM_2$. Let $(L_t, t \in \mathbb{T})$ be the time-dependent differential operator defined by

$$L_t v := \langle b_t, \nabla v\rangle + \frac{1}{2}\operatorname{tr}(a_t \nabla^2 v)$$

for any function $v : \mathbb{R}^d \to \mathbb{R}$ with two weak derivatives. By Zvonkin transform [ZY21, Theorem 2.1], there exists a constant $\lambda_0 > 0$ (depending on Θ_3) such that for $j \in \{1, \ldots, d\}$ the PDE (in distributional sense)

$$\begin{cases} (\partial_t + L_t)u^{(j)} &= \lambda_0 u^{(j)} - b^{0,j}, \\ u^j(T, \cdot) &= 0, \end{cases}$$

has a unique solution $u^{(j)} : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}$ with the properties

$$\|\nabla^2 u^{(j)}\|_{\tilde{L}^{p_0}_{q_0}} \lesssim 1 \quad \text{and} \quad \|u^{(j)}\|_\infty + \|\nabla u^{(j)}\|_\infty \le \frac{1}{2}. \tag{2.27}$$

Above, $b^{0,j}$ is the j-th row of $b^{(0)}$. We define $u : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d$ by $u := (u^{(1)}, \ldots, u^{(d)})^\top$. By (2.27),

$$\begin{aligned} \|\nabla^2 u\|_{\tilde{L}^{p_0}_{q_0}} &\lesssim 1, \\ \|u\|_\infty + \|\nabla u\|_\infty &\le \frac{1}{2}. \end{aligned} \tag{2.28}$$

Above, ∇u and $\nabla^2 u$ are the first and second-order weak derivatives (w.r.t spatial variable) of u. In particular, ∇u is the Jacobian matrix of u. By generalized Itô's lemma [ZY21, Lemma 3.3],

$$\begin{aligned} \mathrm{d}u^{(j)}(t, X^i_t) &= \{\partial_t + L_t\}u^{(j)}(t, X^i_t)\,\mathrm{d}t + \{(\nabla u^{(j)})^\top \sigma\}(t, X^i_t)\,\mathrm{d}B_t \\ &= \{\lambda_0 u^{(j)} - b^{0,j}\}(t, X^i_t)\,\mathrm{d}t + \{(\nabla u^{(j)})^\top \sigma\}(t, X^i_t)\,\mathrm{d}B_t, \end{aligned}$$

which can be written in matrix form as

$$\mathrm{d}u(t, X^i_t) = \{\lambda_0 u - b^{(0)}\}(t, X^i_t)\,\mathrm{d}t + \{(\nabla u)\sigma\}(t, X^i_t)\,\mathrm{d}B_t. \tag{2.29}$$

Let $Y^i_t := X^i_t + u(t, X^i_t)$. By (2.10) and (2.29),

$$\mathrm{d}Y^i_t = \{\lambda_0 u + b^{(1)}\}(t, X^i_t)\,\mathrm{d}t + \{(I_d + \nabla u)\sigma\}(t, X^i_t)\,\mathrm{d}B_t.$$

By maximal inequality (see e.g. [Xia+20, Lemma 2.1(i)]),

$$
\begin{aligned}
|\sigma(t,x)-\sigma(t,y)| &\lesssim |x-y|\{\mathcal{M}|\nabla\sigma_t|(x)+\mathcal{M}|\nabla\sigma_t|(y)+\|\sigma\|_\infty\}\\
&\lesssim |x-y|\sum_{i=1}^{l}\{\mathcal{M}f_i(t,\cdot)(x)+\mathcal{M}f_i(t,\cdot)(y)\}, \qquad (2.30)\\
|\nabla u(t,x)-\nabla u(t,y)| &\lesssim |x-y|\{\mathcal{M}|\nabla^2 u(t,\cdot)|(x)+\mathcal{M}|\nabla^2 u(t,\cdot)|(y)+\|\nabla u\|_\infty\}\\
&\lesssim |x-y|\{1+\mathcal{M}|\nabla^2 u(t,\cdot)|(x)+\mathcal{M}|\nabla^2 u(t,\cdot)|(y)\}. \qquad (2.31)
\end{aligned}
$$

Above, $\mathcal{M}$ is the local Hardy–Littlewood maximal operator defined for $g:\mathbb{R}^d\to\mathbb{R}_+$ by $\mathcal{M}g(x) \coloneqq \sup_{r\in(0,1)}\frac{1}{|B(0,r)|}\int_{B(0,r)} g(x+y)\,\mathrm{d}y$. Let

$$
\begin{aligned}
H_t &\coloneqq \{\lambda_0 u+b^{(1)}\}(t,X_t^1)-\{\lambda_0 u+b^{(1)}\}(t,X_t^2),\\
G_t &\coloneqq \{(I_d+\nabla u)\sigma\}(t,X_t^1)-\{(I_d+\nabla u)\sigma\}(t,X_t^2),\\
h_t &\coloneqq \sum_{j=1}^{2}\mathcal{M}|\nabla^2 u(t,\cdot)|(X_t^j)+\sum_{j=1}^{2}\sum_{i=1}^{l}\mathcal{M}f_i(t,\cdot)(X_t^j).
\end{aligned}
$$

Recall that $b^{(1)}$ is Lipschitz in space uniformly in time. By (2.28),

$$
|H_t|\lesssim |X_t^1-X_t^2|. \qquad (2.32)
$$

By (2.28), (2.30) and (2.31),

$$
|G_t|\lesssim |X_t^1-X_t^2|(1+h_t). \qquad (2.33)
$$

Let $Z_t \coloneqq Y_t^1-Y_t^2$. Then $\mathrm{d}Z_t=H_t\,\mathrm{d}t+G_t\,\mathrm{d}B_t$. We consider $g:\mathbb{R}^d\to\mathbb{R}_+$ defined by $g(x)\coloneqq |x|^{2p}$. Then $\nabla g(x)=2p|x|^{2(p-1)}x$ and $\nabla^2 g(x)=2p\{2(p-1)|x|^{2(p-2)}xx^\top+|x|^{2(p-1)}I_d\}$. Thus

$$
|\nabla g(x)|\lesssim |x|^{2p-1} \quad\text{and}\quad |\nabla^2 g(x)|\lesssim |x|^{2(p-1)}. \qquad (2.34)
$$

By (2.28),

$$
\frac{1}{2}|X_t^1-X_t^2|\le |Z_t|\le \frac{3}{2}|X_t^1-X_t^2|. \qquad (2.35)
$$

By Itô's lemma,

$$
\begin{aligned}
\mathrm{d}|Z_t|^{2p} &= [(\nabla g(Z_t))^\top H_t+\frac{1}{2}\operatorname{tr}\{\nabla^2 g(Z_t)G_tG_t^\top\}]\,\mathrm{d}t+(\nabla g(Z_t))^\top G_t\,\mathrm{d}B_t\\
&\lesssim |Z_t|^{2p}(1+|h_t|^2)\,\mathrm{d}t+(\nabla g(Z_t))^\top G_t\,\mathrm{d}B_t \qquad (2.36)\\
&\le |Z_t|^{2p}\,\mathrm{d}A_t+\mathrm{d}M_t. \qquad (2.37)
\end{aligned}
$$

Above, $A_t\coloneqq\int_0^t(1+|h_s|^2)\,\mathrm{d}s$ and $M_t\coloneqq\int_0^t(\nabla g(Z_s))^\top G_s\,\mathrm{d}B_s$. Also, (2.36) follows from (2.32), (2.33), (2.34) and (2.35). Then

$$
|X_t^1-X_t^2|^{2p}\lesssim |X_0^1-X_0^2|^{2p}+\int_0^t|X_s^1-X_s^2|^{2p}\,\mathrm{d}A_s+M_t.
$$

Let $\eta_t\coloneqq\sup\{|X_s^1-X_s^2|^p : s\in[0,t]\}$ for $t\in\mathbb{T}$. By (2.37) and stochastic Gronwall's lemma (see e.g. [Xia+20, Lemma 3.7]), there exists a constant $c_1>0$ (depending on p) such that

$$
\begin{aligned}
\mathbb{E}[\eta_T|\mathcal{F}_0] &\lesssim \left\{\mathbb{E}\left[\exp\left(c_1\int_0^T(1+|h_t|^2)\,\mathrm{d}t\right)\middle|\mathcal{F}_0\right]\right\}^{c_1}|X_0^1-X_0^2|^p\\
&\coloneqq \gamma^{c_1}|X_0^1-X_0^2|^p.
\end{aligned}
$$

We have

$$\begin{aligned}\gamma &\lesssim \mathbb{E}\Big[\exp\Big(c_1\int_0^T |h_t|^2\,\mathrm{d}t\Big)\Big|\mathcal{F}_0\Big]\\
&\lesssim \mathbb{E}\Big[\exp\Big(c_1\int_0^T\Big\{\sum_{j=1}^2\{\mathcal{M}|\nabla^2 u(t,\cdot)|\}^2(X_t^j)+\sum_{j=1}^2\sum_{i=1}^l\{\mathcal{M}f_i(t,\cdot)\}^2(X_t^j)\Big\}\,\mathrm{d}t\Big)\Big|\mathcal{F}_0\Big]\\
&\lesssim \sum_{j=1}^2\mathbb{E}\Big[\exp\Big(c_1\int_0^T\{\mathcal{M}|\nabla^2 u(t,\cdot)|\}^2(X_t^j)\,\mathrm{d}t\Big)\Big|\mathcal{F}_0\Big]\\
&\quad+\sum_{j=1}^2\sum_{i=1}^l\mathbb{E}\Big[\exp\Big(c_1\int_0^T\{\mathcal{M}f_i(t,\cdot)\}^2(X_t^j)\,\mathrm{d}t\Big)\Big|\mathcal{F}_0\Big]\quad\text{by AM–GM inequality}\\
&=: \sum_{j=1}^2 I_j+\sum_{j=1}^2\sum_{i=1}^l K_{i,j}.\end{aligned}$$

Recall that $(p_i,q_i)\in\mathcal{K}$ and thus $(\frac{p_i}{2},\frac{q_i}{2})\in\bar{\mathcal{K}}$ for $i\in\{1,\dots,l\}$. By Proposition 2.17(1), there exists a constant $c_2>0$ (depending on Θ_3) such that

$$\begin{aligned}I_j &\le \exp(c_2(1+\|\{\mathcal{M}|\nabla^2 u(t,\cdot)|\}^2\|_{\tilde{L}^{p_0/2}_{q_0/2}}^{c_2}))\\
&=\exp(c_2(1+\|\mathcal{M}|\nabla^2 u(t,\cdot)|\|_{\tilde{L}^{p_0}_{q_0}}^{2c_2}))\\
&\lesssim \exp(c_2(1+\|\nabla^2 u(t,\cdot)\|_{\tilde{L}^{p_0}_{q_0}}^{2c_2}))\quad\text{by [Xia+20, Lemma 2.1(ii)].}\end{aligned}$$

By Proposition 2.17(1), there exists a constant $c_3>0$ (depending on Θ_3) such that

$$\begin{aligned}K_{i,j} &\le \exp(c_3(1+\|\{\mathcal{M}f_i(t,\cdot)\}^2\|_{\tilde{L}^{p_i/2}_{q_i/2}}^{c_3}))\\
&=\exp(c_3(1+\|\mathcal{M}f_i(t,\cdot)\|_{\tilde{L}^{p_i}_{q_i}}^{2c_3}))\\
&\lesssim \exp(c_3(1+\|f_i(t,\cdot)\|_{\tilde{L}^{p_i}_{q_i}}^{2c_3}))\quad\text{by [Xia+20, Lemma 2.1(ii)].}\end{aligned}$$

Then $\mathbb{E}[\eta_T]\lesssim\mathbb{E}[|X_0^1-X_0^2|^p]$. This completes the proof. □

We recall Duhamel presentation and L^p estimate of marginal density:

Theorem 2.19. *[Wan23c, Lemma 4.1] Let Assumption 2.15 hold. Let ν admit a density ℓ_ν and ℓ_t be the density of X_t in (2.10).*

1. *For $p\in(1,\infty]$, there exists a constant $c>0$ (depending on Θ_3,p) such that for $t\in(0,T]$ and $1\le k\le\bar{k}\le\infty$:*
$$\|\ell_t\|_{\tilde{L}^{(p\bar{k}^*)^*}}\le ct^{-\frac{d(\bar{k}-k)}{2k\bar{k}p}}\|\ell_\nu\|_{\tilde{L}^k}^{\frac{1}{p}}.$$

2. *Let $v:\mathbb{T}\times\mathbb{R}^d\to\mathbb{R}^d$ be another drift such that (v,σ) satisfies Assumption 2.12. We assume there exists $1\le g\in\tilde{L}^{\bar{p}}_{\bar{q}}$ with $(\bar{p},\bar{q})\in\bar{\mathcal{K}}$ such that $|b_t(x)-v_t(x)|\le g(t,x)$ for $t\in\mathbb{T}$ and $x\in\mathbb{R}^d$. It holds for $t\in\mathbb{T}$ and $x\in\mathbb{R}^d$:*
$$\begin{aligned}\ell_t(x)&=\int_{\mathbb{R}^d}p_{0,t}^{v,\sigma}(y,x)\ell_\nu(y)\,\mathrm{d}y\\
&\quad+\int_0^t\int_{\mathbb{R}^d}\ell_s(y)\langle b_s(y)-v_s(y),\nabla_y p_{s,t}^{v,\sigma}(y,x)\rangle\,\mathrm{d}y\,\mathrm{d}s.\end{aligned}$$

3. *There exists a constant $c > 0$ (depending on Θ_3) such that for $k \in [p_0^*, \infty]$:*

$$\sup_{t \in \mathbb{T}} \|\ell_t\|_{\tilde{L}^k} \le c \|\ell_\nu\|_{\tilde{L}^k}.$$

The existence of (ℓ_t) is guaranteed by Theorem 4.5(1). We remind that p_0 is a parameter in Θ_3. Recall that r^* is the Hölder conjugate of $r \in [1, \infty]$.

Proof. We write $M_1 \lesssim M_2$ if there exists a constant $c > 0$ (depending on Θ_3, p) such that $M_1 \le c M_2$. By Assumption 2.15(4), $(b^{(1)}, \sigma)$ satisfies Assumption 2.12. By Theorem 2.13(1), there exist a constant $\kappa > 0$ (depending on Θ_3) and a family $(\psi_{s,t})_{0 \le s < t \le T}$ of C^∞-diffeomorphisms on $\mathbb{R}^d$ such that for $i \in \{0, 1\}, 0 \le s < t \le T$ and $x, y \in \mathbb{R}^d$:

$$|\nabla_x^i p_{s,t}^{b^{(1)},\sigma}(x, y)| \lesssim (t - s)^{-\frac{i}{2}} p_{t-s}^\kappa (\psi_{s,t}(x) - y). \tag{2.38}$$

1. It suffices to consider $p \in (1, \infty)$. Then $p^* \in (1, \infty)$. Let R_T and $(\bar{X}_t, t \in \mathbb{T})$ be as in the proof of Theorem 4.5. Let $K_r := \mathbb{E}[|R_T|^r]$ for $r \in \mathbb{R}$. As in the proof of [FF13, Proposition 2.4], we have $K_r < \infty$. Notice that K_r depends on Θ_3, r. Let $z \in \mathbb{R}^d$ and $f \in L_+^0(\mathbb{R}^d)$. We have

$$\begin{aligned}
\int_{\mathbb{R}^d} \{\ell_t 1_{B(z,1)} f\}(x) \, \mathrm{d}x &= \bar{\mathbb{E}}[(1_{B(z,1)} f)(\bar{X}_t)] \quad \text{by (4.7)} \\
&= \mathbb{E}[R(1_{B(z,1)} f)(\bar{X}_t)] \quad \text{by Girsanov's theorem} \\
&\le |K_{p^*}|^{\frac{1}{p^*}} (\mathbb{E}[\{1_{B(z,1)} |f|^p\}(\bar{X}_t)])^{\frac{1}{p}} \quad \text{by Hölder's inequality} \\
&\lesssim (\mathbb{E}[\{1_{B(z,1)} |f|^p\}(\bar{X}_t)])^{\frac{1}{p}} \\
&= \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} p_{0,t}^{b^{(1)},\sigma}(x, y) \{1_{B(z,1)} |f|^p\}(y) \ell_\nu(x) \, \mathrm{d}x \, \mathrm{d}y \right)^{\frac{1}{p}} \quad \text{by (4.6)} \\
&\lesssim \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} p_t^\kappa(\psi_{0,t}(x) - y) \{1_{B(z,1)} |f|^p\}(y) \ell_\nu(x) \, \mathrm{d}x \, \mathrm{d}y \right)^{\frac{1}{p}} \quad \text{by (2.38)} \\
&= \|\ell_\nu \hat{P}_{0,t}^\kappa \{1_{B(z,1)} |f|^p\}\|_{L^1}^{\frac{1}{p}}.
\end{aligned}$$

Let $1 \le k \le \bar{k} \le \infty$. Applying Lemma 2.14(2) with $(p, \bar{p}) = (\bar{k}^*, k^*)$, we have

$$\|\ell_\nu \hat{P}_{0,t}^\kappa \{1_{B(z,1)} |f|^p\}\|_{L^1} \lesssim t^{-\frac{d(\bar{k}-k)}{2k\bar{k}}} \|\ell_\nu\|_{\tilde{L}^k} \||f|^p\|_{\tilde{L}^{\bar{k}^*}}.$$

Then

$$\int_{\mathbb{R}^d} \{\ell_t 1_{B(z,1)} f\}(x) \, \mathrm{d}x \lesssim t^{-\frac{d(\bar{k}-k)}{2k\bar{k}p}} \|\ell_\nu\|_{\tilde{L}^k}^{\frac{1}{p}} \|f\|_{\tilde{L}^{p\bar{k}^*}}.$$

The claim then follows from duality.

2. Notice that $(p_{s,t}^{v,\sigma})_{0 \le s < t \le T}$ is well-defined. Let $f \in C_c^\infty(\mathbb{R}^d)$. Applying Itô's lemma on $[0, t] \times \mathbb{R}^d \to \mathbb{R}$, $(s, x) \mapsto (P_{s,t}^{v,\sigma} f)(x)$, we have

$$\begin{aligned}
\mathrm{d}\{(P_{s,t}^{v,\sigma} f)(X_s)\} &= \{(\partial_s + L_s^{b,\sigma})(P_{s,t}^{v,\sigma} f)\}(X_s) \, \mathrm{d}s + \mathrm{d}M_s \\
&= \{(\partial_s + L_s^{v,\sigma}) + (L_s^{b,\sigma} - L_s^{v,\sigma})\}(P_{s,t}^{v,\sigma} f)(X_s) \, \mathrm{d}s + \mathrm{d}M_s \\
&= \{(L_s^{b,\sigma} - L_s^{v,\sigma})(P_{s,t}^{v,\sigma} f)\}(X_s) \, \mathrm{d}s + \mathrm{d}M_s.
\end{aligned} \tag{2.39}$$

Above, $P^{v,\sigma}_{s,t} f$ is defined by (2.12); $M_0 = 0$ and $\mathrm{d}M_s = \{\nabla (P^{v,\sigma}_{s,t} f)(X_s)\}^\top \sigma(s, X_s) \, \mathrm{d}B_s$ for $s \in [0, t]$; and (2.39) is due to (2.13). Then

$$f(X_t) = (P^{v,\sigma}_{0,t} f)(X_0) + \int_0^t \langle b_s(X_s) - v_s(X_s), \nabla (P^{v,\sigma}_{s,t} f)(X_s)\rangle \, \mathrm{d}s + \int_0^t \{\nabla (P^{v,\sigma}_{s,t} f)(X_s)\}^\top \sigma(s, X_s) \, \mathrm{d}B_s. \tag{2.40}$$

Clearly, f and thus $P^{v,\sigma}_{0,t} f$ are bounded. Let's prove that $\|\nabla (P^{v,\sigma}_{s,t} f)\|_\infty < \infty$. It suffices to verify that $P^{v,\sigma}_{s,t} f$ is Lipschitz. We consider the SDE

$$\mathrm{d}Y^x_{s,t} = v(t, Y^x_{s,t}) \, \mathrm{d}t + \sigma(t, Y^x_{s,t}) \, \mathrm{d}B_s, \quad t \in [s, T], Y^x_{s,s} = x.$$

By Lemma 2.18, there exists a constant $c_1 > 0$ such that

$$\begin{aligned} |P^{v,\sigma}_{s,t} f(x) - P^{v,\sigma}_{s,t} f(y)| &= |\mathbb{E}[f(Y^x_{s,t})] - \mathbb{E}[f(Y^y_{s,t})]| \\ &\le \|\nabla f\|_\infty \mathbb{E}[|Y^x_{s,t} - Y^y_{s,t}|] \\ &\le c_1 \|\nabla f\|_\infty |x - y|. \end{aligned}$$

We have

$$\begin{aligned} &\mathbb{E}\bigg[\int_0^t |b_s(X_s) - v_s(X_s)| \times |\nabla (P^{v,\sigma}_{s,t} f)(X_s)| \, \mathrm{d}s \bigg] \\ &\le \|\nabla (P^{v,\sigma}_{s,t} f)\|_\infty \mathbb{E}\bigg[\int_0^t g(t, X_s) \, \mathrm{d}s \bigg] \\ &\lesssim \|\nabla (P^{v,\sigma}_{s,t} f)\|_\infty \|g\|_{\tilde{L}^{\tilde{p}}_{\tilde{q}}(t)} \quad \text{by Proposition 2.17(2).} \end{aligned}$$

So each term in (2.40) is $\mathbb{P}$-integrable. Then

$$\int_{\mathbb{R}^d} \ell_t(x) f(x) \, \mathrm{d}x = \int_{\mathbb{R}^d} \ell_\nu(x) (P^{v,\sigma}_{0,t} f)(x) \, \mathrm{d}x + \int_0^t \int_{\mathbb{R}^d} \ell_s(x) \langle b_s(x) - v_s(x), \nabla (P^{v,\sigma}_{s,t} f)(x)\rangle \, \mathrm{d}x \, \mathrm{d}s. \tag{2.41}$$

By Theorem 2.13(1) and Leibniz integral rule,

$$\nabla (P^{v,\sigma}_{s,t} f)(x) = \nabla_x \int_{\mathbb{R}^d} p^{v,\sigma}_{s,t}(x, y) f(y) \, \mathrm{d}y = \int_{\mathbb{R}^d} \nabla_x p^{v,\sigma}_{s,t}(x, y) f(y) \, \mathrm{d}y.$$

So (2.41) is equivalent to

$$\begin{aligned} \int_{\mathbb{R}^d} \ell_t(x) f(x) \, \mathrm{d}x &= \int_{\mathbb{R}^d} \bigg(\int_{\mathbb{R}^d} p^{v,\sigma}_{0,t}(y, x) \ell_\nu(y) \, \mathrm{d}y \bigg) f(x) \, \mathrm{d}x \\ &\quad + \int_{\mathbb{R}^d} \bigg(\int_0^t \int_{\mathbb{R}^d} \ell_s(y) \langle b_s(y) - v_s(y), \nabla_y p^{v,\sigma}_{s,t}(y, x)\rangle \, \mathrm{d}y \, \mathrm{d}s \bigg) f(x) \, \mathrm{d}x. \end{aligned}$$

The required representation then follows.

3. By Assumption 2.15(4), $(b^{(1)}, \sigma)$ satisfies Assumption 2.12. By Theorem 2.19(2),

$$\ell_t(x) = \int_{\mathbb{R}^d} p^{b^{(1)},\sigma}_{0,t}(y, x) \ell_\nu(y) \, \mathrm{d}y + \int_0^t \int_{\mathbb{R}^d} \ell_s(y) \langle b^{(0)}_s(y), \nabla_y p^{b^{(1)},\sigma}_{s,t}(y, x)\rangle \, \mathrm{d}y \, \mathrm{d}s. \tag{2.42}$$

By (2.38) and (2.42),

$$\begin{aligned}\ell_t(x) &\lesssim \int_{\mathbb{R}^d} p_t^\kappa(\psi_{0,t}(y)-x)\ell_\nu(y)\,\mathrm{d}y \\ &\qquad + \int_0^t (t-s)^{-\frac{1}{2}} \int_{\mathbb{R}^d} \ell_s(y) f_0(s,y) p_{t-s}^\kappa(\psi_{s,t}(y)-x)\,\mathrm{d}y\,\mathrm{d}s \\ &= \tilde{P}_{0,t}^\kappa \ell_\nu(x) + \int_0^t (t-s)^{-\frac{1}{2}} \tilde{P}_{s,t}^\kappa\{\ell_s f_0(s,\cdot)\}(x)\,\mathrm{d}s \quad \text{by (2.15)}.\end{aligned}$$

Then it holds for $l \in [1,\infty]$ that

$$\begin{aligned}\|\ell_t\|_{\tilde{L}^l} &\lesssim \|\tilde{P}_{0,t}^\kappa \ell_\nu\|_{\tilde{L}^l} + \int_0^t \|\tilde{P}_{s,t}^\kappa\{f_0(s,\cdot)(t-s)^{-\frac{1}{2}}\ell_s\}\|_{\tilde{L}^l}\,\mathrm{d}s \\ &=: \|\tilde{P}_{0,t}^\kappa \ell_\nu\|_{\tilde{L}^l} + I_t.\end{aligned}$$

Applying Lemma 2.14(1) with $(p,\bar{p}) = (l,l)$, we have $\|\tilde{P}_{0,t}^\kappa \ell_\nu\|_{\tilde{L}^l} \lesssim \|\ell_\nu\|_{\tilde{L}^l}$. For convenience, we denote $\ell(s,x) := \ell_s(x)$. We fix $l \in [p_0^*,\infty]$. There exists $q_l \in [1,p_0]$ such that $\frac{1}{q_l} = \frac{1}{p_0} + \frac{1}{l}$. Then $(p_0,q_0) \in \mathcal{K}$ implies

$$\delta := \frac{1}{2} + \frac{d(l-q_l)}{2lq_l} = \frac{1}{2} + \frac{d}{2p_0} < \frac{q_0-1}{q_0} < 1.$$

Applying Lemma 2.14(2) with $(p,\bar{p},q) = (q_l,l,1)$, we have

$$I_t = \|\tilde{P}_{\cdot,t}^\kappa f_0(t-\cdot)^{-\frac{1}{2}}\ell\|_{\tilde{L}_1^l(t)} \lesssim \|f_0(t-\cdot)^{-\delta}\ell\|_{\tilde{L}_1^{q_l}(t)}.$$

By Hölder's inequality,

$$\begin{aligned}\|f_0(t-\cdot)^{-\delta}\ell\|_{\tilde{L}_1^{q_l}(t)} &= \int_0^t \|f_0(s,\cdot)(t-s)^{-\delta}\ell_s\|_{\tilde{L}^{q_l}}\,\mathrm{d}s \\ &\le \int_0^t \|f_0(s,\cdot)\|_{\tilde{L}^{p_0}} \|(t-s)^{-\delta}\ell_s\|_{\tilde{L}^l}\,\mathrm{d}s \\ &\le \left(\int_0^t \|f_0(s,\cdot)\|_{\tilde{L}^{p_0}}^{q_0}\,\mathrm{d}s\right)^{\frac{1}{q_0}} \left(\int_0^t \|(t-s)^{-\delta}\ell_s\|_{\tilde{L}^l}^{q_0^*}\,\mathrm{d}s\right)^{\frac{1}{q_0^*}} \\ &= \|f_0\|_{\tilde{L}_{q_0}^{p_0}(t)} \left(\int_0^t \|(t-s)^{-\delta}\ell_s\|_{\tilde{L}^l}^{q_0^*}\,\mathrm{d}s\right)^{\frac{1}{q_0^*}}.\end{aligned}$$

Then

$$\|\ell_t\|_{\tilde{L}^l} \lesssim \|\ell_\nu\|_{\tilde{L}^l} + \|f_0\|_{\tilde{L}_{q_0}^{p_0}} \left(\int_0^t (t-s)^{-\delta q_0^*} \|\ell_s\|_{\tilde{L}^l}^{q_0^*}\,\mathrm{d}s\right)^{\frac{1}{q_0^*}}.$$

Thus

$$\|\ell_t\|_{\tilde{L}^l}^{q_0^*} \lesssim \|\ell_\nu\|_{\tilde{L}^l}^{q_0^*} + \int_0^t (t-s)^{-\delta q_0^*} \|\ell_s\|_{\tilde{L}^l}^{q_0^*}\,\mathrm{d}s. \tag{2.43}$$

Notice that $(p_0,q_0) \in \mathcal{K}$ implies $\delta q_0^* \in (0,1)$. WLOG, we assume $\ell_\nu \in \tilde{L}^k(\mathbb{R}^d)$.

a) We consider the case $k=\infty$. Let $l \in [p_0^*,\infty)$. We apply Theorem 2.19(1) with $k=\bar{k}=\infty$ and $p=l^* \in (1,p_0]$. Then there exists a constant $c_{1,l}>0$ (depending on Θ_3, l) such that

$$\sup_{t\in\mathbb{T}} \|\ell_t\|_{\tilde{L}^l} \le c_{1,l} \|\ell_\nu\|_{\tilde{L}^\infty}^{\frac{1}{l^*}} < \infty. \tag{2.44}$$

By (2.43), (2.44) and Gronwall's lemma,

$$\sup_{t\in\mathbb{T}} \|\ell_t\|_{\tilde{L}^l} \lesssim \|\ell_\nu\|_{\tilde{L}^l}. \tag{2.45}$$

Taking the limit $l \uparrow \infty$ in (2.45), we have $\sup\{\|\ell_t\|_{\tilde{L}^\infty} : t \in \mathbb{T}\} \lesssim \|\ell_\nu\|_{\tilde{L}^\infty}$.

b) We consider the case $k \in [p_0^*, \infty)$. For $\bar{k} \in (k, \infty)$, let $p_{k,\bar{k}} := \frac{k(\bar{k}-1)}{\bar{k}(k-1)} > 1$ and thus

$$k = (p_{k,\bar{k}}\bar{k}^*)^* \quad \text{and} \quad \lim_{\bar{k}\downarrow k} \frac{d(\bar{k}-k)}{2\bar{k}kp_{k,\bar{k}}} = 0.$$

Recall that $\delta q_0^* \in (0,1)$, so there exists $\bar{k} \in (k,\infty)$ such that

$$\varepsilon_{k,\bar{k}} := \frac{d(\bar{k}-k)}{2\bar{k}kp_{k,\bar{k}}} \in (0, 1-\delta q_0^*).$$

By Theorem 2.19(1), there exists a constant $c_{2,k} > 0$ (depending on Θ_3, k) such that

$$\|\ell_t\|_{\tilde{L}^k} \le c_{2,k} t^{-\varepsilon_{k,\bar{k}}} \|\ell_\nu\|_{\tilde{L}^k}^{1/p_{k,\bar{k}}}. \tag{2.46}$$

By (2.43) with $l = k$ and (2.46),

$$\sup_{t\in\mathbb{T}} \|\ell_t\|_{\tilde{L}^k}^{q_0^*} \lesssim \|\ell_\nu\|_{\tilde{L}^k}^{q_0^*} + c_{2,k}\|\ell_\nu\|_{\tilde{L}^k}^{q_0^*/p_{k,\bar{k}}} \sup_{t\in\mathbb{T}} \int_0^t (t-s)^{-\delta q_0^*} s^{-\varepsilon_{k,\bar{k}}} \,\mathrm{d}s.$$

Because $\delta q_0^* + \varepsilon_{k,\bar{k}} < 1$,

$$\sup_{t\in\mathbb{T}} \int_0^t (t-s)^{-\delta q_0^*} s^{-\varepsilon_{k,\bar{k}}} \,\mathrm{d}s < \infty.$$

Then

$$\sup_{t\in\mathbb{T}} \|\ell_t\|_{\tilde{L}^k}^{q_0^*} < \infty. \tag{2.47}$$

By (2.43) with $l = k$, (2.47) and Gronwall's lemma, $\sup\{\|\ell_t\|_{\tilde{L}^k} : t \in \mathbb{T}\} \lesssim \|\ell_\nu\|_{\tilde{L}^k}$. This completes the proof.

□

2.2.3 Auxiliary lemmas

We include here results that will be used later on. Consider the Fokker-Planck equation

$$\partial_t \mu_t = -\partial_{x_i}\{b^i(t,x)\mu_t\} + \frac{1}{2}\partial_{x_i}\partial_{x_j}\{a^{i,j}(t,x)\mu_t\}. \tag{2.48}$$

Lemma 2.20 (Superposition principle)**.** *Assume that a weakly continuous map $\mathbb{T} \to \mathcal{P}(\mathbb{R}^d)$, $t \mapsto \mu_t$ satisfies $\int_{\mathbb{T}}\int_{\mathbb{R}^d}\{|b(t,x)| + |a(t,x)|\}\,\mathrm{d}\mu_t(x)\,\mathrm{d}t < \infty$ and is a distributional solution to (2.48). Then (2.10) has a weak solution whose distribution of X_t is μ_t for $t \in \mathbb{T}$.*

The above form of Lemma 2.20 is taken from [BR20, Section 2]. We refer to [Fig08; Tre16; RXZ20; BRS21] for studies of superposition principle. We denote by $\overset{*}{\rightharpoonup}$ the weak-$*$ convergence on $\mathcal{P}(\mathbb{R}^d)$. This means $\mu_n \overset{*}{\rightharpoonup} \mu$ i.f.f $\int_{\mathbb{R}^d} f\,\mathrm{d}\mu_n \to \int_{\mathbb{R}^d} f\,\mathrm{d}\mu$ for $f \in C_c(\mathbb{R}^d)$.

Lemma 2.21. *Let $\mu_n, \mu \in \mathcal{P}(\mathbb{R}^d)$ for $n \in \mathbb{N}$. If $\mu_n \overset{*}{\rightharpoonup} \mu$ then $\mu_n \rightharpoonup \mu$.*

Proof. Let $f \in C_b(\mathbb{R}^d)$ and $g \in C_c(\mathbb{R}^d)$ such that $0 \le g \le 1$. Then $gf \in C_c(\mathbb{R}^d)$ and $f = (1-g)f + gf$. We have

$$\left| \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu_n - \mu) \right| \le \|f\|_\infty \int_{\mathbb{R}^d} (1-g) \, \mathrm{d}(\mu_n + \mu) + \left| \int_{\mathbb{R}^d} gf \, \mathrm{d}(\mu_n - \mu) \right|.$$

It follows from $\mu_n \xrightarrow{*} \mu$ that $\lim_n \int_{\mathbb{R}^d} gf \, \mathrm{d}(\mu_n - \mu) = 0$. Then

$$\limsup_n \left| \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu_n - \mu) \right| \le \|f\|_\infty \limsup_n \int_{\mathbb{R}^d} (1-g) \, \mathrm{d}(\mu_n + \mu).$$

Notice that

$$\begin{aligned} \limsup_n \int_{\mathbb{R}^d} (1-g) \, \mathrm{d}\mu_n &= 1 - \liminf_n \int_{\mathbb{R}^d} g \, \mathrm{d}\mu_n \\ &= 1 - \int_{\mathbb{R}^d} g \, \mathrm{d}\mu \quad \text{because } \mu_n \xrightarrow{*} \mu \\ &= \int_{\mathbb{R}^d} (1-g) \, \mathrm{d}\mu. \end{aligned}$$

Thus

$$\limsup_n \left| \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu_n - \mu) \right| \le 2\|f\|_\infty \int_{\mathbb{R}^d} (1-g) \, \mathrm{d}\mu. \tag{2.49}$$

Because μ is a probability measure,

$$\sup \left\{ \int_{\mathbb{R}^d} g \, \mathrm{d}\mu : g \in C_c(\mathbb{R}^d) \text{ and } 0 \le g \le 1 \right\} = 1. \tag{2.50}$$

The claim then follows from (2.49) and (2.50). ∎

See e.g. [HLS22] for a generalization of Lemma 2.21. Finally, we recall freezing lemma.

Lemma 2.22. *Let $\mathcal{D}, \mathcal{G}$ be independent sub-σ-algebras of $\mathcal{A}$. Let $(E, \mathcal{E})$ be a measurable space and $Y : \Omega \to E$ measurable w.r.t $\mathcal{D}$. Assume that $\varphi : E \times \Omega \to \mathbb{R}^d$ is measurable w.r.t $\mathcal{E} \otimes \mathcal{G}$ and that $\varphi(Y, \cdot)$ is integrable.*

1. *Let $N := \{x \in E : \varphi(x, \cdot) \text{ not integrable}\}$. Let μ be the distribution of Y on $(E, \mathcal{E})$. Then $N \in \mathcal{E}$ and $\mu(N) = 0$.*

2. *We define $\Phi : E \to \mathbb{R}^d$ by $\Phi(y) := 0$ for $y \in N$ and $\Phi(y) := \mathbb{E}[\varphi(y, \cdot)]$ for $y \in E \setminus N$. Then Φ is measurable and $\mathbb{E}[\varphi(Y, \cdot) | \mathcal{D}] = \Phi(Y)$.*

Proof. WLOG, we assume $d = 1$ and φ is non-negative. The claims that $N \in \mathcal{E}$ and that Φ is measurable follow from Tonelli theorem, measurability and non-negativity of φ.

1. Let ν be the joint distribution of $(Y, \mathrm{id}) : (\Omega, \mathcal{A}) \to (E \times \Omega, \mathcal{E} \otimes \mathcal{G})$. Let $\mathbb{P}_\mathcal{G}$ be the restriction of $\mathbb{P}$ to $(\Omega, \mathcal{G})$. We have $\nu = \mu \otimes \mathbb{P}_\mathcal{G}$ because it holds for $A \in \mathcal{E}$ and $B \in \mathcal{G}$ that

$$\begin{aligned} \nu(A \times B) &= \mathbb{P}[(Y, \mathrm{id}) \in A \times B] \\ &= \mathbb{P}[\{Y \in A\} \cap \{\mathrm{id} \in B\}] \\ &= \mathbb{P}[Y \in A] \times \mathbb{P}[\mathrm{id} \in B] \quad \text{by independence of } \mathcal{D}, \mathcal{G} \\ &= \mu(Y \in A) \times \mathbb{P}_\mathcal{G}[B]. \end{aligned}$$

Then

$$
\begin{aligned}
\mathbb{E}[|\varphi(Y,\cdot)|] &= \int_{E\times\Omega} |\varphi(y,\omega)| \,\mathrm{d}\nu(y,\omega) \quad \text{because } \varphi \text{ is measurable w.r.t } \mathcal{E}\otimes\mathcal{G} \\
&= \int_E \left(\int_\Omega |\varphi(y,\omega)| \,\mathrm{d}\mathbb{P}_{\mathcal{G}}(\omega) \right) \mathrm{d}\mu(y) \quad \text{by Tonelli theorem} \\
&= \int_E \mathbb{E}[|\varphi(y,\cdot)|] \,\mathrm{d}\mu(y). \qquad (2.51)
\end{aligned}
$$

By the integrability of $\varphi(Y,\cdot)$, it holds for μ-a.e. $y \in E$ that $\mathbb{E}[|\varphi(y,\cdot)|] < \infty$. Thus $\mu(N) = 0$.

2. It remains to prove $\mathbb{E}[\varphi(Y,\cdot)|\mathcal{D}] = \Phi(Y)$. First, we verify that $\Phi(Y)$ is integrable. Indeed,

$$
\begin{aligned}
\mathbb{E}[|\Phi(Y)|] &= \int_E |\Phi(y)| \,\mathrm{d}\mu(y) \\
&= \int_E |\mathbb{E}[\varphi(y,\cdot)]| \,\mathrm{d}\mu(y) \\
&\le \int_E \mathbb{E}[|\varphi(y,\cdot)|] \,\mathrm{d}\mu(y) \\
&= \mathbb{E}[|\varphi(Y,\cdot)|] \quad \text{by (2.51)} \\
&< \infty.
\end{aligned}
$$

Let $D \in \mathcal{D}$. We need to prove $\mathbb{E}[\varphi(Y,\cdot)1_D] = \mathbb{E}[\Phi(Y)1_D]$. It suffices to consider the case $\mathbb{P}[D] > 0$. We define a probability measure $\widetilde{\mathbb{P}}$ on $(\Omega,\mathcal{A})$ by $\widetilde{\mathbb{P}}[A] := \frac{\mathbb{P}[A\cap D]}{\mathbb{P}[D]}$ for $A \in \mathcal{A}$. Let $\widetilde{\mathbb{E}}$ be the expectation w.r.t $(\Omega,\mathcal{A},\widetilde{\mathbb{P}})$. By approximation with simple functions, $\mathbb{E}[Z1_D] = \mathbb{P}[D]\widetilde{\mathbb{E}}[Z]$ for $Z \in L^1(\Omega,\mathcal{A},\widetilde{\mathbb{P}})$. So it suffices to prove $\widetilde{\mathbb{E}}[\varphi(Y,\cdot)] = \widetilde{\mathbb{E}}[\Phi(Y)]$. Let $\widetilde{\mu}$ be the distribution of Y on $(E,\mathcal{E})$ under $\widetilde{\mathbb{P}}$. Let $\widetilde{\nu}$ be the joint distribution of $(Y,\mathrm{id}) : (\Omega,\mathcal{A}) \to (E\times\Omega, \mathcal{E}\otimes\mathcal{G})$ under $\widetilde{\mathbb{P}}$. Let $\widetilde{\mathbb{P}}_{\mathcal{G}}$ be the restriction of $\widetilde{\mathbb{P}}$ to $(\Omega,\mathcal{G})$. Let's prove that $\widetilde{\nu} = \widetilde{\mu}\otimes\widetilde{\mathbb{P}}_{\mathcal{G}}$. Indeed, it holds for $A \in \mathcal{E}$ and $B \in \mathcal{G}$ that

$$
\begin{aligned}
\widetilde{\nu}(A\times B) &= \widetilde{\mathbb{P}}[(Y,\mathrm{id}) \in A\times B] \\
&= \widetilde{\mathbb{P}}[\{Y\in A\} \cap \{\mathrm{id}\in B\}] \\
&= \frac{\mathbb{P}[(\{Y\in A\}\cap D) \cap \{\mathrm{id}\in B\}]}{\mathbb{P}[D]} \\
&= \frac{\mathbb{P}[\{Y\in A\}\cap D] \times \mathbb{P}[\{\mathrm{id}\in B\}\cap D]}{(\mathbb{P}[D])^2} \quad \text{by independence of } \mathcal{D},\mathcal{G} \text{ under } \mathbb{P} \\
&= \widetilde{\mathbb{P}}[Y\in A] \times \widetilde{\mathbb{P}}[\mathrm{id}\in B] \\
&= \widetilde{\mu}(Y\in A) \times \widetilde{\mathbb{P}}_{\mathcal{G}}[B].
\end{aligned}
$$

Then

$$\begin{aligned}
\widetilde{\mathbb{E}}[\varphi(Y,\cdot)] &= \int_{E\times\Omega} \varphi(y,\omega)\,\mathrm{d}\widetilde{\nu}(y,\omega) \quad \text{because } \varphi \text{ is measurable w.r.t } \mathcal{E}\otimes\mathcal{G} \\
&= \int_E \left(\int_\Omega \varphi(y,\omega)\,\mathrm{d}\widetilde{\mathbb{P}}_{\mathcal{G}}(\omega) \right) \mathrm{d}\widetilde{\mu}(y) \quad \text{by Fubini theorem} \\
&= \int_E \widetilde{\mathbb{E}}[\varphi(y,\cdot)]\,\mathrm{d}\widetilde{\mu}(y), \\
\widetilde{\mathbb{E}}[\varphi(y,\cdot)] &= \frac{\mathbb{E}[\varphi(y,\cdot)1_D]}{\mathbb{P}[D]} \\
&= \frac{\mathbb{E}[\varphi(y,\cdot)] \times \mathbb{E}[1_D]}{\mathbb{P}[D]} \quad \text{by independence of } \mathcal{D} \text{ and } \mathcal{G} \text{ under } \mathbb{P} \\
&= \mathbb{E}[\varphi(y,\cdot)] = \Phi(y).
\end{aligned}$$

Thus $\widetilde{\mathbb{E}}[\varphi(Y,\cdot)] = \int_E \Phi(y)\,\mathrm{d}\widetilde{\mu}(y) = \widetilde{\mathbb{E}}[\Phi(Y)]$. This completes the proof.

□

Chapter 3

A class of Langevin dynamics

3.1 Introduction

Let $V : \mathbb{R}^d \to \mathbb{R}$ be differentiable and $\pi \in \mathcal{P}(\mathbb{R}^d)$ such that $\mathrm{d}\pi = e^{-V}\,\mathrm{d}x$. This means e^{-V} is the density of π w.r.t Lebesgue measure on $\mathbb{R}^d$. We define $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ by $\varphi(0) := 0$ and $\varphi(s) := s \log s$ for $s > 0$. For $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, their (non-negative) *relative entropy* is defined as

$$\mathcal{H}(\mu|\nu) := \begin{cases} \int_{\mathbb{R}^d} \varphi \circ \varrho \,\mathrm{d}\nu & \text{if} \quad \varrho = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}, \\ \infty & \text{otherwise.} \end{cases}$$

For $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, its *internal energy* is defined as

$$\mathcal{H}(\mu) := \begin{cases} \int_{\mathbb{R}^d} \varphi \circ \varrho \,\mathrm{d}x & \text{if} \quad \varrho = \frac{\mathrm{d}\mu}{\mathrm{d}x}, \\ \infty & \text{otherwise.} \end{cases}$$

If V has at most quadratic growth at infinity, then [ABS21, Proposition 15.6] implies

$$\phi(\mu) := \mathcal{H}(\mu|\pi) = \int_{\mathbb{R}^d} V \,\mathrm{d}\mu + \mathcal{H}(\mu) \quad \text{for} \quad \mu \in \mathcal{P}_2(\mathbb{R}^d).$$

Consider

$$\alpha := \inf\{\phi(\mu) : \mu \in \mathcal{P}_2(\mathbb{R}^d)\}. \tag{3.1}$$

We have explained in Section 1.1 that (3.1) gives rise to the PDE

$$\partial_t \mu_t = \operatorname{div}(\mu_t \nabla V) + \Delta \mu_t. \tag{3.2}$$

For well-posedness of (3.2), we assume the following set of assumptions:

Assumption 3.1.

1. $V : \mathbb{R}^d \to \mathbb{R}$ is differentiable with $\int_{\mathbb{R}^d} e^{-V}\,\mathrm{d}x = 1$.
2. There exists a constant $C > 0$ such that $V(x) \geq -C$ and $|V(x)| + |\nabla V(x)|^2 \leq C(1 + |x|^2)$ for $x \in \mathbb{R}^d$.
3. There exists a continuous non-decreasing and concave function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ such that
   a) $\psi(0) = 0$, $\int_0^1 \frac{\mathrm{d}s}{\psi(s)} = \infty$ and $\psi(s) \geq s$ for $s \geq 0$.
   b) $|\nabla V(x) - \nabla V(y)|^2 \leq 4C^2 \psi(|x - y|^2)$ for $x, y \in \mathbb{R}^d$.

For $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, their relative *Fisher information* is defined by

$$\mathcal{I}(\mu|\nu) := \begin{cases} 4 \int_{\mathbb{R}^d} |\nabla(\sqrt{\varrho})|^2 \,\mathrm{d}\nu & \text{if} \quad \varrho = \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \quad \text{with} \quad \sqrt{\varrho} \in W^{1,2}(\mathbb{R}^d), \\ \infty & \text{otherwise.} \end{cases}$$

For rate of convergence, we assume a log-Sobolev inequality.

Assumption 3.2. There exists a constant $\kappa > 0$ such that

$$\mathcal{H}(\mu|\pi) \leq \frac{\mathcal{I}(\mu|\pi)}{2\kappa} \quad \text{for} \quad \mu \in \mathcal{P}(\mathbb{R}^d).$$

We recall from [Chi22] two conditions under which π satisfies Assumption 3.2. The first by [BÉ85] is when $\nabla^2 V \succeq \kappa I_d$. The second by [HS87] is when $\pi = e^{-\psi}\nu$ for some $\psi \in L^\infty(\mathbb{R}^d)$ and some $\nu \in \mathcal{P}(\mathbb{R}^d)$ satisfying Assumption 3.2.

Our result is the following

Theorem 3.3. *Let Assumption 3.1 hold and $\mu_0 \in D(\phi)$. There exists a unique distributional solution (μ_t) to (3.2) such that $\mu_t \to \mu_0$ in $\mathcal{P}_2(\mathbb{R}^d)$ as $t \downarrow 0$. If, in addition, Assumption 3.2 holds, then $\phi(\mu_t) - \alpha \leq e^{-2\kappa t}\{\phi(\mu_0) - \alpha\}$ for $t \geq 0$.*

The rest of this chapter is dedicated to the proof of Theorem 3.3.

3.2 Convexity along generalized geodesics

We define $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ by $\omega(s) \coloneqq \sqrt{s\psi(s)}$. It follows from $\lim_{s\downarrow 0} \frac{\omega(s)}{\sqrt{s}} = 0$ that $\omega(s) = o(\sqrt{s})$ as $s \downarrow 0$. We have ψ is a modulus of convexity, so is ω. Let's prove that ω is concave. For $x, y \in \mathbb{R}_+$ and $t \in [0,1]$,

$$\begin{aligned} & \omega((1-t)x + ty) \geq (1-t)\omega(x) + t\omega(y) \\ \iff & \sqrt{((1-t)x+ty)\psi((1-t)x+ty)} \geq (1-t)\sqrt{x\psi(x)} + t\sqrt{y\psi(y)} \\ \iff & ((1-t)x+ty)\psi((1-t)x+ty) \geq (1-t)^2 x\psi(x) + t^2 y\psi(y) \\ & \qquad + 2t(1-t)\sqrt{xy\psi(x)\psi(y)}. \end{aligned}$$

Above, the last inequality is true due to the concavity of ψ and AM-GM inequality. Let's prove that ω is an Osgood modulus of continuity of itself. Let $x \geq y \geq 0$. It suffices to prove $\omega(x) - \omega(y) \leq \omega(x-y)$ or equivalently $\omega(x) \leq \omega(x-y) + \omega(y)$. This is true because ω is concave with $\omega(0) = 0$.

We define $\mathcal{V} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ by $\mathcal{V}(\mu) \coloneqq \int_{\mathbb{R}^d} V \, \mathrm{d}\mu$. By Assumption 3.1(2), $\mathcal{V}$ is well-defined. By Lemma 2.4, $\mathcal{V}$ is ω-convex along *any* interpolating curve with constant $\lambda_\omega = -4C < 0$. By [AGS08, Proposition 9.3.9], $\mathcal{H}$ is convex along generalized geodesics. Then $\phi = \mathcal{V} + \mathcal{H}$ is ω-convex along generalized geodesics.

3.3 Existence of a solution

It is clear that ϕ is proper. By [AGS08, Lemma 9.4.3], ϕ is lower semi-continuous. The proximal map $J_t\phi$ at $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ is non-empty for all $t > 0$. This is due to the fact that ϕ is non-negative and lower semi-continuous. By Lemma 2.8, ϕ is regular. By Lemma 2.10(3), there exists a unique gradient flow $\mu \in \mathrm{AC}^2_{\mathrm{loc}}((0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ for ϕ such that $\mu_t \to \mu_0$ as $t \downarrow 0$. We fix $g \in C_c^\infty(\mathbb{R}^d)$. By Definition 2.5(3), there exists a time-dependent Borel vector field v with the property: it holds for a.e. $t > 0$ that $v_t \in -\partial\phi(\mu_t)$ and

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} g \, \mathrm{d}\mu_t = \int_{\mathbb{R}^d} \langle \nabla g, v_t \rangle \, \mathrm{d}\mu_t. \tag{3.3}$$

By Lemma 2.9, it holds for a.e. $t > 0$ that

$$\int_{\mathbb{R}^d} \langle v_t, \mathrm{id} - \mathbf{t}_{\mu_t}^{\nu} \rangle \, \mathrm{d}\mu_t + \frac{\lambda_\omega}{2} \omega(W_2^2(\nu, \mu_t)) \leq \phi(\nu) - \phi(\mu_t) \quad \text{for} \quad \nu \in D(\phi). \tag{3.4}$$

We fix $t > 0$ such that (3.3) and (3.4) hold. Let $f_s := \mathrm{id} + s\nabla g$ for $s \in \mathbb{R}$. Then $f_s = \nabla(\frac{1}{2}|\cdot|^2 + sg)$. There exists $\delta \in (0,1)$ such that $\frac{1}{2}|\cdot|^2 + sg$ is convex for $|s| \le \delta$. By Brenier theorem (see e.g. [Vil03, Theorem 2.12]), f_s is the unique optimal map that transports μ_t to $\nu_s := (f_s)_\sharp \mu_t$ for $|s| \le \delta$. Hence

$$W_2(\nu_s, \mu_t) = \begin{cases} \frac{|s|}{\delta} W_2(\mu_t, \nu_\delta) & \text{if} \quad s \in [0, \delta], \\ \frac{|s|}{\delta} W_2(\mu_t, \nu_{-\delta}) & \text{if} \quad s \in [-\delta, 0]. \end{cases}$$

Recall that $\omega(s) = \mathcal{O}(\sqrt{s})$ as $s \downarrow 0$, so

$$\lim_{s \to 0} \frac{\omega(W_2^2(\nu_s, \mu_t))}{s} = 0. \tag{3.5}$$

By (3.4),

$$\begin{aligned} -\int_{\mathbb{R}^d} \langle \nabla g, v_t \rangle \,\mathrm{d}\mu_t + \frac{\lambda_\omega}{2} \frac{\omega(W_2^2(\nu_s, \mu_t))}{s} &\le \frac{\phi(\nu_s) - \phi(\mu_t)}{s} \quad \text{for} \quad s \in [0, \delta], \\ -\int_{\mathbb{R}^d} \langle \nabla g, v_t \rangle \,\mathrm{d}\mu_t + \frac{\lambda_\omega}{2} \frac{\omega(W_2^2(\nu_s, \mu_t))}{s} &\ge \frac{\phi(\nu_s) - \phi(\mu_t)}{s} \quad \text{for} \quad s \in [-\delta, 0]. \end{aligned} \tag{3.6}$$

By Taylor formula,

$$\begin{aligned} & V \circ f_s(x) - V(x) \\ &= s\langle \nabla V(x), \nabla g(x) \rangle + s \int_0^1 \langle \nabla V(x + rs\nabla g(x)) - \nabla V(x), \nabla g(x) \rangle \,\mathrm{d}r. \end{aligned}$$

We have

$$\begin{aligned} & \frac{\mathcal{V}(\nu_s) - \mathcal{V}(\mu_t)}{s} \\ &= \int_{\mathbb{R}^d} \frac{V \circ f_s - V}{s} \,\mathrm{d}\mu_t \quad \text{by change of variables formula} \\ &= \int_{\mathbb{R}^d} \langle \nabla g, \nabla V \rangle \,\mathrm{d}\mu_t + \int_{\mathbb{R}^d} \int_0^1 \langle \nabla V(x + rs\nabla g(x)) - \nabla V(x), \nabla g(x) \rangle \,\mathrm{d}r \,\mathrm{d}\mu_t(x). \end{aligned}$$

We have

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} \int_0^1 \langle \nabla V(x + rs\nabla g(x)) - \nabla V(x), \nabla g(x) \rangle \,\mathrm{d}r \,\mathrm{d}\mu_t(x) \right| \\ &\le \int_{\mathbb{R}^d} \int_0^1 |\nabla V(x + rs\nabla g(x)) - \nabla V(x)| \times |\nabla g(x)| \,\mathrm{d}r \,\mathrm{d}\mu_t(x) \\ &\le 2C \|\nabla g\|_\infty \sqrt{\psi(s\|\nabla g\|_\infty^2)} \quad \text{by Assumption 3.1(3)}. \end{aligned}$$

We have ψ is continuous with $\psi(0) = 0$, so

$$\lim_{s \to 0} \frac{\mathcal{V}(\nu_s) - \mathcal{V}(\mu_t)}{s} = \int_{\mathbb{R}^d} \langle \nabla g, \nabla V \rangle \,\mathrm{d}\mu_t.$$

By [Mag23, Theorem 13.9(iii)],

$$\lim_{s \to 0} \frac{\mathcal{H}(\nu_s) - \mathcal{H}(\mu_t)}{s} = -\int_{\mathbb{R}^d} \Delta g \,\mathrm{d}\mu_t.$$

Then

$$\lim_{s \to 0} \frac{\phi(\nu_s) - \phi(\mu_t)}{s} = \int_{\mathbb{R}^d} \{\langle \nabla g, \nabla V \rangle - \Delta g\} \,\mathrm{d}\mu_t. \tag{3.7}$$

By (3.5), (3.6) and (3.7),

$$\int_{\mathbb{R}^d} \langle \nabla g, v_t \rangle \, \mathrm{d}\mu_t = \int_{\mathbb{R}^d} \{\Delta g - \langle \nabla g, \nabla V \rangle\} \, \mathrm{d}\mu_t. \tag{3.8}$$

By (3.3) and (3.8),

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} g \, \mathrm{d}\mu_t = \int_{\mathbb{R}^d} \{\Delta g - \langle \nabla g, \nabla V \rangle\} \, \mathrm{d}\mu_t.$$

Then (μ_t) is indeed a distributional solution to (3.2).

3.4 Uniqueness of a solution

We consider the SDE

$$\mathrm{d}X_t = -\nabla V(X_t) \, \mathrm{d}t + \sqrt{2} \, \mathrm{d}B_t. \tag{3.9}$$

Because ∇V is continuous, it is locally integrable. By [KR05, Theorem 2.1], (3.9) has weak uniqueness. Another way to obtain this is by noticing $\sup_{x \in \mathbb{R}^d} \frac{|\nabla V(x)|}{1+|x|} < \infty$ and applying [BL08, Theorem 2]. Let (ν_t) be another distributional solution to (3.2) such that (ν_t) is locally bounded and $\nu_t \to \mu_0$ as $t \downarrow 0$. We have

1. $\mu_t, \nu_t \in \mathcal{P}(\mathbb{R}^d)$ for $t \geq 0$.
2. $x \mapsto \nabla V(x)$ is measurable and
$$\int_0^T \int_{\mathbb{R}^d} |\nabla V| \, \mathrm{d}\{\mu_t + \nu_t\} \, \mathrm{d}t < \infty \quad \text{for} \quad T \geq 0.$$
3. $t \mapsto \mu_t$ and $t \mapsto \nu_t$ are weakly continuous.

By Lemma 2.20, (3.9) has two weak solutions $(X_t, t \geq 0)$ and $(Y_t, t \geq 0)$ such that the distribution of X_t is μ_t and that of Y_t is ν_t. The weak uniqueness of (3.9) implies $(\mu_t) = (\nu_t)$.

3.5 Regularity of a solution

In the rest of this chapter, "it holds a.e." means "it holds for almost every $t > 0$". From the previous section, it holds a.e. that $\mu_t \in D(\phi)$. Indeed, we can obtain stronger result.

Lemma 3.4. *$(\phi(\mu_t))$ is locally absolutely continuous.*

Proof. By [AGS08, Theorem 8.3.1], it holds a.e. that $|\mu'|(t) \leq \|v_t\|_{L^2(\mu_t)}$. By [AGS08, Lemma 10.1.5], it holds a.e. that $|\partial\phi|(\mu_t) \leq \|v_t\|_{L^2(\mu_t)}$. By definition of a gradient flow, $t \mapsto \|v_t\|_{L^2(\mu_t)}$ belongs to $L^2_{\mathrm{loc}}(0, \infty)$. Then

$$t \mapsto |\partial\phi|(\mu_t)|\mu'|(t) \quad \text{belongs to} \quad L^1_{\mathrm{loc}}(0, \infty). \tag{3.10}$$

We fix $0 < a < b < \infty$. Let $\mathcal{P}_0 := \{\mu_t : t \in [a, b]\}$. Then $(\mathcal{P}_0, W_2)$ is compact. We need to prove $[a, b] \ni t \mapsto \phi(\mu_t)$ is absolutely continuous. For $t \in [a, b]$, the global slope $\mathfrak{l}^0_\phi(\mu_t)$ of $\phi : \mathcal{P}_0 \to (-\infty, \infty]$ is defined as

$$\mathfrak{l}^0_\phi(\mu_t) := \sup_{\nu \in \mathcal{P}_0 \setminus \{\mu_t\}} \frac{(\phi(\mu_t) - \phi(\nu))^+}{W_2(\mu_t, \nu)}. \tag{3.11}$$

By [Cra17, Proposition 2.7],

$$|\partial\phi|(\mu_t) = \sup_{\nu \neq \mu_t} \left(\frac{\phi(\mu_t) - \phi(\nu)}{W_2(\mu_t, \nu)} + \frac{\lambda_\omega \omega(W_2^2(\mu_t, \nu))}{W_2(\mu_t, \nu)} \right)^+. \tag{3.12}$$

By [AGS08, Theorem 1.2.5], $\mathfrak{l}_\phi^0$ is a strong upper gradient for $\phi : \mathcal{P}_0 \to (-\infty, \infty]$. By (2.7), it suffices to prove that $[a, b] \ni t \mapsto \mathfrak{l}_\phi^0(\mu_t)|\mu'|(t)$ is integrable. Let $R \coloneqq \operatorname{diam}(\mathcal{P}_0) < \infty$. The map $(0, R] \ni s \mapsto \frac{\omega(s^2)}{s}$ is continuous and $\lim_{s\downarrow 0} \frac{\omega(s^2)}{s} = 0$. So $S \coloneqq \sup_{s\in(0,R]} \frac{\omega(s^2)}{s} < \infty$. By (3.11), (3.12) and the fact that $\lambda_\omega < 0$, it holds for $t \in [a, b]$ that

$$\mathfrak{l}_\phi^0(\mu_t) \le |\partial\phi|(\mu_t) - \lambda_\omega S. \tag{3.13}$$

The claim then follows from (3.10) and (3.13). □

As a consequence of Lemma 3.4, it holds a.e. that μ_t admits a density, which will be denoted by ρ_t. We will verify that φ satisfies *doubling condition* (10.4.23) in [AGS08, Lemma 10.4.4].

Lemma 3.5. *There exists a constant $c > 0$ such that $\varphi(s + r) \le c\{1 + \varphi(s) + \varphi(r)\}$ for $s, r \ge 0$.*

Proof. The minimizer of φ is $\frac{1}{e}$ and its minimum value is $-\frac{1}{e} \le -\frac{2}{5}$. Then $1 + \varphi(s) + \varphi(r) \ge \frac{1}{5}$ for $s, r \in \mathbb{R}_+$. It holds for $2 \le s \le r$ that $\varphi(s + r) = (s + r)\ln(s + r) \le 2r\ln(2r) \le 2r\ln(r^2) = 4r\ln r \le 4\{1 + \varphi(s) + \varphi(r)\}$. Let $c_1 \coloneqq 5\max\{\varphi(s + r) : 0 \le s \le r \le 2\}$. The claim then follows by picking $c \coloneqq 4 \vee c_1$. □

Next we obtain differentiability of ρ_t and explicit form of $\partial^\circ\phi(\mu_t)$.

Lemma 3.6. *We fix $t > 0$ such that $\mu_t \in D(|\partial\phi|)$. Then $\rho_t \in W^{1,1}(\mathbb{R}^d)$ and $\rho_t\nabla V + \nabla\rho_t = \rho_t w$ a.e. on $\mathbb{R}^d$ for some vector field $w \in L^2(\mu_t)$. In this case, $w = \partial^\circ\phi(\mu_t)$.*

Proof. 1. We fix $g \in C_c^\infty(\mathbb{R}^d; \mathbb{R}^d)$. Let $f_s \coloneqq \mathrm{id} + sg$ and $\nu_s \coloneqq (f_s)_\sharp\mu_t$ for $s \in \mathbb{R}$. By (3.7),

$$\lim_{s\to 0} \frac{\phi(\nu_s) - \phi(\mu_t)}{s} = \int_{\mathbb{R}^d} \{\langle g, \nabla V\rangle - \operatorname{div} g\}\, \mathrm{d}\mu_t.$$

It holds for $s \ge 0$ that $W_2(\mu_t, \nu_s) \le s\|g\|_{L^2(\mu_t)}$. We have

$$\begin{aligned}\lim_{s\downarrow 0} \frac{\phi(\nu_s) - \phi(\mu_t)}{s} &\le \limsup_{s\downarrow 0} \frac{(\phi(\nu_s) - \phi(\mu_t))^+}{s} \\ &\le \|g\|_{L^2(\mu_t)} \limsup_{s\downarrow 0} \frac{(\phi(\nu_s) - \phi(\mu_t))^+}{W_2(\mu_t, \nu_s)} \\ &\le \|g\|_{L^2(\mu_t)}|\partial\phi|(\mu_t) \quad \text{by definition of } |\partial\phi|(\mu_t).\end{aligned}$$

Then

$$\int_{\mathbb{R}^d} \{\langle g, \nabla V\rangle - \operatorname{div} g\}\, \mathrm{d}\mu_t \le \|g\|_{L^2(\mu_t)}|\partial\phi|(\mu_t). \tag{3.14}$$

By Cauchy–Schwarz inequality,

$$-\int_{\mathbb{R}^d} \operatorname{div} g\, \mathrm{d}\mu_t \le \{|\partial\phi|(\mu_t) + \|\nabla V\|_{L^2(\mu_t)}\}\|g\|_{L^2(\mu_t)}. \tag{3.15}$$

By Assumption 3.1(2), $\|\nabla V\|_{L^2(\mu_t)} < \infty$. Then ρ_t is a function of bounded variation. Thus its distributional derivative $D\rho_t = (D_1\rho_t, \ldots, D_d\rho_t)$ is a finite $\mathbb{R}^d$-valued Radon measure on $\mathbb{R}^d$. Then (3.15) is equivalent to

$$\sum_{i=1}^d \int_{\mathbb{R}^d} g_i\, \mathrm{d}\{D_i\rho_t\} \le \{|\partial\phi|(\mu_t) + \|\nabla V\|_{L^2(\mu_t)}\}\|g\|_{L^2(\mu_t)}.$$

By Riesz representation theorem, there exists a unique vector field $z \in L^2(\mu_t)$ such that

$$\sum_{i=1}^d \int_{\mathbb{R}^d} g_i \, \mathrm{d}\{D_i \rho_t\} = \int_{\mathbb{R}^d} \langle g, z \rangle \, \mathrm{d}\mu_t.$$

Then $\rho_t \in W^{1,1}(\mathbb{R}^d)$ with its weak gradient $\nabla \rho_t = \rho_t z$. Thus (3.14) is equivalent to

$$\int_{\mathbb{R}^d} \langle g, \rho_t \nabla V + \nabla \rho_t \rangle \, \mathrm{d}x \le \|g\|_{L^2(\mu_t)} |\partial \phi|(\mu_t).$$

By Riesz representation theorem, there exits a unique vector field $w \in L^2(\mu_t)$ such that $\|w\|_{L^2(\mu_t)} \le |\partial \phi|(\mu_t)$ and

$$\int_{\mathbb{R}^d} \langle g, \rho_t \nabla V + \nabla \rho_t \rangle \, \mathrm{d}x = \int_{\mathbb{R}^d} \langle g, w \rangle \, \mathrm{d}\mu_t.$$

In particular,

$$\rho_t \nabla V + \nabla \rho_t = \rho_t w \quad \text{a.e. on } \mathbb{R}^d.$$

2. To prove $w = \partial^o \phi(\mu_t)$, it remains to check $w \in \partial \phi(\mu_t)$. There exist vector fields $w_1, w_2 \in L^2(\mu_t)$ such that $\rho_t \nabla V = \rho_t w_1$ and $\nabla \rho_t = \rho_t w_2$. It holds for μ_t-a.e. that $w = w_1 + w_2$. It suffices to check $w_1 \in \partial \mathcal{V}(\mu_t)$ and $w_2 \in \partial \mathcal{H}(\mu_t)$. We fix $\nu \in D(\phi)$. Let $\mathbf{t} := \mathbf{t}_{\mu_t}^{\nu}$ be the unique optimal map that transports μ_t to ν. By Lemma 2.9, it suffices to prove

$$\int_{\mathbb{R}^d} \langle w_1, \mathbf{t} - \mathrm{id} \rangle \, \mathrm{d}\mu_t + \frac{\lambda_\omega}{2} \omega(W_2^2(\mu_t, \nu)) \le \mathcal{V}(\nu) - \mathcal{V}(\mu_t), \tag{3.16}$$

$$\int_{\mathbb{R}^d} \langle w_2, \mathbf{t} - \mathrm{id} \rangle \, \mathrm{d}\mu_t \le \mathcal{H}(\nu) - \mathcal{H}(\mu_t). \tag{3.17}$$

Let γ be the unique constant-speed geodesic from μ_t to ν. Then $\gamma_s = ((1-s)\,\mathrm{id} + s\mathbf{t})_\sharp \mu_t$ for $s \in [0,1]$.

a) Next we will verify (3.16). Recall that $\mathcal{V}$ is ω-convex along any interpolating curve, so it holds for $s \in [0,1]$ that

$$\begin{aligned}\mathcal{V}(\gamma_s) &\le (1-s)\mathcal{V}(\gamma_0) + s\mathcal{V}(\gamma_1) \\ &\quad - \frac{\lambda_\omega}{2}\{(1-s)\omega(s^2 W_2^2(\gamma_0,\gamma_1)) + s\omega((1-s)^2 W_2^2(\gamma_0,\gamma_1))\}.\end{aligned}$$

Then

$$\begin{aligned}\mathcal{V}(\gamma_1) - \mathcal{V}(\gamma_0) &\ge \frac{\mathcal{V}(\gamma_s) - \mathcal{V}(\gamma_0)}{s} \\ &\quad + \frac{\lambda_\omega}{2}\{(1-s)\frac{\omega(s^2 W_2^2(\gamma_0,\gamma_1))}{s} + \omega((1-s)^2 W_2^2(\gamma_0,\gamma_1))\}.\end{aligned}$$

Thus

$$\mathcal{V}(\gamma_1) - \mathcal{V}(\gamma_0) \ge \lim_{s \downarrow 0} \frac{\mathcal{V}(\gamma_s) - \mathcal{V}(\gamma_0)}{s} + \frac{\lambda_\omega}{2}\omega(W_2^2(\gamma_0,\gamma_1)). \tag{3.18}$$

By Leibniz integral rule,

$$\lim_{s \downarrow 0} \frac{\mathcal{V}(\gamma_s) - \mathcal{V}(\gamma_0)}{s} = \int_{\mathbb{R}^d} \langle \nabla V, \mathbf{t} - \mathrm{id} \rangle \, \mathrm{d}\gamma_0. \tag{3.19}$$

Clearly, (3.18) and (3.19) imply (3.16).

b) Next we will verify (3.17). By Lemma 3.5 and as in [AGS08, Theorem 10.4.6], there exists a sequence $(\nu_n) \subset D(\phi)$ of measures with compact support such that $\nu_n \to \nu$ and $\mathcal{H}(\nu_n) \to \mathcal{H}(\nu)$. By [Vil09, Corollary 5.23], we can assume WLOG that $\mathbf{t}_{\mu_t}^{\nu_n} \to \mathbf{t}$ in probability measure μ_t as $n \to \infty$. By Vitali convergence theorem (see e.g. [Bre11, Exercise 4.14]),

$$\int_{\mathbb{R}^d} \langle w_2, \mathbf{t}_{\mu_t}^{\nu_n} - \mathrm{id} \rangle \, \mathrm{d}\mu_t \xrightarrow{n\to\infty} \int_{\mathbb{R}^d} \langle w_2, \mathbf{t} - \mathrm{id} \rangle \, \mathrm{d}\mu_t.$$

By [AGS08, Lemma 5.4.1], we indeed have stronger convergence

$$\int_{\mathbb{R}^d} |\mathbf{t}_{\mu_t}^{\nu_n} - \mathbf{t}|^2 \, \mathrm{d}\mu_t \xrightarrow{n\to\infty} 0.$$

Hence it suffices to consider ν with compact support. Recall that $\mathcal{H}$ is convex along generalized geodesics. Similar to (3.18), we have

$$\mathcal{H}(\gamma_1) - \mathcal{H}(\gamma_0) \geq \lim_{s\downarrow 0} \frac{\mathcal{H}(\gamma_s) - \mathcal{H}(\gamma_0)}{s}. \tag{3.20}$$

As in the proof of [AGS08, Proposition 9.3.9], the requirements of [AGS08, Lemma 10.4.4] are satisfied. By [AGS08, Equation (10.4.22)],

$$\lim_{s\downarrow 0} \frac{\mathcal{H}(\gamma_s) - \mathcal{H}(\gamma_0)}{s} = -\int_{\mathbb{R}^d} \mathrm{tr}(\tilde{\nabla}(\mathbf{t} - \mathrm{id})) \, \mathrm{d}\gamma_0. \tag{3.21}$$

Above, $\tilde{\nabla}(\mathbf{t} - \mathrm{id})$ is the approximate differential of $\mathbf{t} - \mathrm{id}$ in the sense of [AGS08, Definition 5.5.1]. By [AGS08, Lemma 10.4.5] and the compactness of $\operatorname{supp} \nu$, there exists an increasing sequence of Lipschitz functions $g_n : \mathbb{R}^d \to [0,1]$ with compact support such that (g_n) converges to 1 pointwise on $\mathbb{R}^d$ and

$$-\int_{\mathbb{R}^d} \mathrm{tr}(\tilde{\nabla}(\mathbf{t} - \mathrm{id})) \, \mathrm{d}\gamma_0 \geq \limsup_{n\to\infty} \int_{\mathbb{R}^d} g_n \langle \nabla \rho_t, \mathbf{t} - \mathrm{id} \rangle \, \mathrm{d}x.$$

Then

$$\begin{aligned} -\int_{\mathbb{R}^d} \mathrm{tr}(\tilde{\nabla}(\mathbf{t} - \mathrm{id})) \, \mathrm{d}\gamma_0 &\geq \limsup_{n\to\infty} \int_{\mathbb{R}^d} g_n \langle \rho_t w_2, \mathbf{t} - \mathrm{id} \rangle \, \mathrm{d}x \\ &= \limsup_{n\to\infty} \int_{\mathbb{R}^d} g_n \langle w_2, \mathbf{t} - \mathrm{id} \rangle \, \mathrm{d}\mu_t \\ &= \int_{\mathbb{R}^d} \langle w_2, \mathbf{t} - \mathrm{id} \rangle \, \mathrm{d}\mu_t. \end{aligned} \tag{3.22}$$

Clearly, (3.20), (3.21) and (3.22) imply (3.17).

□

3.6 Rate of convergence

By Lemma 2.10(2) and Lemma 3.4, (μ_t) is 2-curve of maximal slope w.r.t $|\partial\phi|$, and it holds a.e. that $v_t = -\partial^o \phi(\mu_t)$. A monotone function is differentiable a.e. (see e.g. [Tao11, Theorem 1.6.25]). By (2.8), it holds a.e. that $\mu_t \in D(|\partial\phi|)$. By Lemma 3.6, it holds a.e. that $\rho_t \in W^{1,1}(\mathbb{R}^d)$ and $\rho_t \nabla V + \nabla \rho_t = -\rho_t v_t$ a.e. on $\mathbb{R}^d$. In particular, it holds a.e. that $v_t = -(\nabla V + \frac{\nabla \rho_t}{\rho_t})$ μ_t-a.e. on

$\mathbb{R}^d$. We have $\mathrm{d}\pi = e^{-V}\,\mathrm{d}x$ and $\mathrm{d}\mu_t = \rho_t\,\mathrm{d}x$, so $g_t \coloneqq \frac{\mathrm{d}\mu_t}{\mathrm{d}\pi} = e^V \rho_t$. Then

$$\begin{aligned}\mathcal{I}(\mu_t|\pi) &= \int_{\{g_t>0\}} \frac{|\nabla g_t|^2}{g_t}\,\mathrm{d}\pi \\ &= \|\nabla(\ln g_t)\|^2_{L^2(\mu_t)} \\ &= \|v_t\|^2_{L^2(\mu_t)}.\end{aligned}$$

By chain rule [AS07, Equation (4.55)], it holds a.e. that

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi(\mu_t) = -\|v_t\|^2_{L^2(\mu_t)} = -\mathcal{I}(\mu_t|\pi).$$

By Assumption 3.2, $\mathcal{I}(\mu_t|\pi) \geq 2\kappa\mathcal{H}(\mu_t|\pi) = 2\kappa\phi(\mu_t)$. Let (ν_n) be a minimizing sequence of (3.1), i.e., $\phi(\nu_n) \downarrow \alpha$ as $n \to \infty$. Notice that $\phi \geq 0$, so it holds a.e. that

$$\frac{\mathrm{d}}{\mathrm{d}t}\{\phi(\mu_t) - \phi(\nu_n)\} \leq -2\kappa\{\phi(\mu_t) - \phi(\nu_n)\}.$$

By Gronwall's lemma, it holds for $t > 0$ that $\phi(\mu_t) - \phi(\nu_n) \leq e^{-2\kappa t}\{\phi(\mu_0) - \phi(\nu_n)\}$. Taking the limit $n \to \infty$, we have $\phi(\mu_t) - \alpha \leq e^{-2\kappa t}\{\phi(\mu_0) - \alpha\}$ for $t > 0$. This completes the proof.

Chapter 4

Well-posedness of MV-SDEs with density-dependent drift

4.1 Introduction

Let $T \in (0, \infty)$ and $\mathbb{T}$ be the interval $[0, T]$. Let

$$b : \mathbb{T} \times \mathbb{R}^d \times \mathbb{R}_+ \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d,$$
$$\sigma : \mathbb{T} \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^m,$$

be measurable. We consider the SDE

$$\begin{cases} \mathrm{d}X_t = b(t, X_t, \ell_t(X_t), \mu_t)\,\mathrm{d}t + \sigma(t, X_t, \mu_t)\,\mathrm{d}B_t, \\ \nu \text{ is the distribution of } X_0,\ \mu_t \text{ is that of } X_t, \\ \text{and } \ell_t \text{ is the density of } X_t. \end{cases} \tag{4.1}$$

We recall notions of a solution:

Definition 4.1. 1. A *strong solution* to (4.1) is a continuous $\mathbb{R}^d$-valued process (X_t) on $(\Omega, \mathcal{A}, \mathbb{P})$ such that for $t \in \mathbb{T}$: X_t is $\mathcal{F}_t$-adapted, X_t has a distribution $\mu_t \in \mathcal{P}_p(\mathbb{R}^d)$, X_t admits a density ℓ_t, and

$$X_t = X_0 + \int_0^t b(s, X_s, \ell_s(X_s), \mu_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s, \mu_s)\,\mathrm{d}B_s \quad \mathbb{P}\text{-a.s.},$$
$$\int_0^t \mathbb{E}[|b(s, X_s, \ell_s(X_s), \mu_s)| + |\sigma(s, X_s, \mu_s)|^2]\,\mathrm{d}s < \infty.$$

2. A *weak solution* to (4.1) is a continuous $\mathbb{R}^d$-valued process (X_t) on some PS $(\Omega, \mathcal{A}, \mathbb{P})$ on which there exist some m-BM (B_t) and some AF $(\mathcal{F}_t)$ such that the conditions in (1) are satisfied.

3. SDE (4.1) has *strong uniqueness* if, whenever the PS, the AF and the m-BM are fixed, two strong solutions (X_t) and (Y_t) such that $X_0 = Y_0$ coincide $\mathbb{P}$-a.s. on the path space $C(\mathbb{T}; \mathbb{R}^d)$. SDE (4.1) has *weak uniqueness* if two weak solutions with the same initial distribution induce the same distribution on $C(\mathbb{T}; \mathbb{R}^d)$.

4. SDE (4.1) is *strongly well-posed* if it has strong solution and strong uniqueness. SDE (4.1) is *weakly well-posed* if it has weak solution and weak uniqueness. SDE (4.1) is *well-posed* if it is both strongly and weakly well-posed.

Recall that $M_p(\varrho)$ is the p-th moment of $\varrho \in \mathcal{P}(\mathbb{R}^d)$. Let $a := \sigma\sigma^\top$. We denote $b_t(x,r,\varrho) := b(t,x,r,\varrho), \sigma_t(x,\varrho) := \sigma(t,x,\varrho)$ and $a_t(x,\varrho) := a(t,x,\varrho)$. Below, we introduce the main assumption about initial distribution and coefficients of (4.1).

Assumption 4.2. There exist constants $\beta \in (0,1), p \in [1,\infty), C > 0$ such that for $t \in \mathbb{T}; x, y \in \mathbb{R}^d; r, r' \in \mathbb{R}_+$ and $\varrho, \varrho' \in \mathcal{P}_p(\mathbb{R}^d)$:

1. a_t is invertible and σ_t is weakly differentiable w.r.t space variable.
2. There exists $1 \le f_0 \in \tilde{L}^{p_0}_{q_0}$ with $(p_0, q_0) \in \mathcal{K}$ such that $|b_t(x,r,\varrho)| \le f_0(t,x)$.
3. There exists $1 \le f_i \in \tilde{L}^{p_i}_{q_i}$ with $i \in \{1,2,\ldots,l\}$ and $(p_i,q_i) \in \mathcal{K}$ such that $|\nabla\sigma_t(x)| \le \sum_{i=1}^l f_i(t,x)$.
4. $\nu \in \mathcal{P}_p(\mathbb{R}^d)$ has a density $\ell_\nu \in L^\infty(\mathbb{R}^d)$.
5. The following conditions hold:

$$\begin{aligned} \|\sigma_t\|_\infty + \|a_t^{-1}\|_\infty &\le C, \\ |b_t(x,r,\varrho) - b_t(x,r',\varrho)| &\le C\{|r-r'| + W_p(\varrho,\varrho')\}, \\ |\sigma_t(x,\varrho) - \sigma_t(y,\varrho')| &\le C\{|x-y|^\beta + W_p(\varrho,\varrho')\}. \end{aligned}$$

We remind that the space $\tilde{L}^{p_0}_{q_0}$ and the set $\mathcal{K}$ have been defined in Section 2.2.1 and Section 2.2.2 respectively. We gather parameters about (b,σ) in Assumption 4.2:

$$\Theta_1 := (p,d,T,\beta,C,l,(p_i,q_i,f_i)_{i=0}^l).$$

There is no continuity condition w.r.t spatial variable of b. Assumption 4.2(2) means that marginal density and marginal distribution do not affect local integrability of the drift. If b is bounded, then it satisfies Assumption 4.2(2).

Our results are the following:

Theorem 4.3 (Existence)**.** *Let Assumption 4.2 hold.*

1. *SDE (4.1) has a strong solution (X_t) whose marginal distribution is denoted by (μ_t) and marginal density is denoted by (ℓ_t).*
2. *There exist constants $c_1 > 0$ (depending on Θ_1), $c_2 > 0$ (depending on Θ_1, ν), and $\delta \in (0, \frac{1}{2})$ (depending on q_0) such that*

$$\begin{aligned} \sup_{t\in\mathbb{T}} \|\ell_t\|_\infty &\le c_1 \|\ell_\nu\|_\infty, \\ W_p(\mu_s,\mu_t) &\le c_2|t-s|^\delta \quad \textit{for} \quad s,t \in \mathbb{T}. \end{aligned}$$

Theorem 4.4 (Uniqueness)**.** *Let Assumption 4.2 hold. Assume in addition that $p = 1, \|b\|_\infty \le C$ and $\sigma_t(x,\varrho) = \sigma_t(x)$.*

1. *For $k \in \{1,2\}$, let $(X^k_t, t \in \mathbb{T})$ be a weak solution to (4.1), ν_k its initial distribution and $(\ell^k_t, t \in \mathbb{T})$ its marginal density. We assume that ν_k satisfies Assumption 4.2(4). There exists an increasing function $\Lambda : \mathbb{R}_+ \to \mathbb{R}_+$ (depending on Θ_1) such that*

$$\begin{aligned} &\sup_{t\in\mathbb{T}} \int_{\mathbb{R}^d} (1+|x|)|\ell^1_t(x) - \ell^2_t(x)|\,\mathrm{d}x \\ &\le \Lambda(\|\ell_{\nu_1}\|_\infty + M_1(\nu_1)) \int_{\mathbb{R}^d} (1+|x|)|\ell_{\nu_1}(x) - \ell_{\nu_2}(x)|\,\mathrm{d}x. \end{aligned}$$

2. *SDE (4.1) has both weak and strong uniqueness.*

4.2 Outline of the proofs

We will summarize the main ideas behind our mollifying argument. Let (ρ^n) be a sequence of mollifiers. We consider the SDE

$$\begin{cases} \mathrm{d}X_t^n = b(t, X_t^n, (\rho^n * \mu_t^n)(X_t^n), \mu_t^n)\,\mathrm{d}t + \sigma(t, X_t^n, \mu_t^n)\,\mathrm{d}B_t, \\ \nu \text{ is the distribution of } X_0^n, \text{ and } \mu_t^n \text{ is that of } X_t^n. \end{cases} \tag{4.2}$$

Above, $*$ is the convolution operator, i.e.,

$$(\rho^n * \varrho)(x) \coloneqq \int_{\mathbb{R}^d} \rho^n(x-y)\,\mathrm{d}\varrho(y) \quad \text{for} \quad \varrho \in \mathcal{P}(\mathbb{R}^d).$$

Then (4.2) is well-posed and each X_t^n admits a density ℓ_t^n. The sequence (ℓ^n) is locally Hölder continuous on $(0, T] \times \mathbb{R}^d$. By Arzelà–Ascoli theorem, we can extract a subsequence that converges to some function $\ell : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}_+$ uniformly on every strip $[R^{-1}, T] \times B(0, R)$ where $R > 0$. We then verify that $\ell_t \coloneqq \ell(t, \cdot)$ is indeed a density whose induced distribution $\mu_t \in \mathcal{P}_p(\mathbb{R}^d)$. Also, μ_t^n converges to μ_t in W_p (as $n \to \infty$) uniformly for $t \in [R^{-1}, T]$ where $R > 0$.

By Itô's lemma, $(\mu_t^n, t \in \mathbb{T})$ satisfies the Fokker-Planck equation

$$\partial_t \ell_t^n(x) = -\partial_{x_i}\{b(t, x, (\rho^n * \ell_t^n)(x), \mu_t^n)\ell_t^n(x)\} + \frac{1}{2}\partial_{x_i}\partial_{x_j}\{a^{i,j}(t, x, \mu_t^n)\ell_t^n(x)\}.$$

By the convergence of ℓ_t^n to ℓ_t, that of μ_t^n to μ_t, and the continuity of $b(t, x, r, \varrho)$ in (r, ϱ), we deduce that (μ_t) satisfies

$$\partial_t \ell_t(x) = -\partial_{x_i}\{b(t, x, \ell_t(x), \mu_t)\ell(t, x)\} + \frac{1}{2}\partial_{x_i}\partial_{x_j}\{a^{i,j}(t, x, \mu_t)\ell_t(x)\}. \tag{4.3}$$

Notice that (4.3) is the Fokker-Planck equation associated with (4.1). By superposition principle (e.g. [BR20, Section 2]), (4.1) has a weak solution.

4.3 Moment estimates of marginal distribution

In this section, we consider the classical SDE (2.10). First, we establish the following moment estimates:

Theorem 4.5. *Let $p \in [1, \infty)$ and (b, σ) satisfy Assumption 2.15. Let μ_t be the distribution of X_t in (2.10).*

1. *μ_t is absolutely continuous w.r.t Lebesgue measure on $\mathbb{R}^d$.*
2. *There exist constants $c > 0$ (depending on Θ_3, p) and $\delta \in (0, \frac{1}{2})$ (depending on q_0) such that for $0 \le u \le t \le T$:*

$$\mathbb{E}[\sup_{s \in [u,t]} |X_s|^p] \le c(1 + \mathbb{E}[|X_u|^p]), \tag{4.4}$$

$$\mathbb{E}[\sup_{s \in [u,t]} |X_s - X_u|^p] \le c|t-u|^{\delta p}(1 + \mathbb{E}[|X_u|^p]). \tag{4.5}$$

Proof. By Assumption 2.15(2), there exist measurable maps $b^{(0)} : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d$ and $b^{(1)} : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d$ such that $b_t(x) = b_t^{(0)}(x) + b_t^{(1)}(x)$. By Assumption 2.15(3), there exists $f_0 \in \tilde{L}_{q_0}^{p_0}$ with $(p_0, q_0) \in \mathcal{K}$ such that $|b_t^{(0)}(x)| \le f_0(t, x)$. We consider the SDE

$$\mathrm{d}\bar{X}_t = b^{(1)}(t, \bar{X}_t)\,\mathrm{d}t + \sigma(t, \bar{X}_t)\,\mathrm{d}B_t, \tag{4.6}$$

where the distribution of $\bar{X}_0$ is ν. Clearly, $(b^{(1)}, \sigma)$ satisfies Assumption 2.15, so (4.6) is weakly well-posed. We define

$$\begin{aligned}
\xi_t &:= \{\sigma_t^\top a_t^{-1} b_t^{(0)}\}(\bar{X}_t), \\
\bar{B}_t &:= B_t - \int_0^t \xi_s \, \mathrm{d}s, \\
R_t &:= \exp\left(\int_0^t \xi_t^\top \, \mathrm{d}B_t - \frac{1}{2} \int_0^t |\xi_t|^2 \, \mathrm{d}t \right), \\
I_t &:= \mathbb{E}\left[\exp\left(\frac{1}{2} \int_0^t |\xi_t|^2 \, \mathrm{d}t \right) \right].
\end{aligned}$$

By uniform boundedness of $\sigma_t^\top a_t^{-1}$, Remark 2.16 and Proposition 2.17(1), we have $I_T < \infty$. So R_T is an exponential martingale with $\mathbb{E}[R_T] = 1$. By Girsanov's theorem, $(\bar{B}_t, t \in \mathbb{T})$ is an m-BM under the probability measure $\bar{\mathbb{P}} := R_T \mathbb{P}$. We denote by $\bar{\mathbb{E}}$ the expectation w.r.t $\bar{\mathbb{P}}$. Clearly, (4.6) can be written under $\bar{\mathbb{P}}$ as

$$\mathrm{d}\bar{X}_t = b(t, \bar{X}_t) \, \mathrm{d}t + \sigma(t, \bar{X}_t) \, \mathrm{d}\bar{B}_t. \tag{4.7}$$

1. By Assumption 2.15(4) and [MPZ21, Theorem 1.2], the distribution of $\bar{X}_t$ under $\mathbb{P}$ admits a density. Notice that $\bar{\mathbb{P}}$ and $\mathbb{P}$ are equivalent, so the distribution of $\bar{X}_t$ under $\bar{\mathbb{P}}$ also admits a density. Because $\bar{X}_0$ is $\mathcal{F}_0$-measurable, it holds for $\varphi \in C_c^\infty(\mathbb{R}^d)$ that

$$\bar{\mathbb{E}}[\varphi(\bar{X}_0)] = \mathbb{E}[\varphi(\bar{X}_0) R_0] = \mathbb{E}[\varphi(\bar{X}_0)].$$

   Then ν is also the distribution of $\bar{X}_0$ under $\bar{\mathbb{P}}$. By weak uniqueness of (2.10) and (4.7), the distribution of X_t under $\mathbb{P}$ is the same as that of $\bar{X}_t$ under $\bar{\mathbb{P}}$. Thus the distribution of X_t under $\mathbb{P}$ admits a density.

2. We combine localization argument (see e.g. [Bal17, Theorem 9.1]) with Krylov's estimate.

   (a) We fix $p \in [1, \infty)$ and $u \in [0, T)$ such that $\mathbb{E}[|X_u|^p] < \infty$. For $R > 0$, let $\tau_R := \inf\{t \in [u, T] : |X_t| \geq R\}$ be the exit time of X_t from the open ball $B(0, R)$. We adopt the convention that $\tau_R = T$ if $|X_t| < R$ for all $t \in [u, T]$. We denote $X_R(t) := X_{t \wedge \tau_R}$. It holds for $t \in [u, T]$ that

$$\begin{aligned}
X_R(t) &= X_u + \int_u^{t \wedge \tau_R} b(r, X_r) \, \mathrm{d}r + \int_u^{t \wedge \tau_R} \sigma(r, X_r) \, \mathrm{d}B_r \\
&= X_u + \int_u^t b(r, X_r) 1_{\{r < \tau_R\}} \, \mathrm{d}r + \int_u^t \sigma(r, X_r) 1_{\{r < \tau_R\}} \, \mathrm{d}B_r \\
&= X_u + \int_u^t b^{(1)}(r, X_R(r)) 1_{\{r < \tau_R\}} \, \mathrm{d}r + \int_u^t b^{(0)}(r, X_r) 1_{\{r < \tau_R\}} \, \mathrm{d}r \\
&\qquad + \int_u^t \sigma(r, X_R(r)) 1_{\{r < \tau_R\}} \, \mathrm{d}B_r.
\end{aligned}$$

   By Hardy's inequality, it holds for $n \in \mathbb{N}, p \geq 1$ and $x_1, \ldots, x_n \in \mathbb{R}^d$ that $|x_1 + \cdots + x_n|^p \leq n^p(|x_1|^p + \cdots + |x_n|^p)$. We write $M_1 \lesssim M_2$ if there exists a constant $c > 0$

(depending on Θ_3, p) such that $M_1 \leq cM_2$. Then

$$\begin{aligned}\mathbb{E}[\sup_{s\in[u,t]} |X_R(s)|^p] &\lesssim \mathbb{E}[|X_u|^p] + \mathbb{E}\Big[\Big(\int_u^t f_0(r, X_r)\,\mathrm{d}r\Big)^p\Big] \\ &\quad + \mathbb{E}\Big[\Big(\int_u^t |b^{(1)}(r, X_R(r))|\,\mathrm{d}r\Big)^p\Big] \\ &\quad + \mathbb{E}\Big[\sup_{s\in[u,t]}\Big|\int_u^s \sigma(r, X_R(r))1_{\{r<\tau_R\}}\,\mathrm{d}B_r\Big|^p\Big] \\ &=: \mathbb{E}[|X_u|^p] + I_1 + I_2 + I_3.\end{aligned}$$

By Proposition 2.17(2), $I_1 \lesssim 1$. We have $|b^{(1)}(r,x)| \lesssim 1 + |x|$, so

$$\begin{aligned}I_2 &\leq |t-u|^{p-1}\mathbb{E}\Big[\int_u^t |b^{(1)}(r, X_R(r))|^p\,\mathrm{d}r\Big] \quad \text{by Hölder's inequality} \\ &\lesssim |t-u|^{p-1}\mathbb{E}\left[\int_u^t (1 + |X_R(r)|^p)\,\mathrm{d}r\right] \\ &\lesssim 1 + |t-u|^{p-1}\mathbb{E}\left[\int_u^t |X_R(r)|^p\,\mathrm{d}r\right].\end{aligned}$$

By Burkholder-Davis-Gundy inequality (see e.g. [Sch21, Theorem 19.20]) and boundedness of σ (from Assumption 2.15(4)),

$$I_3 \lesssim \mathbb{E}\Big[\Big(\int_u^t |\sigma(r, X_R(r))|^2\,\mathrm{d}r\Big)^{p/2}\Big] \lesssim |t-u|^{\frac{p}{2}}.$$

As a result, it holds for $t \in [u, T]$ that

$$\begin{aligned}\eta_R(t) &:= \mathbb{E}[\sup_{s\in[u,t]} |X_R(s)|^p] \\ &\lesssim 1 + \mathbb{E}[|X_u|^p] + \int_u^t \mathbb{E}[|X_R(r)|^p]\,\mathrm{d}r.\end{aligned}$$

By construction, $|X_R(s)| \leq |X_u| \vee R$ for $s \in [u,t]$. This implies $\eta_R(t) \leq \mathbb{E}[|X_u|^p \vee R^p] < \infty$ for $t \in [u,T]$. It holds for $t \in [u,T]$ that

$$\eta_R(t) \lesssim 1 + \mathbb{E}[|X_u|^p] + \int_u^t \eta_R(r)\,\mathrm{d}r.$$

By Gronwall's lemma, it holds for $t \in [u,T]$ and $R > 0$ that $\eta_R(t) \lesssim 1 + \mathbb{E}[|X_u|^p]$. Because X has continuous sample paths, $\tau_R \uparrow T$ a.s. as $R \to \infty$. Hence

$$\sup_{s\in[u,t]} |X_R(s)|^p \uparrow \sup_{s\in[u,t]} |X_s|^p \quad \text{a.s. as} \quad R \to \infty.$$

The estimate (4.4) then follows from monotone convergence theorem (MCT).

(b) We have

$$\begin{aligned}\mathbb{E}\big[\sup_{s\in[u,t]}|X_s-X_u|^p\big] &\lesssim \mathbb{E}\Big[\Big(\int_u^t f_0(r,X_r)\,\mathrm{d}r\Big)^p\Big] \\ &\quad+\mathbb{E}\Big[\Big(\int_u^t |b^{(1)}(r,X_r)|\,\mathrm{d}r\Big)^p\Big] \\ &\quad+\mathbb{E}\Big[\sup_{s\in[u,t]}\Big|\int_u^s \sigma(r,X_r)\,\mathrm{d}B_r\Big|^p\Big] \\ &=: J_1+J_2+J_3.\end{aligned}$$

There exists $\bar{q}_0\in(2,q_0)$ such that $(p_0,\bar{q}_0)\in\bar{K}$. Let $\delta:=\frac{1}{\bar{q}_0}-\frac{1}{q_0}\in(0,\frac{1}{2})$. By Hölder's inequality,

$$\|f_0\|_{\tilde{L}^{p_0}_{\bar{q}_0}(u,t)}\le (t-u)^{\delta}\|f_0\|_{\tilde{L}^{p_0}_{q_0}(u,t)}.$$

By Proposition 2.17(2),

$$J_1\lesssim \|f_0\|^p_{\tilde{L}^{p_0}_{\bar{q}_0}(u,t)}\le (t-u)^{\delta p}\|f_0\|^p_{\tilde{L}^{p_0}_{q_0}(u,t)}\lesssim (t-u)^{\delta p}.$$

As for I_2 and I_3, we have

$$\begin{aligned}J_2&\lesssim |t-u|^{p-1}\mathbb{E}\left[\int_u^t(1+|X_r|^p)\,\mathrm{d}r\right]\\ &\lesssim |t-u|^p(1+\mathbb{E}[|X_u|^p])\quad\text{by (4.4)},\\ J_3&\lesssim |t-u|^{\frac{p}{2}}.\end{aligned}$$

Then

$$\mathbb{E}\big[\sup_{s\in[u,t]}|X_s-X_u|^p\big]\lesssim |t-u|^{\delta p}(1+\mathbb{E}[|X_u|^p]).$$

The estimate (4.5) then follows. This completes the proof.

□

We consider the following set of assumption:

Assumption 4.6. Let Assumption 2.15 hold with $b^{(1)}=0$.

We gather parameters from Assumption 4.6 in Θ_4. For $\lambda>0$ and $\gamma\in\mathbb{R}$, the heat kernel $p^{\gamma,\lambda}$ is defined for $t>0$ and $x\in\mathbb{R}^d$ by

$$p_t^{\gamma,\lambda}(x):=\frac{1}{t^{(\gamma+d)/2}}e^{-\frac{\lambda|x|^2}{t}}.$$

First, we recall the following estimates:

Lemma 4.7. *[Zha24, Lemma 3.9] Let (b,σ) satisfy Assumption 4.6. Let $\gamma_0:=1-\frac{d}{p_0}-\frac{2}{q_0}$. Then (2.11) has a unique weak solution and $X^x_{s,t}$ admits a density. Moreover, there exist constants $c_1,c_2,c_3>0$ and $\lambda\in(0,1)$ depending on Θ_4 such that*

1. *(Gaussian estimate) for $0\le s<t\le T$ and $x,y\in\mathbb{R}^d$:*

$$c_1p^{0,\lambda^{-1}}_{t-s}(y-x)\le p^{b,\sigma}_{s,t}(x,y)\le c_2p^{0,\lambda}_{t-s}(y-x).$$

2. *(Gradient estimate) for* $0 \le s < t \le T$ *and* $x, y \in \mathbb{R}^d$*:*

$$|\nabla_x p^{b,\sigma}_{s,t}(x,y)| \le c_3 p^{1,\lambda}_{t-s}(y-x).$$

3. *(Hölder estimate in* t *and* y*) for* $\gamma \in (0, \beta \wedge \gamma_0)$*, there exists a constant* $c_4 > 0$ *depending on* (Θ_4, γ) *such that for* $0 \le s < t_1 < t_2 \le T$ *and* $x, y, y_1, y_2 \in \mathbb{R}^d$*:*

$$|p^{b,\sigma}_{s,t_2}(x,y) - p^{b,\sigma}_{s,t_1}(x,y)| \le c_4 |t_1 - t_2|^{\frac{\gamma}{2}} \sum_{i=1}^{2} p^{\gamma,\lambda}_{t_i - s}(x-y),$$
$$|p^{b,\sigma}_{s,t}(x,y_1) - p^{b,\sigma}_{s,t}(x,y_2)| \le c_4 |y_1 - y_2|^{\gamma} \sum_{i=1}^{2} p^{\gamma,\lambda}_{t-s}(x-y_i).$$

We have an immediate corollary:

Corollary 4.8. *Let* (b,σ) *satisfy Assumption 4.6. Let* ν *admit a bounded density* ℓ_ν*. Let* ℓ_t *be the density of* X_t *in (2.10).*

1. *There exists a constant* $c_1 > 0$ *depending on* Θ_4 *such that:*

$$\sup_{t \in \mathbb{T}} \|\ell_t\|_\infty \le c_1 \|\ell_\nu\|_\infty.$$

2. *Let* $\gamma_0 := 1 - \frac{d}{p_0} - \frac{2}{q_0}$*. For* $\gamma \in (0, \beta \wedge \gamma_0)$*, there exists a constant* $c_2 > 0$ *depending on* (Θ_4, γ, ν) *such that for* $0 < s, t \le T$ *and* $x, y \in \mathbb{R}^d$*:*

$$|\ell_t(x) - \ell_s(x)| \le c_2 |t-s|^{\frac{\gamma}{2}} (t^{-\frac{\gamma}{2}} + s^{-\frac{\gamma}{2}}),$$
$$|\ell_t(x) - \ell_t(y)| \le c_2 |x-y|^{\gamma} t^{-\frac{\gamma}{2}}.$$

Proof. We have

$$\ell_t(x) = \int_{\mathbb{R}^d} \ell_\nu(y) p^{b,\sigma}_{0,t}(y,x) \, \mathrm{d}y.$$

The claims then follow from Lemma 4.7 and the fact that ℓ_ν is bounded. □

By Theorem 2.19, we have the following Duhamel presentation:

Lemma 4.9. *Let* (b,σ) *satisfy Assumption 4.6. Let* ν *admit a density* ℓ_ν*. Let* ℓ_t *be the density of* X_t *in (2.10). Then it holds for* $t \in \mathbb{T}$ *and* $x \in \mathbb{R}^d$*:*

$$\ell_t(x) = \int_{\mathbb{R}^d} p^{0,\sigma}_{0,t}(y,x) \ell_\nu(y) \, \mathrm{d}y + \int_0^t \int_{\mathbb{R}^d} \ell_s(y) \langle b_s(y), \nabla_y p^{0,\sigma}_{s,t}(y,x) \rangle \, \mathrm{d}y \, \mathrm{d}s.$$

Third, we establish a decay of tail moment:

Lemma 4.10. *Let* $p \in \{0\} \cup [1, \infty)$ *and* $\nu \in \mathcal{P}_p(\mathbb{R}^d)$*. Let* (b,σ) *satisfy Assumption 4.6 and* (X_t) *satisfy (2.10). Let* μ_t *be the distribution of* X_t*. There exists a function* $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ *depending on* (Θ_4, ν, p) *such that* $\lim_{R \to \infty} \varphi(R) = 0$ *and that*

$$\sup_{t \in \mathbb{T}} \int_{B^c(0,R)} |\cdot|^p \, \mathrm{d}\mu_t \le \phi(R) \quad \textit{for} \quad R \ge 0.$$

Above, $B^c(0,R) := \mathbb{R}^d \setminus B(0,R)$ where $B(0,R) := \{x \in \mathbb{R}^d : |x| \le \mathbb{R}\}$.

Proof. WLOG, we consider $R > 0$. We write $M_1 \lesssim M_2$ if there exists a constant $c > 0$ depending on (Θ_4, ν, p) such that $M_1 \le cM_2$. Let ℓ_t be the density of X_t. Then

$$\ell_t(x) = \int_{\mathbb{R}^d} \ell_\nu(y) p^{b,\sigma}_{0,t}(y,x)\,\mathrm{d}y.$$

By Lemma 4.7(1), there exists a constant $\lambda \in (0,1)$ depending on Θ_4 such that $p^{b,\sigma}_{0,t}(y,x) \lesssim p^{0,\lambda}_t(y-x)$. Then $\ell_t \lesssim \ell_\nu * p^{0,\lambda}_t$. Let Z be a standard normal random variable on $\mathbb{R}^d$. Let Y be a random variable on $\mathbb{R}^d$, independent of Z, with distribution ν. Let $c_t := \sqrt{\frac{t}{2\lambda}}$ and $s := \frac{1}{2}$. Then

$$\begin{aligned}
\int_{B^c_R} |\cdot|^p \,\mathrm{d}\mu_t &\lesssim \mathbb{E}[1_{\{|c_t Z + Y| > R\}} |c_t Z + Y|^p] \\
&\lesssim \mathbb{E}[(1_{\{|c_t Z| > sR\}} + 1_{\{|Y| > (1-s)R\}})(|c_t|^p |Z|^p + |Y|^p)] \\
&\lesssim \mathbb{E}[1_{\{|Z| > \frac{sR}{c_T}\}} |Z|^p] + \mathbb{E}[1_{\{|Z| > \frac{sR}{c_T}\}} |Y|^p] \\
&\qquad + \mathbb{E}[1_{\{|Y| > (1-s)R\}} |Z|^p] + \mathbb{E}[1_{\{|Y| > (1-s)R\}} |Y|^p] \\
&=: I_1(R) + I_2(R) + I_3(R) + I_4(R) \\
&=: \phi(R).
\end{aligned}$$

By Markov's inequality,

$$\begin{aligned}
\mathbb{P}[|Z| > sR/c_T] &\le \frac{c_T \mathbb{E}[|Z|]}{sR}, \\
\mathbb{P}[|Y| > (1-s)R] &\le \frac{\mathbb{E}[|Y|]}{(1-s)R}.
\end{aligned}$$

We have $\mathbb{E}[|Z|^p] + \mathbb{E}|Y|^p] < \infty$. By dominated convergence theorem (DCT),

$$\lim_{R\to\infty} I_1(R) = \lim_{R\to\infty} I_2(R) = \lim_{R\to\infty} I_3(R) = \lim_{R\to\infty} I_4(R) = 0.$$

This completes the proof. □

4.4 Existence and regularity of a solution

This section is dedicated to the proof of Theorem 4.3. We recall that $\Theta_1 = (p, d, T, \beta, C, l, (p_i, q_i, f_i)_{i=0}^l)$ contains parameters about (b, σ) from Assumption 4.2. We write $M_1 \lesssim M_2$ if there exists a constant $c > 0$ (depending on Θ_1, ν) such that $M_1 \le cM_2$. We construct a sequence (ρ^n) of mollifiers as follows. We fix a smooth density $\rho : \mathbb{R}^d \to \mathbb{R}$ whose support is contained in $B(0,1)$. For each $n \in \mathbb{N}$, we define $\rho^n : \mathbb{R}^d \to \mathbb{R}$ by $\rho^n(x) := n^d \rho(nx)$ and consider the McKean-Vlasov SDE

$$\begin{cases} \mathrm{d}X^n_t = b(t, X^n_t, (\rho^n * \mu^n_t)(X^n_t), \mu^n_t)\,\mathrm{d}t + \sigma(t, X^n_t, \mu^n_t)\,\mathrm{d}B_t, \\ \nu \text{ is the distribution of } X^n_0, \text{ and } \mu^n_t \text{ is that of } X^n_t. \end{cases} \tag{4.8}$$

Then (4.8) is a mollified version of (4.1).

4.4.1 Stability estimates for mollified SDEs

We define the map $b^n : \mathbb{T} \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$ by $b^n(t, x, \varrho) := b(t, x, (\rho^n * \varrho)(x), \varrho)$. Then

$$
\begin{aligned}
|b^n(t,x,\varrho) - b^n(t,x,\tilde{\varrho})| &\lesssim \left| \int_{\mathbb{R}^d} \rho^n(x-y) \, \mathrm{d}(\varrho - \tilde{\varrho})(y) \right| + W_p(\varrho, \tilde{\varrho}) && (4.9)\\
&\le \|\nabla \rho^n\|_\infty W_1(\varrho, \tilde{\varrho}) + W_p(\varrho, \tilde{\varrho}) && (4.10)\\
&\le (1 + \|\nabla \rho^n\|_\infty) W_p(\varrho, \tilde{\varrho}). && (4.11)
\end{aligned}
$$

Above, (4.9) is due to Assumption 4.2(5), (4.10) due to Lemma 2.1(2), and (4.11) due to Lemma 2.1(3). It follows that b^n is Lipschitz in distribution variable. We consider the McKean-Vlasov SDE

$$
\begin{cases} \mathrm{d}Y_t = b^n(t, Y_t, \xi_t) \, \mathrm{d}t + \sigma(t, Y_t, \xi_t) \, \mathrm{d}B_t, \\ \nu \text{ is the distribution of } Y_0, \text{ and } \xi_t \text{ is that of } Y_t. \end{cases} \tag{4.12}
$$

It follows from [HW22, Theorem 1.1(1)] that (4.12) is well-posed.

Remark 4.11. The Lipschitz continuity of $b^n(t, x, \cdot)$ is just for the application of [HW22, Theorem 1.1(1)], and its Lipschitz constant does not play any role below.

By (4.8), $(X_t^n, t \in \mathbb{T})$ satisfies (4.12). As a consequence, (4.8) is well-posed. We define the maps $\bar{b}^n : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d$ and $\bar{\sigma}^n : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^m$ by

$$
\begin{aligned}
\bar{b}^n(t, x) &:= b^n(t, x, \mu_t^n), \\
\bar{\sigma}^n(t, x) &:= \sigma(t, x, \mu_t^n).
\end{aligned}
$$

We have

$$
\mathrm{d}X_t^n = \bar{b}^n(t, X_t^n) \, \mathrm{d}t + \bar{\sigma}^n(t, X_t^n) \, \mathrm{d}B_t. \tag{4.13}
$$

Let $\bar{a}^n := \bar{\sigma}^n (\bar{\sigma}^n)^\top$. We denote $\bar{b}_t^n := \bar{b}^n(t, \cdot), \bar{\sigma}_t^n := \bar{\sigma}^n(t, \cdot)$ and $\bar{a}_t^n := \bar{a}^n(t, \cdot)$.

Remark 4.12. All pairs $(\bar{b}^n, \bar{\sigma}^n)_{n \in \mathbb{N}}$ satisfy Assumption 4.6 for the same set of parameters.

By Theorem 4.5(1), each X_t^n admits a density denoted by ℓ_t^n. By Theorem 4.5(2), there exists a constant $\delta \in (0, \frac{1}{2})$ depending on q_0 such that

$$
\sup_{n \in \mathbb{N}} \sup_{t \in \mathbb{T}} M_p(\mu_t^n) \lesssim 1, \tag{4.14}
$$

$$
\sup_{n \in \mathbb{N}} W_p(\mu_s^n, \mu_t^n) \lesssim |t - s|^\delta \quad \text{for} \quad s, t \in \mathbb{T}. \tag{4.15}
$$

Let $\gamma_0 := 1 - \frac{d}{p_0} - \frac{2}{q_0}$ and $\gamma := \frac{\beta \wedge \gamma_0}{2}$. By Corollary 4.8(1),

$$
\sup_{n \in \mathbb{N}} \sup_{t \in \mathbb{T}} \|\ell_t^n\|_\infty \lesssim \|\ell_\nu\|_\infty. \tag{4.16}
$$

By Corollary 4.8(2), it holds for $t \in (0, T]$ that

$$
\sup_{n \in \mathbb{N}} \sup_{\substack{s, r \in [t, T] \\ s \neq r}} \sup_{x \in \mathbb{R}^d} \frac{|\ell_s^n(x) - \ell_r^n(x)|}{|s - r|^{\frac{\gamma}{2}}} \lesssim t^{-\frac{\gamma}{2}}, \tag{4.17}
$$

$$
\sup_{n \in \mathbb{N}} \sup_{s \in [t, T]} \sup_{\substack{x, y \in \mathbb{R}^d \\ x \neq y}} \frac{|\ell_s^n(x) - \ell_s^n(y)|}{|x - y|^\gamma} \lesssim t^{-\frac{\gamma}{2}}. \tag{4.18}
$$

By Lemma 4.10, there exists a function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ depending on (Θ_1, ν) such that $\lim_{R\to\infty} \phi(R) = 0$ and that

$$\sup_{n\in\mathbb{N}} \sup_{t\in\mathbb{T}} \int_{B^c(0,R)} (1 + |\cdot|^p) \, \mathrm{d}\mu_t^n \le \phi(R) \quad \text{for} \quad R \ge 0. \tag{4.19}$$

By (4.15), the map $\mathbb{T} \to \mathcal{P}_p(\mathbb{R}^d)$, $t \mapsto \mu_t^n$ is continuous.

4.4.2 Convergence of marginal densities of mollified SDEs

By (4.17), (4.18), Arzelà–Ascoli theorem and diagonal extraction, there exist a sub-sequence (also denoted by (ℓ^n) for simplicity) and a continuous function $\ell : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}_+$ such that

$$\lim_n \sup_{t\in[R^{-1},T]} \sup_{x\in B(0,R)} |\ell_t^n(x) - \ell_t(x)| = 0 \quad \text{for} \quad R > T^{-1}. \tag{4.20}$$

Above, $\ell_t := \ell(t, \cdot)$. Clearly, $\ell_0 = \ell_\nu$ and

$$\sup_{t\in\mathbb{T}} \|\ell_t\|_\infty \lesssim \|\ell_\nu\|_\infty. \tag{4.21}$$

We remark that the constant in (4.21) depends on Θ_1. Next we verify that ℓ_t is indeed a density for $t \in (0, T]$. We have

$$\begin{aligned}\int_{B(0,R)} \ell_t^n(x) \, \mathrm{d}x &= 1 - \int_{B^c(0,R)} \ell_t^n(x) \, \mathrm{d}x \\ &\gtrsim 1 - \phi(R) \quad \text{by (4.19)}.\end{aligned}$$

By (4.20), (4.21) and DCT,

$$\int_{B(0,R)} \ell_t(x) \, \mathrm{d}x = \lim_{n\to\infty} \int_{B(0,R)} \ell_t^n(x) \, \mathrm{d}x.$$

It follows that

$$1 - \phi(R) \lesssim \int_{B(0,R)} \ell_t(x) \, \mathrm{d}x \le 1.$$

Then

$$\int_{\mathbb{R}^d} \ell_t(x) \, \mathrm{d}x = \lim_{R\to\infty} \int_{B(0,R)} \ell_t(x) \, \mathrm{d}x = 1.$$

Let $\mu_t \in \mathcal{P}(\mathbb{R}^d)$ be the probability measure induced by ℓ_t, i.e.,

$$\mu_t(B) := \int_B \ell_t(x) \, \mathrm{d}x \quad \text{for} \quad B \in \mathcal{B}(\mathbb{R}^d).$$

By (4.20), $\mu_t^n \overset{*}{\rightharpoonup} \mu_t$ as $n \to \infty$. By Lemma 2.21, $\mu_t^n \rightharpoonup \mu_t$ as $n \to \infty$. We have

$$\begin{aligned}\int_{B^c(0,R)} |\cdot|^p \, \mathrm{d}\mu_t &= \lim_{K\to\infty} \int_{B^c(0,R)\cap B(0,K)} |\cdot|^p \, \mathrm{d}\mu_t \quad \text{by MCT} \\ &= \lim_{K\to\infty} \lim_{n\to\infty} \int_{B^c(0,R)\cap B(0,K)} |\cdot|^p \, \mathrm{d}\mu_t^n \quad \text{by (4.20)} \\ &\lesssim \phi(R) \quad \text{by (4.19)}.\end{aligned}$$

Then

$$\sup_{t\in\mathbb{T}} \int_{B^c(0,R)} |\cdot|^p \, \mathrm{d}\mu_t \lesssim \phi(R). \tag{4.22}$$

Clearly, $|\cdot|^p$ is continuous and bounded from below. By Lemma 2.21 and Portmanteau's theorem,

$$\begin{aligned}\sup_{t\in\mathbb{T}} M_p(\mu_t) &\le \sup_{t\in\mathbb{T}} \liminf_n M_p(\mu_t^n) \\ &\lesssim 1 \quad \text{by (4.14)}.\end{aligned}$$

Then $\mu_t \in \mathcal{P}_p(\mathbb{R}^d)$ for $t \in \mathbb{T}$. Moreover,

$$\begin{aligned} W_p^p(\mu_s,\mu_t) &= \sup\left\{\int_{\mathbb{R}^d} \varphi \,\mathrm{d}\mu_s + \int_{\mathbb{R}^d} \psi \,\mathrm{d}\mu_t : (\varphi,\psi)\in\Phi_p\right\} && (4.23)\\ &= \sup\left\{\lim_n \left(\int_{\mathbb{R}^d} \varphi \,\mathrm{d}\mu_s^n + \int_{\mathbb{R}^d} \psi \,\mathrm{d}\mu_t^n\right) : (\varphi,\psi)\in\Phi_p\right\} && (4.24)\\ &\le \limsup_n \sup\left\{\int_{\mathbb{R}^d} \varphi \,\mathrm{d}\mu_s^n + \int_{\mathbb{R}^d} \psi \,\mathrm{d}\mu_t^n : (\varphi,\psi)\in\Phi_p\right\} \\ &= \limsup_n W_p^p(\mu_s^n,\mu_t^n) && (4.25)\\ &\lesssim |t-s|^{\delta p}. && (4.26)\end{aligned}$$

Above, (4.23) and (4.25) are due to Lemma 2.1(1); (4.24) due to Lemma 2.21; and (4.26) due to (4.15). It follows that

$$W_p(\mu_s,\mu_t) \lesssim |t-s|^{\delta} \quad \text{for} \quad s,t\in\mathbb{T}.$$

Next we establish an essential result about convergence:

Lemma 4.13. *We have for $R > T^{-1}$ that*

$$\lim_n \sup_{t\in[R^{-1},T]} W_p(\mu_t^n,\mu_t) = 0.$$

Proof. We have

$$\begin{aligned} W_p^p(\mu_t^n,\mu_t) &\lesssim \int_{\mathbb{R}^d} |x|^p |\ell_t^n(x) - \ell_t(x)| \,\mathrm{d}x \quad \text{by Lemma 2.2} \\ &\le \int_{B(0,k)} |x|^p |\ell_t^n(x) - \ell_t(x)| \,\mathrm{d}x + \int_{B^c(0,k)} |x|^p (\ell_t^n(x) + \ell_t(x)) \,\mathrm{d}x \\ &=: I(t,n,k) + J(t,n,k) \quad \text{for} \quad k>0.\end{aligned}$$

By (4.20),

$$\lim_n \sup_{t\in[R^{-1},T]} I(t,n,k) = 0.$$

By (4.19) and (4.22),

$$\sup_{n\in\mathbb{N}} \sup_{t\in\mathbb{T}} J(t,n,k) \lesssim \phi(k).$$

As such,

$$\begin{aligned} &\limsup_n \sup_{t\in[R^{-1},T]} W_p^p(\mu_t^n,\mu_t) \\ &\lesssim \limsup_n \sup_{t\in[R^{-1},T]} I(t,n,k) + \limsup_n \sup_{t\in\mathbb{T}} J(t,n,k) \\ &\lesssim \phi(k).\end{aligned}$$

The claim then follows by taking the limit $k\to\infty$. □

4.4.3 Existence of a weak solution

Notice that $\rho^n * \mu^n_t = \rho^n * \ell^n_t$. The Fokker-Planck equation (in distributional sense) associated with (4.8) is

$$\partial_t \ell^n_t(x) = -\partial_{x_i}\{b(t,x,(\rho^n * \ell^n_t)(x), \mu^n_t)\ell^n_t(x)\} + \frac{1}{2}\partial_{x_i}\partial_{x_j}\{a^{i,j}(t,x,\mu^n_t)\ell^n_t(x)\}.$$

This means for each $(\varphi,\psi) \in C^\infty_c(0,T) \times C^\infty_c(\mathbb{R}^d)$ that

$$\begin{aligned}
&-\int_{\mathbb{T}}\int_{\mathbb{R}^d} \varphi'(t)\psi(x)\,\mathrm{d}\mu^n_t(x)\,\mathrm{d}t \\
=&\sum_{i=1}^d \int_{\mathbb{T}}\int_{\mathbb{R}^d} b(t,x,(\rho^n * \ell^n_t)(x),\mu^n_t)\varphi(t)\partial_{x_i}\psi(x)\,\mathrm{d}\mu^n_t(x)\,\mathrm{d}t \\
&\quad + \frac{1}{2}\sum_{i,j=1}^d \int_{\mathbb{T}}\int_{\mathbb{R}^d} a^{i,j}(t,x,\mu^n_t)\varphi(t)\partial_{x_i}\partial_{x_j}\psi(x)\,\mathrm{d}\mu^n_t(x)\,\mathrm{d}t.
\end{aligned} \tag{4.27}$$

Above, $a^{i,j}$ is the entry in the i-th row and j-th column of a. We recall from Lemma 4.13 and (4.20) that

$$W_p(\mu^n_t,\mu_t) \xrightarrow{n\to\infty} 0, \tag{4.28}$$

$$\sup_{x\in B(0,R)} |\ell^n_t(x) - \ell_t(x)| \xrightarrow{n\to\infty} 0 \quad \text{for} \quad (t,R) \in \mathbb{T}\times\mathbb{R}_+. \tag{4.29}$$

We fix $(\varphi,\psi) \in C^\infty_c(0,T) \times C^\infty_c(\mathbb{R}^d)$. By (4.28), the boundedness of a, and the continuity of a w.r.t distribution variable,

$$\int_{\mathbb{R}^d} a^{i,j}(t,x,\mu^n_t)\partial_{x_i}\partial_{x_j}\psi(x)\,\mathrm{d}\mu^n_t(x) \xrightarrow{n\to\infty} \int_{\mathbb{R}^d} a^{i,j}(t,x,\mu_t)\partial_{x_i}\partial_{x_j}\psi(x)\,\mathrm{d}\mu_t(x).$$

Let $S \coloneqq B(0,1) + \operatorname{supp}\psi$. By triangle inequality,

$$\begin{aligned}
\|1_S\{(\rho^n * \ell^n_t) - \ell_t\}\|_\infty &\le \|1_S\{\rho^n * (\ell^n_t - \ell_t)\}\|_\infty + \|1_S(\rho^n * \ell_t - \ell_t)\|_\infty \\
&\le \|\rho^n * \{1_S(\ell^n_t - \ell_t)\}\|_\infty + \|1_S(\rho^n * \ell_t - \ell_t)\|_\infty \\
&\le \|1_S(\ell^n_t - \ell_t)\|_\infty + \|1_S(\rho^n * \ell_t - \ell_t)\|_\infty.
\end{aligned}$$

By (4.29), $\|1_S(\ell^n_t - \ell_t)\|_\infty \to 0$ as $n\to\infty$. By [Bre11, Proposition 4.21], $\|1_S(\rho^n * \ell_t - \ell_t)\|_\infty \to 0$ as $n\to\infty$. It follows that $\|1_S\{(\rho^n * \ell^n_t) - \ell_t\}\|_\infty \to 0$ as $n\to\infty$. This, together with (4.28) and Assumption 4.2(5), implies

$$\sup_{x\in S} |b(t,x,(\rho^n * \ell^n_t)(x),\mu^n_t) - b(t,x,\ell_t(x),\mu_t)| \xrightarrow{n\to\infty} 0. \tag{4.30}$$

Recall that $|b| \le f_0$ and $f_0 \in \tilde{L}^{p_0}_{q_0}$. It follows from (4.28), (4.30) and DCT that

$$\int_{\mathbb{R}^d} b(t,x,(\rho^n * \ell^n_t)(x),\mu^n_t)\partial_{x_i}\psi(x)\,\mathrm{d}\mu^n_t(x) \xrightarrow{n\to\infty} \int_{\mathbb{R}^d} b(t,x,\ell_t(x),\mu_t)\partial_{x_i}\psi(x)\,\mathrm{d}\mu_t(x).$$

Taking the limit $n \to \infty$ in (4.27), we get

$$
\begin{aligned}
&- \int_{\mathbb{T}} \int_{\mathbb{R}^d} \varphi'(t)\psi(x) \, \mathrm{d}\mu_t(x) \, \mathrm{d}t \\
=& \sum_{i=1}^d \int_{\mathbb{T}} \int_{\mathbb{R}^d} b(t,x,\ell_t(x),\mu_t)\varphi(t)\partial_{x_i}\psi(x) \, \mathrm{d}\mu_t(x) \, \mathrm{d}t \\
&+ \frac{1}{2} \sum_{i,j=1}^d \int_{\mathbb{T}} \int_{\mathbb{R}^d} a^{i,j}(t,x,\mu_t)\varphi(t)\partial_{x_i}\partial_{x_j}\psi(x) \, \mathrm{d}\mu_t(x) \, \mathrm{d}t.
\end{aligned}
$$

So ℓ satisfies the Fokker-Planck equation

$$
\partial_t \ell_t(x) = -\partial_{x_i}\{b(t,x,\ell_t(x),\mu_t)\ell_t(x)\} + \frac{1}{2}\partial_{x_i}\partial_{x_j}\{a^{i,j}(t,x,\mu_t)\ell_t(x)\}.
$$

Moreover, ℓ satisfies the following integrability estimate:

Lemma 4.14. *There exists a constant $c > 0$ (depending on Θ_1) such that*

$$
\int_{\mathbb{T}} \int_{\mathbb{R}^d} \{|b(t,x,\ell_t(x),\mu_t)| + |a(t,x,\mu_t)|\} \, \mathrm{d}\mu_t(x) \, \mathrm{d}t \le c(1 + \|f_0\|_{\tilde{L}^{p_0}_{q_0}}).
$$

Proof. By (4.13),

$$
\mathrm{d}X^n_t = \bar{b}^n(t,X^n_t) \, \mathrm{d}t + \bar{\sigma}^n(t,X^n_t) \, \mathrm{d}B_t.
$$

Recall that Assumption 4.6 is a special case of Assumption 2.15. By Remark 4.12, all pairs $(\bar{b}^n, \bar{\sigma}^n)_{n\in\mathbb{N}}$ satisfy Assumption 2.15 for the same set of parameters. Then

$$
\begin{aligned}
\int_{\mathbb{T}} \int_{\mathbb{R}^d} f_0(t,x) \, \mathrm{d}\mu^n_t(x) \, \mathrm{d}t &= \mathbb{E}\Big[\int_0^T f_0(t,X^n_t) \, \mathrm{d}t\Big] \quad \text{by Tonelli's theorem} \\
&\lesssim 1 + \|f_0\|_{\tilde{L}^{p_0}_{q_0}} \quad \text{by Proposition 2.17(2).} \qquad (4.31)
\end{aligned}
$$

We have

$$
\begin{aligned}
\int_{\mathbb{T}} \int_{\mathbb{R}^d} f_0(t,x) \, \mathrm{d}\mu_t(x) \, \mathrm{d}t &= \int_{\mathbb{T}} \lim_k \int_{\mathbb{R}^d} 1_{B(0,k)}(x) f_0(t,x)\ell_t(x) \, \mathrm{d}x \, \mathrm{d}t && (4.32)\\
&\le \liminf_k \int_{\mathbb{T}} \int_{\mathbb{R}^d} 1_{B(0,k)}(x) f_0(t,x)\ell_t(x) \, \mathrm{d}x \, \mathrm{d}t && (4.33)\\
&= \liminf_k \int_{\mathbb{T}} \lim_n \int_{\mathbb{R}^d} 1_{B(0,k)}(x) f_0(t,x)\ell^n_t(x) \, \mathrm{d}x \, \mathrm{d}t && (4.34)\\
&\le \liminf_k \liminf_n \int_{\mathbb{T}} \int_{\mathbb{R}^d} 1_{B(0,k)}(x) f_0(t,x)\ell^n_t(x) \, \mathrm{d}x \, \mathrm{d}t && (4.35)\\
&\le \liminf_n \int_{\mathbb{T}} \int_{\mathbb{R}^d} f_0(t,x)\ell^n_t(x) \, \mathrm{d}x \, \mathrm{d}t \\
&\lesssim 1 + \|f_0\|_{\tilde{L}^{p_0}_{q_0}} \quad \text{by (4.31).} && (4.36)
\end{aligned}
$$

Above, (4.32) is due to MCT; (4.33) and (4.35) are due to Fatou's lemma. We will justify how (4.34) follows from $f_0 \in \tilde{L}^{p_0}_{q_0}$ and DCT:

1. From (4.16), we get $1_{B(0,k)}(x) f_0(t,x)\ell^n_t(x) \lesssim 1_{B(0,k)}(x) f_0(t,x)$.

2. From (4.29), we get $1_{B(0,k)}(x) f_0(t,x)\ell^n_t(x) \to 1_{B(0,k)}(x) f_0(t,x)\ell_t(x)$ (as $n \to \infty$) for all $x \in \mathbb{R}^d$.

We denote by I the LHS of the inequality in the statement of Lemma 4.14. Then

$$\begin{aligned} I &\lesssim 1 + \int_{\mathbb{T}} \int_{\mathbb{R}^d} f_0(t,x) \, \mathrm{d}\mu_t(x) \, \mathrm{d}t \\ &\lesssim 1 + \|f_0\|_{\tilde{L}^{p_0}_{q_0}} \quad \text{by (4.36)}. \end{aligned}$$

This completes the proof. □

We have

1. The maps $(t,x) \mapsto b(t,x,\ell_t(x),\mu_t)$ and $(t,x) \mapsto a(t,x,\mu_t)$ are measurable.
2. By Lemma 4.14,
$$\int_{\mathbb{T}} \int_{\mathbb{R}^d} \{|b(t,x,\ell_t(x),\mu_t)| + |a(t,x,\mu_t)|\} \, \mathrm{d}\mu_t(x) \, \mathrm{d}t < \infty.$$
3. The map $\mathbb{T} \to \mathcal{P}_p(\mathbb{R}^d)$, $t \mapsto \mu_t$ is continuous by (4.26).

By superposition principle [Fig08; Tre16; BRS21] as in [BR20, Section 2], (4.1) has a weak solution whose marginal distribution is exactly (μ_t).

4.4.4 Existence of a strong solution

By the previous subsection, there exists a PS $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$ on which there exist an m-BM $(\bar{B}_t)$, an AF $(\bar{\mathcal{F}}_t)$ and a continuous $(\bar{\mathcal{F}}_t)$-adapted process $(\bar{X}_t)$ such that

$$\begin{cases} \mathrm{d}\bar{X}_t = b(t,\bar{X}_t,\ell_t(\bar{X}_t),\mu_t) \, \mathrm{d}t + \sigma(t,\bar{X}_t,\mu_t) \, \mathrm{d}\bar{B}_t, \\ \nu \text{ is the distribution of } \bar{X}_0, \ \mu_t \text{ is that of } \bar{X}_t, \\ \text{and } \ell_t \text{ is the density of } \bar{X}_t. \end{cases}$$

Above, the distribution of $\bar{X}_0$ is ν, that of $\bar{X}_t$ is μ_t, and the density of $\bar{X}_t$ is ℓ_t. We define the map $\bar{b} : \mathbb{T} \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$ by $\bar{b}(t,x,\varrho) := b(t,x,\ell_t(x),\varrho)$. We consider the McKean-Vlasov SDE

$$\begin{cases} \mathrm{d}Y_t = \bar{b}(t,Y_t,\mu'_t) \, \mathrm{d}t + \sigma(t,Y_t,\mu'_t) \, \mathrm{d}B_t, \\ \nu \text{ is the distribution of } Y_0, \text{ and } \mu'_t \text{ is that of } Y_t. \end{cases} \tag{4.37}$$

We recall that (B_t) is the fixed m-BM on the fixed PS $(\Omega, \mathcal{A}, \mathbb{P})$ introduced in Section 2.2. By [HW22, Theorem 1.1(1)], (4.37) is well-posed. On the other hand, $(\bar{X}_t)$ also satisfies (4.37). Then $\mu_t = \mu'_t$ and thus the density of Y_t is also ℓ_t. In particular,

$$\mathrm{d}Y_t = b(t,Y_t,\ell_t(Y_t),\mu_t) \, \mathrm{d}t + \sigma(t,Y_t,\mu_t) \, \mathrm{d}B_t.$$

This completes the proof.

4.5 Uniqueness of a solution

This section is dedicated to the proof of Theorem 4.4. For $k \in \{1,2\}$, we consider the SDE

$$\begin{cases} \mathrm{d}X^k_t = b(t,X^k_t,\ell^k_t(X^k_t),\mu^k_t) \, \mathrm{d}t + \sigma(t,X^k_t) \, \mathrm{d}B^k_t, \\ \nu_k \text{ is the distribution of } X^k_0, \ \mu^k_t \text{ is that of } X^k_t, \\ \text{and } \ell^k_t \text{ is the density of } X^k_t. \end{cases} \tag{4.38}$$

Above, $(B^k_t, t \geq 0)$ is an m-BM. We define measurable maps $b^k : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d$ by $b^k(t,x) := b(t,x,\ell^k_t(x),\mu^k_t)$.

4.5.1 Uniqueness of marginal density

Clearly, (b^k, σ) satisfies Assumption 4.6. We denote $b_t^k(x) := b^k(t, x)$. By Lemma 4.9,

$$\ell_t^k(x) = \int_{\mathbb{R}^d} p_{0,t}^{0,\sigma}(y, x) \ell_{\nu_k}(y) \, \mathrm{d}y + \int_0^t \int_{\mathbb{R}^d} \ell_s^k(y) \langle b_s^k(y), \nabla_y p_{s,t}^{0,\sigma}(y, x) \rangle \, \mathrm{d}y \, \mathrm{d}s.$$

Then

$$\begin{aligned} |\ell_t^2(x) - \ell_t^1(x)| \le & \int_{\mathbb{R}^d} p_{0,t}^{0,\sigma}(y, x) |\ell_{\nu_1}(y) - \ell_{\nu_2}(y)| \, \mathrm{d}y \\ & + \int_0^t \int_{\mathbb{R}^d} |b_s^2(y)| \times |\ell_s^2(y) - \ell_s^1(y)| \times |\nabla_y p_{s,t}^{0,\sigma}(y, x)| \, \mathrm{d}y \, \mathrm{d}s \\ & + \int_0^t \int_{\mathbb{R}^d} \ell_s^1(y) |b_s^2(y) - b_s^1(y)| \times |\nabla_y p_{s,t}^{0,\sigma}(y, x)| \, \mathrm{d}y \, \mathrm{d}s. \end{aligned}$$

We write $M_1 \preccurlyeq M_2$ if there exists a constant $c > 0$ (depending on Θ_1) such that $M_1 \preccurlyeq cM_2$. Thus

$$\begin{aligned} |\ell_t^2(x) - \ell_t^1(x)| \preccurlyeq & \int_{\mathbb{R}^d} p_{0,t}^{0,\sigma}(y, x) |\ell_{\nu_1}(y) - \ell_{\nu_2}(y)| \, \mathrm{d}y \\ & + \int_0^t \int_{\mathbb{R}^d} |\ell_s^2(y) - \ell_s^1(y)| \times |\nabla_y p_{s,t}^{0,\sigma}(y, x)| \, \mathrm{d}y \, \mathrm{d}s \\ & + \int_0^t \int_{\mathbb{R}^d} \ell_s^1(y) |b_s^2(y) - b_s^1(y)| \times |\nabla_y p_{s,t}^{0,\sigma}(y, x)| \, \mathrm{d}y \, \mathrm{d}s. \end{aligned} \tag{4.39}$$

By Corollary 4.8(1),

$$\sup_{t \in \mathbb{T}} \|\ell_t^1\|_\infty \preccurlyeq \|\ell_{\nu_1}\|_\infty. \tag{4.40}$$

By Assumption 4.2(5),

$$|b_s^2(y) - b_s^1(y)| \preccurlyeq |\ell_s^2(y) - \ell_s^1(y)| + W_p(\mu_s^2, \mu_s^1). \tag{4.41}$$

By (4.39), (4.40) and (4.41),

$$\begin{aligned} |\ell_t^2(x) - \ell_t^1(x)| \preccurlyeq & \int_{\mathbb{R}^d} p_{0,t}^{0,\sigma}(y, x) |\ell_{\nu_1}(y) - \ell_{\nu_2}(y)| \, \mathrm{d}y \\ & + (1 + \|\ell_{\nu_1}\|_\infty) \int_0^t \int_{\mathbb{R}^d} |\ell_s^2(y) - \ell_s^1(y)| \times |\nabla_y p_{s,t}^{0,\sigma}(y, x)| \, \mathrm{d}y \, \mathrm{d}s \\ & + \int_0^t W_p(\mu_s^2, \mu_s^1) \int_{\mathbb{R}^d} \ell_s^1(y) |\nabla_y p_{s,t}^{0,\sigma}(y, x)| \, \mathrm{d}y \, \mathrm{d}s \\ =: & \, I_1(t, x) + (1 + \|\ell_{\nu_1}\|_\infty) I_2(t, x) + I_3(t, x). \end{aligned}$$

The pair $(0, \sigma)$ satisfies Assumption 4.6. By Lemma 4.7, there exists a constant $\lambda > 0$ (depending on Θ_1) such that for $i \in \{0, 1\}, 0 \le s < t \le T$ and $x, y \in \mathbb{R}^d$:

$$|\nabla_y^i p_{s,t}^{0,\sigma}(y, x)| \preccurlyeq p_{t-s}^{i,\lambda}(y - x). \tag{4.42}$$

Then

$$\begin{aligned} \int_{\mathbb{R}^d} (|x|^p + 1) |\nabla_y^i p_{s,t}^{0,\sigma}(y, x)| \, \mathrm{d}x & \preccurlyeq \int_{\mathbb{R}^d} (|x|^p + 1) p_{t-s}^{i,\lambda}(y - x) \, \mathrm{d}x \quad \text{by (4.42)} \\ & \preccurlyeq (t - s)^{-\frac{i}{2}} (|y|^p + 1). \end{aligned} \tag{4.43}$$

We define a measurable map $f : \mathbb{T} \to \mathbb{R}_+$ by

$$f(s) := \int_{\mathbb{R}^d} (|x|^p + 1)|\ell_s^2(x) - \ell_s^1(x)| \, \mathrm{d}x.$$

By (4.4), f is bounded. First,

$$\begin{aligned}
&\int_{\mathbb{R}^d} (|x|^p + 1) I_1(t, x) \, \mathrm{d}x \\
&= \int_0^t \int_{\mathbb{R}^d} |\ell_{\nu_1}(y) - \ell_{\nu_2}(y)| \int_{\mathbb{R}^d} (|x|^p + 1) p_{0,t}^{0,\sigma}(y, x) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}s \\
&\preccurlyeq \int_0^t \int_{\mathbb{R}^d} |\ell_{\nu_1}(y) - \ell_{\nu_2}(y)|(|y|^p + 1) \, \mathrm{d}y \, \mathrm{d}s \quad \text{by (4.43)} \\
&\preccurlyeq \int_{\mathbb{R}^d} (|y|^p + 1)|\ell_{\nu_1}(y) - \ell_{\nu_2}(y)| \, \mathrm{d}y = f(0).
\end{aligned}$$

Second,

$$\begin{aligned}
&\int_{\mathbb{R}^d} (|x|^p + 1) I_2(t, x) \, \mathrm{d}x \\
&= \int_0^t \int_{\mathbb{R}^d} |\ell_s^2(y) - \ell_s^1(y)| \int_{\mathbb{R}^d} (|x|^p + 1)|\nabla_y p_{s,t}^{0,\sigma}(y, x)| \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}s \\
&\preccurlyeq \int_0^t (t - s)^{-\frac{1}{2}} \int_{\mathbb{R}^d} |\ell_s^2(y) - \ell_s^1(y)|(|y|^p + 1) \, \mathrm{d}y \, \mathrm{d}s \quad \text{by (4.43)} \\
&= \int_0^t (t - s)^{-\frac{1}{2}} f(s) \, \mathrm{d}s.
\end{aligned}$$

Third,

$$\begin{aligned}
&\int_{\mathbb{R}^d} (|x|^p + 1) I_3(t, x) \, \mathrm{d}x \\
&= \int_0^t W_p(\mu_s^2, \mu_s^1) \int_{\mathbb{R}^d} \ell_s^1(y) \int_{\mathbb{R}^d} (|x|^p + 1)|\nabla_y p_{s,t}^{0,\sigma}(y, x)| \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}s \\
&\preccurlyeq \int_0^t (t - s)^{-\frac{1}{2}} W_p(\mu_s^2, \mu_s^1) \int_{\mathbb{R}^d} \ell_s^1(y)(|y|^p + 1) \, \mathrm{d}y \, \mathrm{d}s \quad \text{by (4.43)} \\
&\preccurlyeq (1 + M_p(\nu_1)) \int_0^t (t - s)^{-\frac{1}{2}} W_p(\mu_s^2, \mu_s^1) \, \mathrm{d}s \quad \text{by (4.4)} \\
&\preccurlyeq (1 + M_p(\nu_1)) \int_0^t (t - s)^{-\frac{1}{2}} |f(s)|^{\frac{1}{p}} \, \mathrm{d}s \quad \text{by Lemma 2.2.}
\end{aligned}$$

To sum up,

$$f(t) \preccurlyeq f(0) + (1 + \|\ell_{\nu_1}\|_\infty + M_p(\nu_1)) \int_0^t (T - s)^{-\frac{1}{2}} (f(s) + |f(s)|^{\frac{1}{p}}) \, \mathrm{d}s.$$

Because $p = 1$, we get

$$f(t) \preccurlyeq f(0) + (1 + \|\ell_{\nu_1}\|_\infty + M_1(\nu_1)) \int_0^t (T - s)^{-\frac{1}{2}} f(s) \, \mathrm{d}s.$$

By Gronwall's lemma,

$$\sup_{t \in \mathbb{T}} f(t) \preccurlyeq f(0) \exp\left\{2\sqrt{T}(1 + \|\ell_{\nu_1}\|_\infty + M_1(\nu_1))\right\}. \tag{4.44}$$

This implies the existence of the function Λ as required in Theorem 4.4(1).

4.5.2 Weak and strong uniqueness of a solution

By (4.38),

$$\mathrm{d}X_t^k = b^k(t, X_t^k)\,\mathrm{d}t + \sigma(t, X_t^k)\,\mathrm{d}B_t^k.$$

Now we let $\nu := \nu_1 = \nu_2$. By (4.44), $\ell_t^1 = \ell_t^2$ and $\mu_t^1 = \mu_t^2$ for $t \in \mathbb{T}$. Then $\boldsymbol{b} := b^1 = b^2$. We consider the SDE

$$\begin{cases} \mathrm{d}Y_t = \boldsymbol{b}(t, Y_t)\,\mathrm{d}t + \sigma(t, Y_t)\,\mathrm{d}B_t, \\ \nu \text{ is the distribution of } Y_0. \end{cases} \tag{4.45}$$

By [HW22, Theorem 1.1(1)], (4.45) is well-posed. On the other hand, (X_t^1) and (X_t^2) satisfy (4.45). It follows that (4.1) has both weak and strong uniqueness.

Chapter 5

Euler-Maruyama scheme for MV-SDEs with density-dependent drift

5.1 Introduction

Let $T \in (0, \infty)$ and $\mathbb{T}$ be the interval $[0, T]$. Consider a measurable function

$$b : \mathbb{T} \times \mathbb{R}^d \times \mathbb{R}_+ \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d.$$

This chapter is about time discretization of the SDE

$$\begin{cases} \mathrm{d}X_t = b(t, X_t, \ell_t(X_t), \mu_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \\ \nu \text{ is the distribution of } X_0,\ \mu_t \text{ is that of } X_t, \\ \text{and } \ell_t \text{ is the density of } X_t. \end{cases} \tag{5.1}$$

The Euler-Maruyama scheme $X^n := (X_t^n, t \in \mathbb{T})$ for (5.1) is constructed as follows. We fix $n \in \mathbb{N}$ and $\varepsilon_n := T/n$. Let $t_k := k\varepsilon_n$ for $k \in [\![0, n]\!]$. Let $\tau_t^n := t_k$ if $t \in [t_k, t_{k+1})$ for some $k \in [\![0, n-1]\!]$. For $t \in \mathbb{T}$, let

$$\begin{cases} X_t^n := X_0 + \int_0^t b(s, X^n_{\tau_s^n}, \ell^n_{\tau_s^n}(X^n_{\tau_s^n}), \mu^n_{\tau_s^n}) 1_{(\varepsilon_n, T]}(s)\,\mathrm{d}s + \sqrt{2} B_t, \\ \mu_t^n \text{ is the distribution of } X_t^n, \text{ and } \ell_t^n \text{ is the density of } X_t^n. \end{cases} \tag{5.2}$$

We consider the following set of assumptions:

Assumption 5.1. There exist constants $\alpha \in (0,1), p \in [1, \infty), C > 0$ such that for $t \in \mathbb{T}; x, y \in \mathbb{R}^d; r, r' \in \mathbb{R}_+$ and $\varrho, \varrho' \in \mathcal{P}_p(\mathbb{R}^d)$:

1. $|b(t, x, r, \varrho)| \le C$.

2. $\nu \in \mathcal{P}_p(\mathbb{R}^d)$ admits a density $\ell_\nu \in C_b^\alpha(\mathbb{R}^d)$.

3. $|b(t, x, r, \varrho) - b(t, x, r', \rho')| \le C\{|r - r'| + W_p(\varrho, \varrho')\}$.

We gather parameters in Assumption 5.1:

$$\Theta_1 := (d, T, \alpha, p, C).$$

First, we prove the following estimates about Hölder regularity:

Theorem 5.2. *Let Assumption 5.1 hold. There exist constants $c_1 > 0$ (depending on Θ_1), $c_2 > 0$ (depending on Θ_1, ν) and $\delta \in (0, \frac{1}{2})$ such that for $0 \le s < t \le T$:*

$$\sup_n \sup_{t \in \mathbb{T}} \|\ell_t^n\|_{C_b^\alpha} \le c_1 \|\ell_\nu\|_{C_b^\alpha}, \tag{5.3}$$

$$\sup_n \|\ell_t^n - \ell_s^n\|_\infty \le c_2 (t-s)^{\frac{\alpha}{2}}, \tag{5.4}$$

$$\sup_n W_p(\mu_t^n, \mu_s^n) \le c_1 (t-s)^\delta, \tag{5.5}$$

$$\sup_n \int_{\mathbb{R}^d} (1+|x|^p)|\ell_t^n(x) - \ell_s^n(x)| \, \mathrm{d}x \le c_2 (t-s)^{\frac{\alpha}{2}} s^{-\frac{\alpha}{2}}. \tag{5.6}$$

If we assume, in addition, that $\int_{\mathbb{R}^d} (1+|x|^p)\sqrt{\ell_\nu(x)} \, \mathrm{d}x \le C$, then

$$\sup_n \int_{\mathbb{R}^d} (1+|x|^p)|\ell_t^n(x) - \ell_s^n(x)| \, \mathrm{d}x \le c_2 (t-s)^{\frac{\alpha}{4}}. \tag{5.7}$$

Estimate (5.6) will be used in the proof of Theorem 5.4, but it explodes as $s \downarrow 0$. As in [SH24], we use time cutoff $1_{(\varepsilon_n, T]}$ in (5.2) to tackle this issue. Well-posedness of (5.1) has been obtained in Chapter 4 via mollifying argument, which in turn relies on [HW22, Theorem 1.1(1)]. We will give an alternative proof of weak existence as a direct application of Theorem 5.2.

Theorem 5.3. *Let Assumption 5.1 hold. Then (5.1) has a weak solution whose marginal density satisfies the estimates in Theorem 5.2.*

Finally, we derive the following rate of convergence:

Theorem 5.4. *Let Assumption 5.1 hold and $p = 1$. There exists a constant $c > 0$ (depending on Θ_1, ν) such that*

$$\sup_{t \in \mathbb{T}} \int_{\mathbb{R}^d} (1+|x|)|\ell_t^n(x) - \ell_t(x)| \, \mathrm{d}x \le c n^{-\frac{\alpha}{2}}.$$

5.2 Some crude estimates

We define $b^n : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d$ by

$$b^n(t,x) := b(t, x, \ell_{\tau_t^n}^n(x), \mu_{\tau_t^n}^n) 1_{(\varepsilon_n, T]}(t). \tag{5.8}$$

Then X^n satisfies

$$\mathrm{d}X_t^n = b^n(t, X_{\tau_t^n}^n) \, \mathrm{d}t + \sqrt{2} \, \mathrm{d}B_t. \tag{5.9}$$

First, we verify that the scheme is well-defined.

Lemma 5.5. *Let Assumption 5.1 hold. Then each X_t^n in (5.2) has a distribution $\mu_t^n \in \mathcal{P}_p(\mathbb{R}^d)$ and admits a density.*

Proof. We fix $t \in (t_k, t_{k+1}]$ for some $k \in [\![0, n-1]\!]$. By induction argument, we assume WLOG that the statement holds for t_k. We have

$$X_t^n = X_{t_k}^n + \int_{t_k}^t b^n(s, X_{t_k}^n) \, \mathrm{d}s + \sqrt{2}(B_t - B_{t_k}).$$

It follows from $M_p(\mu_{t_k}^n) + \|b^n\|_\infty < \infty$ that $M_p(\mu_t^n) < \infty$. It remains to prove that X_t^n admits a density. Let Σ_t^1 be the σ-algebra generated by $(B_s - B_r, t_k \le r < s \le t)$. Let Σ_t^2 be the σ-algebra generated by $X_{t_k}^n$. We define $F_t : \mathbb{R}^d \times \Omega \to \mathbb{R}^d$ by

$$F_t(x, \cdot) := x + \int_{t_k}^t b^n(s, x) \, \mathrm{d}s + \sqrt{2}(B_t - B_{t_k}).$$

Because b^n is bounded, F_t is well-defined and measurable w.r.t $\mathcal{B}(\mathbb{R}^d) \otimes \Sigma_t^1$. First, $X_t^n = F_t(X_{t_k}^n, \cdot)$. Second, $F_t(x, \cdot)$ has a non-degenerate normal distribution. We fix a Lebesgue-null set $A \in \mathcal{B}(\mathbb{R}^d)$. We need to prove $\mathbb{P}[X_t^n \in A] = 0$. We define $\bar{F}_t : \mathbb{R}^d \to \mathbb{R}^d$ by $\bar{F}_t(x) := \mathbb{P}[F_t(x, \cdot) \in A]$. Then $\bar{F}_t = 0$. Because $\mathbb{F}$ is admissible, Σ_t^1 is independent of Σ_t^2. We have

$$
\begin{aligned}
\mathbb{P}[X_t^n \in A] &= \mathbb{P}[F_t(X_{t_k}^n, \cdot) \in A] \\
&= \mathbb{E}[\mathbb{P}[F_t(X_{t_k}^n, \cdot) \in A | \Sigma_t^2]] \\
&= \mathbb{E}[\bar{F}_t(X_{t_k}^n)] \quad \text{by Lemma 2.22} \\
&= 0.
\end{aligned}
$$

This completes the proof. □

We consider the standard Gaussian heat kernel defined for $t > 0$ and $x \in \mathbb{R}^d$ by

$$
p_t(x) := \frac{1}{(4\pi t)^{\frac{d}{2}}} \exp\left(-\frac{|x|^2}{4t}\right).
$$

Then

$$
p_t(x) \le 2^{\frac{d}{2}} p_{2t}(x), \tag{5.10}
$$

$$
p_t(x+y) \le 2^{\frac{d}{2}} \exp\left(\frac{|y|^2}{4t}\right) p_{2t}(x). \tag{5.11}
$$

The following estimates are classical. See e.g. [HRZ21, Lemma 2.1] for their proofs.

Lemma 5.6. 1. *There exists a constant $c > 0$ (depending on d) such that for $t > 0$ and $x, y \in \mathbb{R}^d$:*

$$
|\nabla p_t(x)| \le c t^{-\frac{1}{2}} p_t(x). \tag{5.12}
$$

2. *For $\alpha \in (0,1)$, there exists a constant $c > 0$ (depending on d, T, α) such that for $0 < t \le T; x, y \in \mathbb{R}^d$ and $i \in \{0,1\}$:*

$$
|\nabla^i p_t(x) - \nabla^i p_t(y)| \le c|x-y|^\alpha t^{-\frac{i+\alpha}{2}} \{p_{4t}(x) + p_{4t}(y)\}. \tag{5.13}
$$

3. *For $\alpha \in (0,1)$, there exists a constant $c > 0$ (depending on d, T, α) such that for $0 \le s < t \le T; x \in \mathbb{R}^d$ and $i \in \{0,1\}$:*

$$
|\nabla^i p_t(x) - \nabla^i p_s(x)| \le c|t-s|^{\frac{\alpha}{2}} \{t^{-\frac{i+\alpha}{2}} p_{2t}(x) + s^{-\frac{i+\alpha}{2}} p_{2s}(x)\}. \tag{5.14}
$$

The associated semigroup $(P_t, t > 0)$ is defined for $x \in \mathbb{R}^d$ and $f \in L_+^0(\mathbb{R}^d) \cup L_b^0(\mathbb{R}^d)$ by

$$
P_t f(x) := \int_{\mathbb{R}^d} p_t(x-y) f(y) \, \mathrm{d}y.
$$

Second, we have some crude estimates.

Lemma 5.7. *Let Assumption 5.1 hold. Let $t \in (t_k, t_{k+1}]$ for some $k \in [\![0, n-1]\!]$. There exists a* *constant $c \ge 1$ (depending on Θ_1) such that for $\varphi \in L^0_+(\mathbb{R}^d \times \mathbb{R}^d) \cup L^0_b(\mathbb{R}^d \times \mathbb{R}^d)$:*

$$\mathbb{E}[\varphi(X^n_{t_k}, X^n_t)] = \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(x,y) \ell^n_{t_k}(x) p_{t-t_k}\Big(x - y + \int_{t_k}^t b^n(s,x)\,\mathrm{d}s\Big)\,\mathrm{d}x\,\mathrm{d}y \tag{5.15}$$

$$\le c \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(x,y) \ell^n_{t_k}(x) p_{2(t-t_k)}(x-y)\,\mathrm{d}x\,\mathrm{d}y, \tag{5.16}$$

$$\|\ell^n_t\|_\infty \le c \|\ell^n_{t_k}\|_\infty. \tag{5.17}$$

Proof. Let Σ^1_t, Σ^2_t and F_t be defined as in the proof of Lemma 5.5. We have

$$\begin{aligned}\mathbb{E}[\varphi(X^n_{t_k}, X^n_t)] &= \mathbb{E}[\varphi(X^n_{t_k}, F_t(X^n_{t_k}, \cdot))] \\ &= \mathbb{E}[\mathbb{E}[\varphi(X^n_{t_k}, F_t(X^n_{t_k}, \cdot))|\Sigma^2_t]].\end{aligned} \tag{5.18}$$

We define $\hat{F}_t : \mathbb{R}^d \to \mathbb{R}^d$ by $\hat{F}_t(x) := \mathbb{E}[\varphi(x, F_t(x,\cdot))]$. Notice that $F_t(x,\cdot)$ has a normal distribution with mean $x + \int_{t_k}^t b^n(s,x)\,\mathrm{d}s$ and covariance matrix $2tI_d$. Consequently, the density of $F_t(x,\cdot)$ is $y \mapsto p_{t-t_k}(x - y + \int_{t_k}^t b^n(s,x)\,\mathrm{d}s)$. There exists a constant $c_1 \ge 1$ (depending on Θ_1) such that

$$\hat{F}_t(x) = \int_{\mathbb{R}^d} \varphi(x,y) p_{t-t_k}\Big(x - y + \int_{t_k}^t b^n(s,x)\,\mathrm{d}s\Big)\,\mathrm{d}y \tag{5.19}$$

$$\le c_1 \int_{\mathbb{R}^d} \varphi(x,y) p_{2(t-t_k)}(x-y)\,\mathrm{d}y \quad \text{by (5.11) and Assumption 5.1(1).} \tag{5.20}$$

WLOG, we assume $\varphi \in L^0_b(\mathbb{R}^d \times \mathbb{R}^d)$. Then

$$\mathbb{E}[\varphi(X^n_{t_k}, X^n_t)] = \mathbb{E}[\hat{F}_t(X^n_{t_k})] \quad \text{by (5.18) and Lemma 2.22}$$

$$= \int_{\mathbb{R}^d} \ell^n_{t_k}(x) \hat{F}_t(x)\,\mathrm{d}x \tag{5.21}$$

$$\le c_1 \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(x,y) \ell^n_{t_k}(x) p_{2(t-t_k)}(x-y)\,\mathrm{d}x\,\mathrm{d}y \quad \text{by (5.20)}. \tag{5.22}$$

Thus (5.16) follows. Clearly, (5.19) and (5.21) imply (5.15). The case $\varphi \in L^0_+(\mathbb{R}^d \times \mathbb{R}^d)$ follows from a truncated argument and MCT. It holds for $f \in L^0_+(\mathbb{R}^d)$ that

$$\begin{aligned}\mathbb{E}[f(X^n_t)] &\le c_1 \int_{\mathbb{R}^d \times \mathbb{R}^d} f(y) \ell^n_{t_k}(x) p_{2(t-t_k)}(x-y)\,\mathrm{d}x\,\mathrm{d}y \quad \text{by (5.22)} \\ &\le c_1 \|\ell^n_{t_k}\|_\infty \int_{\mathbb{R}^d \times \mathbb{R}^d} f(y) p_{2(t-t_k)}(x-y)\,\mathrm{d}x\,\mathrm{d}y \\ &= c_1 \|\ell^n_{t_k}\|_\infty \|f\|_{L^1}.\end{aligned}$$

By duality, $\|\ell^n_t\|_\infty \le c_1 \|\ell^n_{t_k}\|_\infty$. This completes the proof. □

Finally, we obtain Duhamel representation of marginal density.

Lemma 5.8. *Let Assumption 5.1 hold. It holds for $x \in \mathbb{R}^d$ that*

$$\ell^n_t(x) = P_t \ell_\nu(x) + \int_0^t \mathbb{E}[\langle b^n(s, X^n_{\tau^n_s}), \nabla p_{t-s}(X^n_s - x)\rangle]\,\mathrm{d}s, \tag{5.23}$$

$$\ell_t(x) = P_t \ell_\nu(x) + \int_0^t \mathbb{E}[\langle b(s, X_s, \ell_s(X_s), \mu_s), \nabla p_{t-s}(X_s - x)\rangle]\,\mathrm{d}s. \tag{5.24}$$

Proof. Let's prove (5.23). We fix $t \in (0,T]$ and $f \in C^\infty_c(\mathbb{R}^d)$. We define

$$g : [0,t] \times \mathbb{R}^d \to \mathbb{R},\ (s,y) \mapsto (P_{t-s} f)(y).$$

Then $(\partial_s + \Delta) g = 0$. We have

$$\begin{aligned} \mathrm{d}g(s, X^n_s) &= \langle b^n(s, X^n_{\tau^n_s}), \nabla g(s, X^n_s)\rangle \,\mathrm{d}s \qquad\qquad \text{by Itô's lemma} \\ &\quad + (\partial_s + \Delta) g(s, X^n_s) \,\mathrm{d}s + \mathrm{d}M_s \\ &= \langle b^n(s, X^n_{\tau^n_s}), \nabla g(s, X^n_s)\rangle \,\mathrm{d}s + \mathrm{d}M_s. \end{aligned}$$

Above, $M_0 = 0$ and $\mathrm{d}M_s = \{\nabla g(s, X^n_s)\}^\top \,\mathrm{d}B_s$. Then

$$f(X^n_t) = g(0, X_0) + \int_0^t \langle b^n(s, X^n_{\tau^n_s}), \nabla g(s, X^n_s)\rangle \,\mathrm{d}s + M_t.$$

To apply Fubini's theorem, we next verify $\sup_{s\in\mathbb{T}} \|\nabla g(s,\cdot)\|_\infty < \infty$. It suffices to prove that the Lipschitz constant of $g(s,\cdot)$ is bounded uniformly in time. We consider the Gaussian process governed by the SDE

$$\mathrm{d}Y^x_{s,t} = \sqrt{2}\,\mathrm{d}B_t, \quad t \in [s, T], Y^x_{s,s} = x.$$

We have

$$\begin{aligned} |g(s,x) - g(s,y)| &= |\mathbb{E}[f(Y^x_{s,t})] - \mathbb{E}[f(Y^y_{s,t})]| \\ &\le \|\nabla f\|_\infty \mathbb{E}[|Y^x_{s,t} - Y^y_{s,t}|] \\ &\le c_1 \|\nabla f\|_\infty |x - y|. \end{aligned}$$

Above, the constant $c_1 > 0$ is given by Lemma 2.18. By Fubini's theorem,

$$\begin{aligned} \int_{\mathbb{R}^d} f(x) \ell^n_t(x) \,\mathrm{d}x &= \mathbb{E}[g(0, X_0)] + \int_0^t \mathbb{E}[\langle b^n(s, X^n_{\tau^n_s}), \nabla g(s, X^n_s)\rangle] \,\mathrm{d}s \\ &=: I_1 + I_2. \end{aligned}$$

By Leibniz integral rule,

$$\nabla g(s,y) = \nabla_y \int_{\mathbb{R}^d} p_{t-s}(y-x) f(x) \,\mathrm{d}x = \int_{\mathbb{R}^d} \nabla p_{t-s}(y-x) f(x) \,\mathrm{d}x.$$

We have

$$\begin{aligned} I_1 &= \int_{\mathbb{R}^d} \ell_\nu(y) \{P_t f\}(y) \,\mathrm{d}y \\ &= \int_{\mathbb{R}^d} f(x) \left(\int_{\mathbb{R}^d} p_t(y-x) \ell_\nu(y) \,\mathrm{d}y \right) \mathrm{d}x \\ &= \int_{\mathbb{R}^d} f(x) \{P_t \ell_\nu\}(x) \,\mathrm{d}x, \\ I_2 &= \int_0^t \mathbb{E}[\langle b^n(s, X^n_{\tau^n_s}), \int_{\mathbb{R}^d} f(x) \nabla p_{t-s}(X^n_s - x) \,\mathrm{d}x \rangle] \,\mathrm{d}s \\ &= \int_{\mathbb{R}^d} f(x) \int_0^t \mathbb{E}[\langle b^n(s, X^n_{\tau^n_s}), \nabla p_{t-s}(X^n_s - x)\rangle] \,\mathrm{d}s \,\mathrm{d}x. \end{aligned}$$

Then (5.23) follows. The proof of (5.24) is a straightforward modification of above reasoning. □

For brevity, we adopt the following conventions in the remaining of this chapter:

1. We write $M_1 \lesssim M_2$ if there exists a constant $c > 0$ (depending on Θ_1) such that $M_1 \le c M_2$.
2. We write $M_1 \preccurlyeq M_2$ if there exists a constant $c > 0$ (depending on Θ_1, ν) such that $M_1 \le c M_2$.

5.3 Stability estimates of the scheme

This section is devoted to the proof of Theorem 5.2 where we use techniques from [Wan23c; HRZ21]. By (5.2),

$$X_t^n - X_s^n = \int_s^t b^n(r, X_{\tau_r^n}^n)\,\mathrm{d}r + \sqrt{2}(B_t - B_s).$$

Then

$$\mathbb{E}[|X_t^n - X_s^n|^p] \lesssim (t-s)^p + \mathbb{E}[|B_{t-s}|^p] \lesssim (t-s)^{\frac{p}{2}}.$$

Thus (5.5) follows.

5.3.1 Bound supremum norm of marginal density

By Lemma 5.8 and Assumption 5.1(1), there exists a constant $c_1 > 0$ (depending on Θ_1) such that

$$\begin{aligned} \ell_t^n(x) &\le \|\ell_\nu\|_\infty + c_1 \int_0^t \mathbb{E}[|\nabla p_{t-s}(X_s^n - x)|]\,\mathrm{d}s \\ &=: \|\ell_\nu\|_\infty + c_1 \int_0^t E_s\,\mathrm{d}s. \end{aligned} \tag{5.25}$$

Step 1: We are going to prove the existence of $m \in [\![1, n]\!]$ such that $\frac{m}{n}$ depends only on Θ_1 and that $\|\ell_{t_k}^n\|_\infty \le 2\|\ell_\nu\|_\infty$ for $k \in [\![0, m]\!]$. The idea is to pick m such that the following induction argument is valid. We fix $k \in [\![1, m]\!]$ and assume that $\|\ell_{t_i}^n\|_\infty \le 2\|\ell_\nu\|_\infty$ for $i \in [\![0, k-1]\!]$. Let $s \in (t_i, t_{i+1})$ for some $i \in [\![0, k-1]\!]$. We have

$$\begin{aligned} E_s &\lesssim \int_{\mathbb{R}^d \times \mathbb{R}^d} \ell_{t_i}^n(y) p_{2(s-t_i)}(y-z) |\nabla p_{t_k - s}(x-z)|\,\mathrm{d}y\,\mathrm{d}z \quad \text{by (5.16)} \\ &\lesssim \frac{1}{\sqrt{t_k - s}} \int_{\mathbb{R}^d \times \mathbb{R}^d} \ell_{t_i}^n(y) p_{2(s-t_i)}(y-z) p_{t_k - s}(x-z)\,\mathrm{d}y\,\mathrm{d}z \quad \text{by (5.12)} \\ &\lesssim \frac{1}{\sqrt{t_k - s}} \int_{\mathbb{R}^d \times \mathbb{R}^d} \ell_{t_i}^n(y) p_{2(s-t_i)}(y-z) p_{2(t_k - s)}(x-z)\,\mathrm{d}z\,\mathrm{d}y \quad \text{by (5.10)} \\ &= \frac{1}{\sqrt{t_k - s}} \int_{\mathbb{R}^d} \ell_{t_i}^n(y) p_{2(t_k - t_i)}(x-y)\,\mathrm{d}y \quad \text{by Chapman–Kolmogorov equation} \\ &\le \frac{2\|\ell_\nu\|_\infty}{\sqrt{t_k - s}} \quad \text{by inductive hypothesis.} \end{aligned} \tag{5.26}$$

By (5.25) and (5.26), there exists a constant $c_2 > 0$ (depending on Θ_1) such that

$$\begin{aligned} \ell_{t_k}^n(x) &\le \|\ell_\nu\|_\infty \left[1 + c_1 c_2 \int_0^{t_k} \frac{\mathrm{d}s}{\sqrt{t_k - s}}\right] \\ &= \|\ell_\nu\|_\infty (1 + 2 c_1 c_2 \sqrt{t_k}). \end{aligned}$$

It suffices to choose m such that

$$c_1 c_2 \sqrt{t_k} \le \frac{1}{2}. \tag{5.27}$$

Clearly, there exists a constant $c_3 \in (0,1)$ (depending on Θ_1) such that if $\frac{m}{n} \le c_3$ then (5.27) holds.

Step 2: Repeating **Step 1** at most $\lceil 1/c_3 \rceil$ times, we have $\|\ell_{t_k}^n\|_\infty \le 2^{\lceil 1/c_3 \rceil} \|\ell_\nu\|_\infty$ for $k \in [\![0, n]\!]$.

Step 3: By (5.17), there exists a constant $c_4 \ge 1$ (depending on Θ_1) such that if $t \in (t_k, t_{k+1})$

then $\|\ell_t^n\|_\infty \le c_4 \|\ell_{t_k}^n\|_\infty$. Thus

$$\sup_{t\in\mathbb{T}} \|\ell_t^n\|_\infty \le c_4 2^{\lceil 1/c_3 \rceil} \|\ell_\nu\|_\infty. \tag{5.28}$$

5.3.2 Hölder continuity in space

We fix $x, x' \in \mathbb{R}^d$. By Lemma 5.8 and Assumption 5.1(1),

$$\begin{aligned}|\ell_t^n(x) - \ell_t^n(x')| &\lesssim |P_t\ell_\nu(x) - P_t\ell_\nu(x')| \\ &\qquad + \int_0^t \mathbb{E}[|\nabla p_{t-s}(X_s^n - x) - \nabla p_{t-s}(X_s^n - x')|]\,\mathrm{d}s \\ &=: I_1 + \int_0^t I_2^n(s)\,\mathrm{d}s. \end{aligned}\tag{5.29}$$

We consider the Gaussian process governed by the SDE

$$\mathrm{d}Y_t^x = \sqrt{2}\,\mathrm{d}B_t, \quad t \in \mathbb{T}, Y_0^x = x.$$

By [HW22, Theorem 1.1(2)],

$$\mathbb{E}\big[\sup_{t\in\mathbb{T}} |Y_t^x - Y_t^{x'}|\big] \lesssim |x - x'|. \tag{5.30}$$

On the other hand,

$$\begin{aligned} I_1 &= |\mathbb{E}[\ell_\nu(Y_t^x)] - \mathbb{E}[\ell_\nu(Y_t^{x'})]| \\ &\le [\ell_\nu]_\alpha \mathbb{E}[|Y_t^x - Y_t^{x'}|^\alpha] \\ &\le [\ell_\nu]_\alpha (\mathbb{E}[|Y_t^x - Y_t^{x'}|])^\alpha \quad \text{by Jensen's inequality} \\ &\lesssim [\ell_\nu]_\alpha |x - x'|^\alpha \quad \text{by (5.30)}. \end{aligned}\tag{5.31}$$

Next we bound $I_2^n(s)$. We have

$$\begin{aligned} I_2^n(s) &\lesssim s^{-\frac{1+\alpha}{2}} |x - x'|^\alpha \mathbb{E}[p_{4(t-s)}(X_s^n - x) + p_{4(t-s)}(X_s^n - x')] \quad \text{by (5.13)} \\ &= s^{-\frac{1+\alpha}{2}} |x - x'|^\alpha \int_{\mathbb{R}^d} \ell_s^n(y)\{p_{4(t-s)}(y - x) + p_{4(t-s)}(y - x')\}\,\mathrm{d}y \\ &\lesssim \|\ell_\nu\|_\infty s^{-\frac{1+\alpha}{2}} |x - x'|^\alpha \quad \text{by (5.28)}. \end{aligned}\tag{5.32}$$

By (5.29), (5.31) and (5.32),

$$\begin{aligned} \frac{|\ell_t^n(x) - \ell_t^n(x')|}{|x - x'|^\alpha} &\lesssim [\ell_\nu]_\alpha + \|\ell_\nu\|_\infty \int_0^t s^{-\frac{1+\alpha}{2}}\,\mathrm{d}s \\ &\lesssim \|\ell_\nu\|_{C_b^\alpha} \quad \text{because } \alpha \in (0,1). \end{aligned}\tag{5.33}$$

Then (5.28) and (5.33) imply (5.3).

5.3.3 Hölder continuity in time

We fix $x \in \mathbb{R}^d$ and $0 \le s < t \le T$.

1. First, we will prove (5.4). By Lemma 5.8,

$$\begin{aligned} \ell_t^n(x) &= P_t\ell_\nu(x) + \int_0^t \mathbb{E}[\langle b^n(r, X^n_{\tau^n_r}), \nabla p_{t-r}(X^n_r - x)\rangle]\,\mathrm{d}r, \\ \ell_s^n(x) &= P_s\ell_\nu(x) + \int_0^s \mathbb{E}[\langle b^n(r, X^n_{\tau^n_r}), \nabla p_{s-r}(X^n_r - x)\rangle]\,\mathrm{d}r. \end{aligned} \tag{5.34}$$

By (5.34) and Assumption 5.1(1),

$$\begin{aligned} |\ell_t^n(x) - \ell_s^n(x)| &\lesssim |P_t\ell_\nu(x) - P_s\ell_\nu(x)| \\ &\qquad + \int_0^s \mathbb{E}[|\nabla p_{t-r}(X^n_r - x) - \nabla p_{s-r}(X^n_r - x)|]\,\mathrm{d}r \\ &\qquad\qquad + \int_s^t \mathbb{E}[|\nabla p_{t-r}(X^n_r - x)|]\,\mathrm{d}r \\ &=: I_1(x) + I_2^n(x) + I_3^n(x). \end{aligned} \tag{5.35}$$

It holds for $i \in \{0, 1\}$ that

$$\begin{aligned} &\nabla^i p_{t-r}(y-x) - \nabla^i p_{s-r}(y-x) \\ &= \nabla^i_y \int_{\mathbb{R}^d} p_{s-r}(y-z)p_{t-s}(z-x)\,\mathrm{d}z - \nabla^i p_{s-r}(y-x) \\ &= \nabla^i_y \int_{\mathbb{R}^d} p_{s-r}(y-z)p_{t-s}(z-x)\,\mathrm{d}z - \nabla^i_y \int_{\mathbb{R}^d} p_{s-r}(y-x)p_{t-s}(z-x)\,\mathrm{d}z \\ &= \int_{\mathbb{R}^d} \{\nabla^i p_{s-r}(y-z) - \nabla^i p_{s-r}(y-x)\} p_{t-s}(z-x)\,\mathrm{d}z. \end{aligned} \tag{5.36}$$

First,

$$\begin{aligned} I_1(x) &= \left| \int_{\mathbb{R}^d} \ell_\nu(y)\{p_t(y-x) - p_s(y-x)\}\,\mathrm{d}y \right| \\ &= \left| \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} \ell_\nu(y)\{p_s(y-z) - p_s(y-x)\}\,\mathrm{d}y \right] p_{t-s}(z-x)\,\mathrm{d}z \right| \quad \text{by (5.36)} \\ &\le \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \ell_\nu(y)\{p_s(y-z) - p_s(y-x)\}\,\mathrm{d}y \right| p_{t-s}(z-x)\,\mathrm{d}z \\ &= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} p_s(y)\{\ell_\nu(y+z) - \ell_\nu(y+x)\}\,\mathrm{d}y \right| p_{t-s}(z-x)\,\mathrm{d}z \\ &\preccurlyeq \int_{\mathbb{R}^d} |z-x|^\alpha p_{t-s}(z-x)\,\mathrm{d}z \quad \text{because } \ell_\nu \in C_b^\alpha(\mathbb{R}^d) \\ &\lesssim (t-s)^{\frac{\alpha}{2}}. \end{aligned} \tag{5.37}$$

Second,

$$\begin{aligned} I_2^n(x) &= \int_0^s \mathbb{E}\left[\left| \int_{\mathbb{R}^d} \{\nabla p_{s-r}(X^n_r - z) - \nabla p_{s-r}(X^n_r - x)\} p_{t-s}(z-x)\,\mathrm{d}z \right| \right] \mathrm{d}r \quad \text{by (5.36)} \\ &\le \int_0^s \int_{\mathbb{R}^d} \mathbb{E}[|\nabla p_{s-r}(X^n_r - z) - \nabla p_{s-r}(X^n_r - x)|] p_{t-s}(z-x)\,\mathrm{d}z\,\mathrm{d}r \\ &\lesssim \int_0^s (s-r)^{-\frac{1+\alpha}{2}} \int_{\mathbb{R}^d} |z-x|^\alpha \mathbb{E}[p_{4(s-r)}(X^n_r - z)] p_{t-s}(z-x)\,\mathrm{d}z\,\mathrm{d}r \qquad \text{by (5.13)} \\ &\qquad + \int_0^s (s-r)^{-\frac{1+\alpha}{2}} \int_{\mathbb{R}^d} |z-x|^\alpha \mathbb{E}[p_{4(s-r)}(X^n_r - x)] p_{t-s}(z-x)\,\mathrm{d}z\,\mathrm{d}r \\ &=: I_{21}^n(x) + I_{22}^n(x). \end{aligned}$$

We have

$$
\begin{aligned}
I_{21}^n(x) &= \int_0^s (s-r)^{-\frac{1+\alpha}{2}} \int_{\mathbb{R}^d\times\mathbb{R}^d} |z-x|^\alpha \ell_s^n(y) p_{4(s-r)}(y-z) p_{t-s}(z-x)\,\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}r \\
&\preccurlyeq \int_0^s (s-r)^{-\frac{1+\alpha}{2}} \int_{\mathbb{R}^d\times\mathbb{R}^d} |z-x|^\alpha p_{4(s-r)}(y-z) p_{t-s}(z-x)\,\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}r \quad \text{by (5.28)} \\
&\lesssim \int_0^s (s-r)^{-\frac{1+\alpha}{2}} \int_{\mathbb{R}^d\times\mathbb{R}^d} |z-x|^\alpha p_{4(s-r)}(y-z) p_{4(t-s)}(z-x)\,\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}r \quad \text{by (5.10)} \\
&= \int_0^s (s-r)^{-\frac{1+\alpha}{2}} \int_{\mathbb{R}^d} |z-x|^\alpha p_{4(t-s)}(z-x)\,\mathrm{d}z\,\mathrm{d}r \\
&\lesssim (t-s)^{\frac{\alpha}{2}} \int_0^s (s-r)^{-\frac{1+\alpha}{2}}\,\mathrm{d}r. \qquad (5.38)
\end{aligned}
$$

Similarly,

$$
I_{22}^n(x) \preccurlyeq (t-s)^{\frac{\alpha}{2}} \int_0^s (s-r)^{-\frac{1+\alpha}{2}}\,\mathrm{d}r. \qquad (5.39)
$$

Third,

$$
\begin{aligned}
I_3^n(x) &= \int_s^t \int_{\mathbb{R}^d} \ell_r^n(y) |\nabla p_{t-r}(y-x)|\,\mathrm{d}y\,\mathrm{d}r \\
&\lesssim \int_s^t \frac{1}{\sqrt{t-r}} \int_{\mathbb{R}^d} \ell_r^n(y) p_{t-r}(y-x)\,\mathrm{d}y\,\mathrm{d}r \quad \text{by (5.12)} \qquad (5.40) \\
&\preccurlyeq \int_s^t \frac{1}{\sqrt{t-r}} \int_{\mathbb{R}^d} p_{t-r}(y-x)\,\mathrm{d}y\,\mathrm{d}r \quad \text{by (5.28)} \\
&= \int_s^t \frac{\mathrm{d}r}{\sqrt{t-r}} \lesssim \sqrt{t-s} \lesssim (t-s)^{\frac{\alpha}{2}}. \qquad (5.41)
\end{aligned}
$$

By (5.37), (5.38), (5.39) and (5.41),

$$
\begin{aligned}
|\ell_t^n(x) - \ell_s^n(x)| &\preccurlyeq (t-s)^{\frac{\alpha}{2}} \left[1 + \sup_{s\in\mathbb{T}} \int_0^s (s-r)^{-\frac{1+\alpha}{2}}\,\mathrm{d}r \right] \\
&\lesssim (t-s)^{\frac{\alpha}{2}} \quad \text{because } \alpha \in (0,1).
\end{aligned}
$$

This implies (5.4).

2. Second, we will prove (5.6). We have

$$
\begin{aligned}
I_1(x) &\le \int_{\mathbb{R}^d} \ell_\nu(y) |p_t(y-x) - p_s(y-x)|\,\mathrm{d}y \\
&\lesssim |t-s|^{\frac{\alpha}{2}} \int_{\mathbb{R}^d} \ell_\nu(y) \{ t^{-\frac{\alpha}{2}} p_{2t}(y-x) + s^{-\frac{\alpha}{2}} p_{2s}(y-x) \}\,\mathrm{d}y \quad \text{by (5.14)} \\
&\lesssim (t-s)^{\frac{\alpha}{2}} s^{-\frac{\alpha}{2}} \int_{\mathbb{R}^d} \ell_\nu(y) \{ p_{2t}(y-x) + p_{2s}(y-x) \}\,\mathrm{d}y.
\end{aligned}
$$

Then

$$
\begin{aligned}
&\int_{\mathbb{R}^d} (1+|x|^p) I_1(x)\,\mathrm{d}x \\
&\lesssim (t-s)^{\frac{\alpha}{2}} s^{-\frac{\alpha}{2}} \int_{\mathbb{R}^d} \ell_\nu(y) \left[\int_{\mathbb{R}^d} (1+|x|^p) \{ p_{2t}(y-x) + p_{2s}(y-x) \}\,\mathrm{d}x \right] \mathrm{d}y \\
&\lesssim (t-s)^{\frac{\alpha}{2}} s^{-\frac{\alpha}{2}} \int_{\mathbb{R}^d} (1+|y|^p) \ell_\nu(y)\,\mathrm{d}y \\
&\preccurlyeq (t-s)^{\frac{\alpha}{2}} s^{-\frac{\alpha}{2}}. \qquad (5.42)
\end{aligned}
$$

We have

$$\begin{aligned} I_2^n(x) &\lesssim |t-s|^{\frac{\alpha}{2}} \int_0^s \mathbb{E}[(t-r)^{-\frac{1+\alpha}{2}} p_{2(t-r)}(X_r^n - x) \qquad \text{by (5.14)} \\ &\qquad + (s-r)^{-\frac{1+\alpha}{2}} p_{2(s-r)}(X_r^n - x)] \,\mathrm{d}r \\ &\lesssim |t-s|^{\frac{\alpha}{2}} \int_0^s (s-r)^{-\frac{1+\alpha}{2}} \mathbb{E}[p_{2(t-r)}(X_r^n - x)] \,\mathrm{d}r. \end{aligned}$$

Then

$$\begin{aligned} &\int_{\mathbb{R}^d} (1+|x|^p) I_2^n(x) \,\mathrm{d}x \\ &\lesssim |t-s|^{\frac{\alpha}{2}} \int_0^s (s-r)^{-\frac{1+\alpha}{2}} \mathbb{E}\Big[\int_{\mathbb{R}^d} (1+|x|^p) p_{2(t-r)}(X_r^n - x) \,\mathrm{d}x \Big] \,\mathrm{d}r \\ &\lesssim |t-s|^{\frac{\alpha}{2}} \int_0^s (s-r)^{-\frac{1+\alpha}{2}} \mathbb{E}[1+|X_r^n|^p] \,\mathrm{d}r \\ &\preccurlyeq |t-s|^{\frac{\alpha}{2}} \int_0^s (s-r)^{-\frac{1+\alpha}{2}} \,\mathrm{d}r \quad \text{by (5.5)} \\ &\lesssim |t-s|^{\frac{\alpha}{2}} \quad \text{because } \alpha \in (0,1). \end{aligned} \tag{5.43}$$

We have

$$\begin{aligned} &\int_{\mathbb{R}^d} (1+|x|^p) I_3^n(x) \,\mathrm{d}x \\ \lesssim &\int_s^t \frac{1}{\sqrt{t-r}} \int_{\mathbb{R}^d} \ell_r^n(y) \Big[\int_{\mathbb{R}^d} (1+|x|^p) p_{t-r}(y-x) \,\mathrm{d}x \Big] \,\mathrm{d}y \,\mathrm{d}r \quad \text{by (5.40)} \\ \lesssim &\int_s^t \frac{1}{\sqrt{t-r}} \Big[\int_{\mathbb{R}^d} (1+|y|^p) \ell_r^n(y) \,\mathrm{d}y \Big] \,\mathrm{d}r \\ \preccurlyeq &\int_s^t \frac{\mathrm{d}r}{\sqrt{t-r}} \quad \text{by (5.5)} \\ \lesssim &\sqrt{t-s} \lesssim (t-s)^{\frac{\alpha}{2}}. \end{aligned} \tag{5.44}$$

We have (5.35) together with (5.42), (5.43) and (5.44) implies (5.6).

3. We assume, in addition, that $\int_{\mathbb{R}^d} (1+|x|^p) \sqrt{\ell_\nu(x)} \,\mathrm{d}x \le C$ holds. Finally, we will prove (5.7). Let Y be a random variable whose density is p_1. Then

$$\begin{aligned} I_1(x) &= |\mathbb{E}[\ell_\nu(x+\sqrt{t}Y) - \ell_\nu(x+\sqrt{s}Y)]| \\ &\le \mathbb{E}[\{|\ell_\nu(x+\sqrt{t}Y) - \ell_\nu(x+\sqrt{s}Y)|^2\}^{\frac{1}{2}}] \\ &\preccurlyeq \mathbb{E}[|(\sqrt{t}-\sqrt{s})Y|^{\frac{\alpha}{2}} \times |\ell_\nu(x+\sqrt{t}Y) - \ell_\nu(x+\sqrt{s}Y)|^{\frac{1}{2}}] \quad \text{because } \ell_\nu \in C_b^\alpha(\mathbb{R}^d) \\ &\lesssim (t-s)^{\frac{\alpha}{4}} \mathbb{E}[|Y|^{\frac{\alpha}{2}} \{|\ell_\nu(x+\sqrt{t}Y)|^{\frac{1}{2}} + |\ell_\nu(x+\sqrt{s}Y)|^{\frac{1}{2}}\}]. \end{aligned}$$

Thus

$$\begin{aligned}
&\int_{\mathbb{R}^d} (1+|x|^p) I_1(x) \, \mathrm{d}x \\
&\preccurlyeq (t-s)^{\frac{\alpha}{4}} \mathbb{E}[|Y|^{\frac{\alpha}{2}} \int_{\mathbb{R}^d} (1+|x|^p) \{ |\ell_\nu(x+\sqrt{t}Y)|^{\frac{1}{2}} + |\ell_\nu(x+\sqrt{s}Y)|^{\frac{1}{2}} \} \, \mathrm{d}x] \\
&\lesssim (t-s)^{\frac{\alpha}{4}} \mathbb{E}[|Y|^{\frac{\alpha}{2}} \int_{\mathbb{R}^d} (1+|x|^p+|Y|^p) \sqrt{\ell_\nu(x)} \, \mathrm{d}x] \\
&\preccurlyeq (t-s)^{\frac{\alpha}{4}} \mathbb{E}[|Y|^{\frac{\alpha}{2}} (1+|Y|^p)] \\
&\le (t-s)^{\frac{\alpha}{4}}.
\end{aligned} \tag{5.45}$$

We have (5.35) together with (5.45), (5.43) and (5.44) implies (5.7). This completes the proof.

5.4 Existence of a weak solution

This section is dedicated to the proof of Theorem 5.3.

5.4.1 Convergence of marginal densities

By Theorem 5.2 and Arzelà–Ascoli theorem, there exist a sub-sequence (also denoted by (ℓ^n) for simplicity) and a continuous function $\ell : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}_+$ such that

$$\lim_n \sup_{t \in \mathbb{T}} \sup_{x \in B(0,R)} |\ell^n_t(x) - \ell_t(x)| = 0 \quad \text{for} \quad R > 0. \tag{5.46}$$

Above, $\ell_t \coloneqq \ell(t, \cdot)$. Clearly, $\ell_0 = \ell_\nu$ and

$$\begin{aligned}
\sup_{t \in \mathbb{T}} \|\ell_t\|_{C^\alpha_b} &\lesssim \|\ell_\nu\|_{C^\alpha_b}, \\
\|\ell_t - \ell_s\|_\infty &\preccurlyeq |t-s|^{\frac{\alpha}{2}} \quad \text{for} \quad s, t \in \mathbb{T}.
\end{aligned}$$

As in Section 4.4.2, we get

1. ℓ_t is a density whose induced distribution $\mu_t \in \mathcal{P}_p(\mathbb{R}^d)$.
2. There exist a function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ depending on (Θ_1, ν) and a constant $\delta \in (0, \frac{1}{2})$ such that $\lim_{R \to \infty} \phi(R) = 0$ and

$$\lim_n \sup_{t \in [R^{-1}, T]} W_p(\mu^n_t, \mu_t) = 0, \tag{5.47}$$

$$W_p(\mu_t, \mu_s) \preccurlyeq |t-s|^\delta, \tag{5.48}$$

$$\sup_{n \in \mathbb{N}} \sup_{t \in \mathbb{T}} \mu^n_t(B^c_R) \preccurlyeq \phi(R), \tag{5.49}$$

$$\sup_{t \in \mathbb{T}} \mu_t(B^c_R) \preccurlyeq \phi(R) \quad \text{for} \quad R > T^{-1}. \tag{5.50}$$

5.4.2 Existence of a weak solution

We fix $(f, g) \in C^\infty_c(0,T) \times C^\infty_c(\mathbb{R}^d)$. We apply Itô's lemma to (5.9) and the map $(t,x) \mapsto f(t)g(x)$. Then

$$\begin{aligned}
\mathbb{E}[\int_0^t f'(s) g(X^n_s) \, \mathrm{d}s] = &- \mathbb{E}[\int_0^t \langle f(s) \nabla g(X^n_s), b^n(s, X^n_{\tau^n_s}) \rangle \, \mathrm{d}s] \\
&- \mathbb{E}[\int_0^t f(s) \Delta g(X^n_s) \, \mathrm{d}s].
\end{aligned} \tag{5.51}$$

Let

$$
\begin{aligned}
I_1^n(s) &:= \mathbb{E}[g(X_s^n)], \\
I_3^n(s) &:= \mathbb{E}[\Delta g(X_s^n)], \\
I_2^n(s) &:= \mathbb{E}[\langle \nabla g(X_s^n), b^n(s, X_{\tau_s^n}^n)\rangle] \\
&= \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \nabla g(y), b^n(s,x)\rangle \ell_{\tau_s^n}^n(x) \qquad \text{by (5.15)} \\
&\qquad \times p_{s-\tau_s^n}\left(x - y + \int_{\tau_s^n}^s b_r^n(x)\,\mathrm{d}r\right)\mathrm{d}x\,\mathrm{d}y.
\end{aligned} \tag{5.52}
$$

By (5.51) and Fubini's theorem,

$$
\int_0^t f'(s) I_1^n(s)\,\mathrm{d}s = -\int_0^t f(s) I_2^n(s)\,\mathrm{d}s - \int_0^t f(s) I_3^n(s)\,\mathrm{d}s. \tag{5.53}
$$

It holds for n large enough that $\operatorname{supp} f \subset (\varepsilon_n, T)$. WLOG, we assume $s \in (\varepsilon_n, T)$ and thus $b^n(s,x) = b(s, x, \ell_{\tau_s^n}^n(x), \mu_{\tau_s^n}^n)$. By (5.47),

$$
\begin{aligned}
\lim_n I_1^n(s) &= \int_{\mathbb{R}^d} g(x)\ell(s,x)\,\mathrm{d}x, \\
\lim_n I_3^n(s) &= \int_{\mathbb{R}^d} \Delta g(x)\ell(s,x)\,\mathrm{d}x.
\end{aligned} \tag{5.54}
$$

Next we consider $I_2^n(s)$. For $x \in \mathbb{R}^d$, we define $\nu^{n,x} \in \mathcal{P}(\mathbb{R}^d)$, $h^n(x) \in \mathbb{R}$ and $I_2(s) \in \mathbb{R}$ by

$$
\mathrm{d}\nu^{n,x}(y) := p_{s-\tau_s^n}\left(x - y + \int_{\tau_s^n}^s b^n(r,x)\,\mathrm{d}r\right)\mathrm{d}y, \tag{5.55}
$$

$$
h^n(x) := \int_{\mathbb{R}^d} \langle \nabla g(y), b(s, x, \ell_{\tau_s^n}^n(x), \mu_{\tau_s^n}^n)\rangle\,\mathrm{d}\nu^{n,x}(y), \tag{5.56}
$$

$$
I_2(s) := \int_{\mathbb{R}^d} \ell_s(x)\langle \nabla g(x), b(s,x,\ell_s(x),\mu_s)\rangle\,\mathrm{d}x. \tag{5.57}
$$

Then

$$
\begin{aligned}
\int_{\mathbb{R}^d} |x-y|\,\mathrm{d}\nu^{n,x}(y) &= \int_{\mathbb{R}^d} |y| p_{s-\tau_s^n}\left(y + \int_{\tau_s^n}^s b^n(r,x)\,\mathrm{d}r\right)\mathrm{d}y, \\
&\lesssim \sqrt{s-\tau_s^n} \quad \text{by Assumption 5.1(1) and (5.11)},
\end{aligned} \tag{5.58}
$$

$$
I_2^n(s) = \int_{\mathbb{R}^d} \ell_{\tau_s^n}^n(x) h^n(x)\,\mathrm{d}x \quad \text{by (5.52), (5.55) and (5.56)}. \tag{5.59}
$$

WLOG, we assume $\|f\|_\infty + \|\nabla g\|_\infty + \|\nabla^2 g\|_\infty \le 1$. By (5.57) and (5.59),

$$
\begin{aligned}
|I_2^n(s) - I_2(s)| \lesssim &\int_{\mathbb{R}^d} |\ell_{\tau_s^n}^n(x) - \ell_s(x)| \times |\langle \nabla g(x), b(s,x,\ell_s(x),\mu_s)\rangle|\,\mathrm{d}x \\
&+ \int_{\mathbb{R}^d} \ell_{\tau_s^n}^n(x) |h^n(x) - \langle \nabla g(x), b(s,x,\ell_s(x),\mu_s)\rangle|\,\mathrm{d}x.
\end{aligned} \tag{5.60}
$$

First,

$$
\begin{aligned}
&|h^n(x) - \langle \nabla g(x), b(s,x,\ell_s(x),\mu_s)\rangle| \\
&\lesssim \int_{\mathbb{R}^d} |\nabla g(y) - \nabla g(x)| \, \mathrm{d}\nu^{n,x}(y) \qquad \text{by (5.56)} \\
&\quad + |b(s,x,\ell^n_{\tau^n_s}(x),\mu^n_{\tau^n_s}) - b(s,x,\ell_s(x),\mu_s)| \\
&\lesssim \int_{\mathbb{R}^d} |x-y| \, \mathrm{d}\nu^{n,x}(y) + |\ell^n_{\tau^n_s}(x) - \ell_s(x)| + W_p(\mu^n_{\tau^n_s},\mu_s) \quad \text{by Assumption 5.1(3)} \\
&\lesssim \sqrt{s-\tau^n_s} + |\ell^n_{\tau^n_s}(x) - \ell_s(x)| + W_p(\mu^n_{\tau^n_s},\mu_s) \quad \text{by (5.58).}
\end{aligned} \tag{5.61}
$$

Let $S := \operatorname{supp} g$. Then S is compact. By (5.60) and (5.61),

$$
\begin{aligned}
|I^n_2(s) - I_2(s)| \lesssim \int_S |\ell^n_{\tau^n_s}(x) - \ell_s(x)| \, \mathrm{d}x + \sqrt{s-\tau^n_s} \\
+ \int_{\mathbb{R}^d} \ell^n_{\tau^n_s}(x) |\ell^n_{\tau^n_s}(x) - \ell_s(x)| \, \mathrm{d}x + W_p(\mu^n_{\tau^n_s},\mu_s).
\end{aligned} \tag{5.62}
$$

By (5.4) and (5.46), it holds for $R>0$ that

$$
\lim_n \sup_{x\in B(0,R)} |\ell^n_{\tau^n_s}(x) - \ell_s(x)| = 0. \tag{5.63}
$$

By (5.5) and (5.47),

$$
\lim_n W_p(\mu^n_{\tau^n_s},\mu_s) = 0. \tag{5.64}
$$

We have

$$
\begin{aligned}
&\limsup_n |I^n_2(s) - I_2(s)| \\
&\lesssim \limsup_n \int_{\mathbb{R}^d} \ell^n_{\tau^n_s}(x) |\ell^n_{\tau^n_s}(x) - \ell_s(x)| \, \mathrm{d}x \quad \text{by (5.62), (5.63) and (5.64)} \\
&\preccurlyeq \limsup_n \int_{\mathbb{R}^d} |\ell^n_{\tau^n_s}(x) - \ell_s(x)| \, \mathrm{d}x \quad \text{by (5.3)} \\
&\le \limsup_n \int_{B(0,R)} |\ell^n_{\tau^n_s}(x) - \ell_s(x)| \, \mathrm{d}x + \limsup_n \int_{B^c_R} |\ell^n_{\tau^n_s}(x) - \ell_s(x)| \, \mathrm{d}x \\
&=: \limsup_n I^n_{41}(s,R) + \limsup_n I^n_{42}(s,R) \quad \text{for} \quad R>0.
\end{aligned}
$$

By (5.63), $\limsup_n I^n_{41}(s,R) = 0$. By (5.49) and (5.50), $\limsup_n I^n_{42}(s,R) \preccurlyeq \phi(R)$. Then $\limsup_n |I^n_2(s) - I_2(s)| \preccurlyeq \phi(R)$ and thus $\lim_n I^n_2(s) = I_2(s)$. This together with (5.53), (5.54) and DCT implies

$$
\begin{aligned}
\int_0^t \int_{\mathbb{R}^d} f'(s) g(x) \ell(s,x) \, \mathrm{d}x \, \mathrm{d}s = &- \int_0^t \int_{\mathbb{R}^d} f(s) \langle \nabla g(x), b(s,x,\ell(s,x),\mu_s) \ell(s,x)\rangle \, \mathrm{d}x \, \mathrm{d}s \\
&- \int_0^t \int_{\mathbb{R}^d} f(s) \Delta g(x) \ell(s,x) \, \mathrm{d}x \, \mathrm{d}s.
\end{aligned}
$$

Hence ℓ satisfies the following Fokker-Planck equation in distributional sense

$$
\partial_t \ell(t,x) = -\sum_{i=1}^d \partial_{x_i} \{ b^i(t,x,\ell(t,x),\mu_t) \ell(t,x)\} + \Delta \ell(t,x).
$$

Clearly,

1. $(t,x) \mapsto b(t,x,\ell(t,x),\mu_t)$ is measurable with

$$\int_{\mathbb{T}} \int_{\mathbb{R}^d} |b(t,x,\ell(t,x),\mu_t)| \, \mathrm{d}\mu_t(x) \, \mathrm{d}t < \infty.$$

2. $\mathbb{T} \to \mathcal{P}_p(\mathbb{R}^d)$, $t \mapsto \mu_t$ is continuous by (5.48).

We apply superposition principle as in [BR20, Section 2] and get that (5.1) has a weak solution whose marginal distribution is exactly $(\mu_t, t \in \mathbb{T})$. This completes the proof.

5.5 Rate of convergence

This section is dedicated to the proof of Theorem 5.4.

5.5.1 Decomposition of error

For $(s,z) \in \mathbb{T} \times \mathbb{R}^d$, we define

$$\begin{aligned}
I_1^n(s,z) &:= \mathbb{E}[\langle b^n(s, X^n_{\tau^n_s}), \nabla p_{t-s}(X^n_s - z) - \nabla p_{t-s}(X^n_{\tau^n_s} - z)\rangle], \\
I_2^n(s,z) &:= \mathbb{E}[\langle b^n(s, X^n_{\tau^n_s}), \nabla p_{t-s}(X^n_{\tau^n_s} - z)\rangle] - \mathbb{E}[\langle b^n(s, X^n_s), \nabla p_{t-s}(X^n_s - z)\rangle], \\
I_3^n(s,z) &:= 1_{(\varepsilon_n, T]}(s) \mathbb{E}[\langle b(s, X^n_s, \ell^n_{\tau^n_s}(X^n_s), \mu^n_{\tau^n_s}) - b(s, X^n_s, \ell^n_s(X^n_s), \mu^n_s), \nabla p_{t-s}(X^n_s - z)\rangle], \\
I_4^n(s,z) &:= 1_{[0, \varepsilon_n]}(s) \mathbb{E}[\langle b(s, X^n_s, \ell^n_s(X^n_s), \mu^n_s), \nabla p_{t-s}(X^n_s - z)\rangle].
\end{aligned}$$

We also define the density-like function $\hat{\ell}^n_t : \mathbb{R}^d \to \mathbb{R}$ by

$$\hat{\ell}^n_t(z) := P_t \ell_\nu(z) + \int_0^t \mathbb{E}[\langle b(s, X^n_s, \ell^n_s(X^n_s), \mu^n_s), \nabla p_{t-s}(X^n_s - z)\rangle] \, \mathrm{d}s.$$

By Lemma 5.8,

$$\ell^n_t(z) - \hat{\ell}^n_t(z) = \int_0^t \{I_1^n(s,z) + I_2^n(s,z) + I_3^n(s,z) - I_4^n(s,z)\} \, \mathrm{d}s.$$

We define $f, \hat{f} : \mathbb{T} \to \mathbb{R}_+$ by

$$\begin{aligned}
f(t) &:= \int_{\mathbb{R}^d} (1 + |z|^p) |\ell^n_t(z) - \ell_t(z)| \, \mathrm{d}z, \\
\hat{f}(t) &:= \int_{\mathbb{R}^d} (1 + |z|^p) |\hat{\ell}^n_t(z) - \ell_t(z)| \, \mathrm{d}z.
\end{aligned}$$

By (5.5) and (5.48), f is bounded. Clearly,

$$f(t) \le \hat{f}(t) + \int_0^t \int_{\mathbb{R}^d} (1 + |z|^p) \{|I_1^n(s,z)| + |I_2^n(s,z)| + |I_3^n(s,z)| + |I_4^n(s,z)|\} \, \mathrm{d}z \, \mathrm{d}s. \tag{5.65}$$

We have

$$\begin{aligned}
|I_1^n(s,z)| &\lesssim \mathbb{E}[|\nabla p_{t-s}(X^n_s - z) - \nabla p_{t-s}(X^n_{\tau^n_s} - z)|] \quad \text{by Assumption 5.1(1)} \\
&\lesssim \frac{1}{\sqrt{t-s}} \mathbb{E}[|X^n_s - X^n_{\tau^n_s}|^\alpha \{p_{4(t-s)}(X^n_s - z) + p_{4(t-s)}(X^n_{\tau^n_s} - z)\}] \quad \text{by (5.13)} \\
&\lesssim \frac{1}{\sqrt{t-s}} \int_{(\mathbb{R}^d)^2} |x-y|^\alpha \{p_{4(t-s)}(y-z) + p_{4(t-s)}(x-z)\} \quad \text{by (5.16)} \\
&\qquad \times \ell^n_{\tau^n_s}(x) p_{2(s-\tau^n_s)}(x-y) \, \mathrm{d}x \, \mathrm{d}y.
\end{aligned}$$

There exists a constant $\kappa > 0$ (depending on Θ_1) such that

$$\int_{\mathbb{R}^d} (1+|z|^p)\{p_{4(t-s)}(y-z) + p_{4(t-s)}(x-z)\}\,\mathrm{d}z \lesssim 1 + |x|^p + |y|^p,$$
$$|x-y|^\alpha p_{2(s-\tau_s^n)}(x-y) \lesssim (s-\tau_s^n)^{\frac{\alpha}{2}} p_{\kappa(s-\tau_s^n)}(x-y).$$

Then

$$\begin{aligned}
&\int_{\mathbb{R}^d} (1+|z|^p)|I_1^n(s,z)|\,\mathrm{d}z \\
&\lesssim \frac{(s-\tau_s^n)^{\frac{\alpha}{2}}}{\sqrt{t-s}} \int_{(\mathbb{R}^d)^2} (1+|x|^p+|y|^p)\ell^n_{\tau_s^n}(x) p_{\kappa(s-\tau_s^n)}(x-y)\,\mathrm{d}x\,\mathrm{d}y \\
&\lesssim \frac{(s-\tau_s^n)^{\frac{\alpha}{2}}}{\sqrt{t-s}} \int_{\mathbb{R}^d} (1+|x|^p)\ell^n_{\tau_s^n}(x)\,\mathrm{d}x \\
&\preccurlyeq \frac{(s-\tau_s^n)^{\frac{\alpha}{2}}}{\sqrt{t-s}} \quad \text{by (5.5).}
\end{aligned} \tag{5.66}$$

We have

$$\begin{aligned}
|I_2^n(s,z)| &\le \int_{\mathbb{R}^d} |\langle b^n(s,x), \nabla p_{t-s}(x-z)\rangle| \times |\ell^n_{\tau_s^n}(x) - \ell^n_s(x)|\,\mathrm{d}x \\
&\lesssim \frac{1_{(\varepsilon_n,T]}(s)}{\sqrt{t-s}} \int_{\mathbb{R}^d} p_{t-s}(x-z)|\ell^n_{\tau_s^n}(x) - \ell^n_s(x)|\,\mathrm{d}x \quad \text{by (5.8), (5.12) and Assumption 5.1(1).}
\end{aligned}$$

Then

$$\begin{aligned}
&\int_{\mathbb{R}^d} (1+|z|^p)|I_2^n(s,z)|\,\mathrm{d}z \\
&\lesssim \frac{1_{(\varepsilon_n,T]}(s)}{\sqrt{t-s}} \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} (1+|z|^p) p_{t-s}(x-z)\,\mathrm{d}z\right] |\ell^n_{\tau_s^n}(x) - \ell^n_s(x)|\,\mathrm{d}x \\
&\lesssim \frac{1_{(\varepsilon_n,T]}(s)}{\sqrt{t-s}} \int_{\mathbb{R}^d} (1+|x|^p)|\ell^n_{\tau_s^n}(x) - \ell^n_s(x)|\,\mathrm{d}x \\
&\preccurlyeq \frac{(s-\tau_s^n)^{\frac{\alpha}{2}} 1_{(\varepsilon_n,T]}(s)}{|\tau_s^n|^{\frac{\alpha}{2}}\sqrt{t-s}} \quad \text{by (5.6)} \\
&\lesssim \frac{(s-\tau_s^n)^{\frac{\alpha}{2}} 1_{(\varepsilon_n,T]}(s)}{(s-\varepsilon_n)^{\frac{\alpha}{2}}\sqrt{t-s}}.
\end{aligned} \tag{5.67}$$

We have

$$\begin{aligned}
|I_3^n(s,z)| &\lesssim \mathbb{E}[|\nabla p_{t-s}(X_s^n - z)|]\{\|\ell^n_{\tau_s^n} - \ell^n_s\|_\infty + W_p(\mu^n_{\tau_s^n}, \mu^n_s)\} \quad \text{by Assumption 5.1(3)} \\
&\lesssim \frac{\|\ell^n_{\tau_s^n} - \ell^n_s\|_\infty + W_p(\mu^n_{\tau_s^n}, \mu^n_s)}{\sqrt{t-s}} \mathbb{E}[p_{t-s}(X_s^n - z)] \quad \text{by (5.12)} \\
&\preccurlyeq \frac{(s-\tau_s^n)^{\frac{\alpha}{2}}}{\sqrt{t-s}} \mathbb{E}[p_{t-s}(X_s^n - z)] \quad \text{by Theorem 5.2.}
\end{aligned}$$

Then

$$\begin{aligned}\int_{\mathbb{R}^d}(1+|z|^p)|I_3^n(s,z)|\,\mathrm{d}z &\preccurlyeq \frac{(s-\tau_s^n)^{\frac{\alpha}{2}}}{\sqrt{t-s}}\mathbb{E}\Big[\int_{\mathbb{R}^d}(1+|z|^p)p_{t-s}(X_s^n-z)\,\mathrm{d}z\Big]\\ &\lesssim \frac{(s-\tau_s^n)^{\frac{\alpha}{2}}}{\sqrt{t-s}}\mathbb{E}[1+|X_s^n|^p]\\ &\preccurlyeq \frac{(s-\tau_s^n)^{\frac{\alpha}{2}}}{\sqrt{t-s}} \quad \text{by (5.5)}. \end{aligned} \tag{5.68}$$

By Assumption 5.1(1) and (5.12),

$$|I_4^n(s,z)| \lesssim \frac{1_{[0,\varepsilon_n]}(s)}{\sqrt{t-s}}\mathbb{E}[p_{t-s}(X_s^n-z)].$$

Then

$$\begin{aligned}\int_{\mathbb{R}^d}(1+|z|^p)|I_4^n(s,z)|\,\mathrm{d}z &\lesssim \frac{1_{[0,\varepsilon_n]}(s)}{\sqrt{t-s}}\mathbb{E}\Big[\int_{\mathbb{R}^d}(1+|z|^p)p_{t-s}(X_s^n-z)\,\mathrm{d}z\Big]\\ &\lesssim \frac{1_{[0,\varepsilon_n]}(s)}{\sqrt{t-s}}\mathbb{E}[1+|X_s^n|^p]\\ &\preccurlyeq \frac{1_{[0,\varepsilon_n]}(s)}{\sqrt{t-s}} \quad \text{by (5.5)}. \end{aligned} \tag{5.69}$$

5.5.2 Bound weighted total variation norm

As in Section 4.5.1, we have

$$\begin{aligned}\hat{f}(t) &\lesssim (1+\|\ell_\nu\|_\infty + M_p(\nu))\int_0^t (T-s)^{-\frac{1}{2}}\{f(s)+|f(s)|^{\frac{1}{p}}\}\,\mathrm{d}s\\ &\preccurlyeq \int_0^t (T-s)^{-\frac{1}{2}}\{f(s)+|f(s)|^{\frac{1}{p}}\}\,\mathrm{d}s. \end{aligned} \tag{5.70}$$

By (5.65), (5.66), (5.67), (5.68), (5.69) and (5.70),

$$\begin{aligned} f(t) &\preccurlyeq (s-\tau_s^n)^{\frac{\alpha}{2}}\int_0^t \left\{\frac{1}{\sqrt{t-s}} + \frac{1_{[0,\varepsilon_n]}(s)}{(s-\tau_s^n)^{\frac{\alpha}{2}}\sqrt{t-s}} + \frac{1_{(\varepsilon_n,T]}(s)}{(s-\varepsilon_n)^{\frac{\alpha}{2}}\sqrt{t-s}}\right\}\mathrm{d}s\\ &\qquad + \int_0^t (T-s)^{-\frac{1}{2}}\{f(s)+|f(s)|^{\frac{1}{p}}\}\,\mathrm{d}s\\ &\lesssim (s-\tau_s^n)^{\frac{\alpha}{2}} + \int_0^t (T-s)^{-\frac{1}{2}}\{f(s)+|f(s)|^{\frac{1}{p}}\}\,\mathrm{d}s. \end{aligned}$$

Because $p=1$ and $s-\tau_s^n \le \frac{1}{n}$,

$$f(t) \preccurlyeq n^{-\frac{\alpha}{2}} + \int_0^t (T-s)^{-\frac{1}{2}} f(s)\,\mathrm{d}s.$$

By Gronwall's lemma,

$$\sup_{t\in\mathbb{T}} f(t) \preccurlyeq n^{-\frac{\alpha}{2}}.$$

This completes the thesis.

Bibliography

[ABS21] Luigi Ambrosio, Elia Brué, and Daniele Semola. *Lectures on optimal transport*. Springer, 2021.

[AG13] Luigi Ambrosio and Nicola Gigli. "A User's Guide to Optimal Transport". In: *Modelling and Optimisation of Flows on Networks: Cetraro, Italy 2009, Editors: Benedetto Piccoli, Michel Rascle*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 1–155.

[AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Second. Springer Science & Business Media, 2008.

[AH10] Krishna B Athreya and Chii-Ruey Hwang. "Gibbs measures asymptotics". In: *Sankhya A* 72 (2010), pp. 191–207.

[AK89] Emile Aarts and Jan Korst. *Simulated annealing and Boltzmann machines: a stochastic approach to combinatorial optimization and neural computing*. John Wiley & Sons, Inc., 1989.

[And82] Brian DO Anderson. "Reverse-time diffusion equation models". In: *Stochastic Processes and their Applications* 12.3 (1982), pp. 313–326.

[AS07] Luigi Ambrosio and Giuseppe Savaré. "Chapter 1 - Gradient Flows of Probability Measures". In: ed. by C.M. Dafermos and E. Feireisl. Vol. 3. Handbook of Differential Equations: Evolutionary Equations. North-Holland, 2007, pp. 1–136.

[ASZ19] Sergio A. Almada Monter, Mykhaylo Shkolnikov, and Jiacheng Zhang. "Dynamics of observables in rank-based models and performance of functionally generated portfolios". In: *The Annals of Applied Probability* 29.5 (Oct. 2019).

[Bal17] Paolo Baldi. *Stochastic Calculus: An Introduction Through Theory and Exercises*. 1st. Universitext. Springer International Publishing, 2017.

[Bar10] Viorel Barbu. *Nonlinear differential equations of monotone types in Banach spaces*. Springer Science & Business Media, 2010.

[Bar16] Viorel Barbu. "Generalized solutions to nonlinear Fokker–Planck equations". en. In: *Journal of Differential Equations* 261.4 (Aug. 2016), pp. 2446–2471.

[Bar18] Viorel Barbu. "Generalized Solutions to Nonlinear Fokker–Planck Equations with Linear Drift". In: *Stochastic Partial Differential Equations and Related Fields*. Ed. by Andreas Eberle et al. Vol. 229. Series Title: Springer Proceedings in Mathematics & Statistics. Cham: Springer International Publishing, 2018, pp. 293–302.

[Bar21] Viorel Barbu. *Semigroup Approach to Nonlinear Diffusion Equations*. en. World Scientific, Oct. 2021.

[BB00] Jean-David Benamou and Yann Brenier. "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem". In: *Numerische Mathematik* 84.3 (2000), pp. 375–393.

[BB18] Adrien Blanchet and Jérôme Bolte. "A family of functional inequalities: Łojasiewicz inequalities and displacement convex functions". In: *Journal of Functional Analysis* 275.7 (2018), pp. 1650–1673.

[BBC75] Philippe Benilan, Haim Brezis, and Michael G. Crandall. "A semilinear equation in $L^1(\mathbb{R}^N)$". en. In: *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze* 2.4 (1975), pp. 523–555.

[BC79] Haiem Brezis and Michael G. Crandall. "Uniqueness of Solutions of the Initial-Value Problem for $u_t - \Delta\varphi(u) = 0$". In: *Journal de Mathématiques Pures et Appliquées* 58.9 (1979), pp. 153–163.

[BC81] Philippe Bénilan and Michael G Crandall. "The continuous dependence on φ of solutions of $u_t - \Delta\varphi(u) = 0$". In: *Indiana University Mathematics Journal* 30.2 (1981), pp. 161–177.

[BCR13] Nadia Belaribi, François Cuvelier, and Francesco Russo. "Probabilistic and deterministic algorithms for space multidimensional irregular porous media equation". en. In: *Stochastic Partial Differential Equations: Analysis and Computations* 1.1 (Mar. 2013), pp. 3–62.

[BÉ85] Dominique Bakry and Michel Émery. "Diffusions hypercontractives". fr. In: *Séminaire de probabilités de Strasbourg* 19 (1985), pp. 177–206.

[Ben+96] Said Benachour et al. "Processus associés à l'équation des milieux poreux". In: *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze* 23.4 (1996), pp. 793–832.

[Ben+98] S. Benachour et al. "Nonlinear self-stabilizing processes – I Existence, invariant probability, propagation of chaos". en. In: *Stochastic Processes and their Applications* 75.2 (July 1998), pp. 173–201.

[Ber+10] Frédéric Bernardin et al. "Stochastic Lagrangian method for downscaling problems in computational fluid dynamics". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 44.5 (Sept. 2010), pp. 885–920.

[Bic02] Klaus Bichteler. *Stochastic integration with jumps.* 89. Cambridge University Press, 2002.

[BJ15] Mireille Bossy and Jean-François Jabir. "Lagrangian stochastic models with specular boundary condition". en. In: *Journal of Functional Analysis* 268.6 (Mar. 2015), pp. 1309–1381.

[BJ18] Mireille Bossy and Jean-François Jabir. "Particle approximation for Lagrangian Stochastic Models with specular boundary condition". In: *Electronic Communications in Probability* 23.none (Jan. 2018).

[BJ19] Mireille Bossy and Jean-François Jabir. "On the Wellposedness of Some McKean Models with Moderated or Singular Diffusion Coefficient". en. In: *Frontiers in Stochastic Analysis–BSDEs, SPDEs and their Applications.* Ed. by Samuel N. Cohen et al. Vol. 289. Series Title: Springer Proceedings in Mathematics & Statistics. Cham: Springer International Publishing, 2019, pp. 43–87.

[BJ22] Oumaima Bencheikh and Benjamin Jourdain. "Convergence in Total Variation of the Euler-Maruyama Scheme Applied to Diffusion Processes with Measurable Drift Coefficient and Additive Noise". en. In: *SIAM Journal on Numerical Analysis* 60.4 (Aug. 2022), pp. 1701–1740.

[BJT11] Mireille Bossy, Jean-François Jabir, and Denis Talay. "On conditional McKean Lagrangian stochastic models". en. In: *Probability Theory and Related Fields* 151.1-2 (Oct. 2011), pp. 319–351.

[BL08] C Le Bris and P-L Lions. “Existence and uniqueness of solutions to Fokker-Planck type equations with irregular coefficients”. In: *Communications in Partial Differential Equations* 33.7 (2008), pp. 1272–1317.

[BM19a] Martin Bauer and Thilo Meyer-Brandis. “Existence and regularity of solutions to multi-dimensional mean-field stochastic differential equations with irregular drift”. In: *arXiv preprint arXiv:1912.05932* (2019).

[BM19b] Martin Bauer and Thilo Meyer-Brandis. *McKean-Vlasov equations on infinite-dimensional Hilbert spaces with irregular drift and additive fractional noise.* arXiv:1912.07427 [math]. Dec. 2019.

[BMP18] Martin Bauer, Thilo Meyer-Brandis, and Frank Proske. “Strong solutions of mean-field stochastic differential equations with irregular drift”. In: *Electronic Journal of Probability* 23.none (Jan. 2018).

[BMV24] Jérôme Bolte, Laurent Miclo, and Stéphane Villeneuve. “Swarm gradient dynamics for global optimization: the mean-field limit case”. In: *Mathematical Programming* 205.1 (2024), pp. 661–701.

[Bog+15] V. I. Bogachev et al., eds. *Fokker-Planck-Kolmogorov equations.* Mathematical surveys and monographs volume 207. Providence, Rhode Island: American Mathematical Society, 2015.

[BR18] Viorel Barbu and Michael Röckner. “Probabilistic Representation for Solutions to Nonlinear Fokker-Planck Equations”. In: *SIAM Journal on Mathematical Analysis* 50.4 (2018), pp. 4246–4260.

[BR20] Viorel Barbu and Michael Röckner. “From nonlinear Fokker-Planck equations to solutions of distribution dependent SDE”. In: *The Annals of Probability* 48.4 (2020), pp. 1902–1920.

[BR21a] Viorel Barbu and Michael Röckner. “Solutions for nonlinear Fokker-Planck equations with measures as initial data and McKean-Vlasov equations”. In: *Journal of Functional Analysis* 280.7 (2021), p. 108926.

[BR21b] Viorel Barbu and Michael Röckner. “Uniqueness for nonlinear Fokker-Planck equations and weak uniqueness for McKean-Vlasov SDEs”. In: *Stochastics and Partial Differential Equations: Analysis and Computations* 9 (2021), pp. 702–713.

[BR23a] Viorel Barbu and Michael Röckner. “Nonlinear Fokker-Planck equations with time-dependent coefficients”. In: *SIAM Journal on Mathematical Analysis* 55.1 (2023), pp. 1–18.

[BR23b] Viorel Barbu and Michael Röckner. “Uniqueness for nonlinear Fokker–Planck equations and for McKean–Vlasov SDEs: The degenerate case”. en. In: *Journal of Functional Analysis* 285.4 (Aug. 2023), p. 109980.

[BR24a] Viorel Barbu and Michael Röckner. *Nonlinear Fokker-Planck Flows and their Probabilistic Counterparts.* en. Vol. 2353. Lecture Notes in Mathematics. Springer Nature Switzerland, 2024.

[BR24b] Viorel Barbu and Michael Röckner. “Nonlinear Fokker–Planck equations with fractional Laplacian and McKean–Vlasov SDEs with Lévy noise”. en. In: *Probability Theory and Related Fields* (Apr. 2024).

[Bra22] Pierre Bras. “Convergence rates of Gibbs measures with degenerate minimum”. In: *Bernoulli* 28.4 (2022), pp. 2431–2458.

[Bre11] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations.* en. New York, NY: Springer New York, 2011.

[Bre87] Yann Brenier. “Décomposition polaire et réarrangement monotone des champs de vecteurs”. In: *Comptes rendus de l’Académie des sciences. Série 1, Mathématique* 305 (1987), pp. 805–808.

[BRR10] Philippe Blanchard, Michael Röckner, and Francesco Russo. “Probabilistic representation for solutions of an irregular porous media type equation”. In: *The Annals of Probability* 38.5 (Sept. 2010).

[BRR11] Viorel Barbu, Michael Röckner, and Francesco Russo. “Probabilistic representation for solutions of an irregular porous media type equation: the degenerate case”. en. In: *Probability Theory and Related Fields* 151.1-2 (Oct. 2011), pp. 1–43.

[BRR17] Viorel Barbu, Michael Röckner, and Francesco Russo. “Doubly probabilistic representation for the stochastic porous media type equation”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 53.4 (Nov. 2017).

[BRS21] Vladimir I. Bogachev, Michael Röckner, and Stanislav V. Shaposhnikov. “On the Ambrosio–Figalli–Trevisan Superposition Principle for Probability Solutions to Fokker–Planck–Kolmogorov Equations”. en. In: *Journal of Dynamics and Differential Equations* 33.2 (June 2021), pp. 715–739.

[BRV98] S. Benachour, B. Roynette, and P. Vallois. “Nonlinear self-stabilizing processes – II: Convergence to invariant probability”. en. In: *Stochastic Processes and their Applications* 75.2 (July 1998), pp. 203–224.

[BSR23] Viorel Barbu, José Luís da Silva, and Michael Röckner. *Nonlocal, nonlinear Fokker-Planck equations and nonlinear martingale problems*. arXiv:2308.06388 [math]. Aug. 2023.

[BT96] Mireille Bossy and Denis Talay. “Convergence rate for the approximation of the limit law of weakly interacting particles: application to the Burgers equation”. In: *The Annals of Applied Probability* 6.3 (Aug. 1996).

[BT97] Mireille Bossy and Denis Talay. “A stochastic particle method for the McKean-Vlasov and the Burgersequation”. en. In: *Mathematics of Computation* 66.217 (Jan. 1997), pp. 157–193.

[Car19] Pierre Cardaliaguet. *The master equation and the convergence problem in mean field games*. Annals of mathematics studies 201. Princeton, NJ: Princeton University Press, 2019.

[CB18] Lénaı-c Chizat and Francis Bach. “On the Global Convergence of Gradient Descent for Over-parameterized Models using Optimal Transport”. In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio et al. Vol. 31. Curran Associates, Inc., 2018.

[CD18] René Carmona and François Delarue. *Probabilistic theory of mean field games with applications I-II*. Springer, 2018.

[CD22a] Louis-Pierre Chaintron and Antoine Diez. “Propagation of chaos: A review of models, methods and applications. I. Models and methods”. In: *Kinetic and Related Models* 15.6 (2022), p. 895.

[CD22b] Louis-Pierre Chaintron and Antoine Diez. “Propagation of chaos: A review of models, methods and applications. II. Applications”. In: *Kinetic and Related Models* 15.6 (2022), p. 1017.

[CF22] Paul-Eric Chaudru De Raynal and Noufel Frikha. “Well-posedness for some non-linear SDEs and related PDE on the Wasserstein space”. en. In: *Journal de Mathématiques Pures et Appliquées* 159 (Mar. 2022), pp. 1–167.

[CGL23] Haoyang Cao, Xin Guo, and Joon Seok Lee. “Approximation of N-player stochastic games with singular controls by mean field games”. In: *Numerical Algebra, Control and Optimization* 13.3-4 (2023), pp. 604–629.

[Cha20] P.E. Chaudru De Raynal. “Strong well posedness of McKean–Vlasov stochastic differential equations with Hölder drift”. en. In: *Stochastic Processes and their Applications* 130.1 (Jan. 2020), pp. 79–107.

[Chi22] Lénaı-c Chizat. “Mean-Field Langevin Dynamics : Exponential Convergence and Annealing”. In: *Transactions on Machine Learning Research* (2022).

[CKS91] Tzuu-Shuh Chiang, Gopinath Kallianpur, and P Sundar. “Propagation of chaos and the McKean-Vlasov equation in duals of nuclear spaces”. In: *Applied Mathematics and Optimization* 24.1 (1991), pp. 55–83.

[Cou63] Philippe Courrege. “Intégrales stochastiques et martingales de carré intégrable”. In: *Séminaire Brelot-Choquet-Deny. Théorie du potentiel* 7 (1963), pp. 1–20.

[CP83] P. Calderoni and M. Pulvirenti. “Propagation of chaos for Burgers’ equation”. en. In: *Annales de l’I.H.P. Physique théorique* 39.1 (1983), pp. 85–97.

[Cra17] Katy Craig. “Nonconvex gradient flow in the Wasserstein metric and applications to constrained nonlocal interactions”. In: *Proceedings of the London Mathematical Society* 114.1 (2017), pp. 60–102.

[CRW23] Fan Chen, Zhenjie Ren, and Songbo Wang. “Entropic Fictitious Play for Mean Field Optimization Problem”. In: *Journal of Machine Learning Research* 24.211 (2023), pp. 1–36.

[CW13] K.L. Chung and R.J. Williams. *Introduction to Stochastic Integration*. Modern Birkhäuser Classics. Springer New York, 2013.

[CX10] Dan Crisan and Jie Xiong. “Approximate McKean-Vlasov representations for a class of SPDEs”. In: *Stochastics An International Journal of Probability and Stochastics Processes* 82.1 (2010), pp. 53–68.

[Daw83] Donald A Dawson. “Critical dynamics and fluctuations for a mean-field model of cooperative behavior”. In: *Journal of Statistical Physics* 31.1 (1983), pp. 29–85.

[Deg18] Pierre Degond. “Mathematical models of collective dynamics and self-organization”. In: *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*. World Scientific. 2018, pp. 3925–3946.

[Der03] Azzouz Dermoune. “Propagation and conditional propagation of chaos for pressureless gas equations”. en. In: *Probability Theory and Related Fields* 126.4 (Aug. 2003), pp. 459–476.

[DGI22] Tiziano De Angelis, Maximilien Germain, and Elena Issoglio. “A numerical scheme for stochastic differential equations with distributional drift”. en. In: *Stochastic Processes and their Applications* 154 (Dec. 2022), pp. 55–90.

[DH23] Chang-Song Deng and Xing Huang. *Well-Posedness for McKean-Vlasov SDEs with Distribution Dependent Stable Noises*. arXiv:2306.10970 [math]. June 2023.

[Din00] Nicolae Dinculeanu. *Vector integration and stochastic integration in Banach spaces*. Vol. 48. John Wiley & Sons, 2000.

[DM70] Catherine Doléans-Dade and Paul-André Meyer. “Intégrales stochastiques par rapport aux martingales locales”. fr. In: *Séminaire de probabilités de Strasbourg* 4 (1970), pp. 77–107.

[Doo53] J.L. Doob. *Stochastic Processes*. Probability and Statistics Series. Wiley, 1953.

[DV95] Donald Dawson and Jean Vaillancourt. “Stochastic McKean-Vlasov equations”. In: *Nonlinear Differential Equations and Applications NoDEA* 2.2 (1995), pp. 199–229.

[EM24] Matthias Erbar and Giulia Meglioli. *Gradient flow for a class of diffusion equations with Dirichlet boundary data*. arXiv:2408.05987 [math]. Aug. 2024.

[EMS18] Murat A Erdogdu, Lester Mackey, and Ohad Shamir. “Global non-convex optimization with discretized diffusions”. In: *Advances in Neural Information Processing Systems* 31 (2018).

[Ern22] Xavier Erny. “Well-posedness and propagation of chaos for McKean–Vlasov equations with jumps and locally Lipschitz coefficients”. en. In: *Stochastic Processes and their Applications* 150 (Aug. 2022), pp. 192–214.

[Eva22] L.C. Evans. *Partial Differential Equations*. Graduate Studies in Mathematics. American Mathematical Society, 2022.

[Fan+22] Xiliang Fan et al. “Distribution dependent SDEs driven by fractional Brownian motions”. en. In: *Stochastic Processes and their Applications* 151 (Sept. 2022), pp. 23–67.

[FBL14] A. Friedman, Z.W. Birnbaum, and E. Lukacs. *Stochastic Differential Equations and Applications: Volume 1*. v. 1. Elsevier Science, 2014.

[FF13] Ennio Fedrizzi and Franco Flandoli. “Hölder flow and differentiability for SDEs with nonregular drift”. In: *Stochastic Analysis and Applications* 31.4 (2013), pp. 708–736.

[FG21] Alessio Figalli and Federico Glaudo. *An invitation to optimal transport, Wasserstein distances, and gradient flows*. eng. EMS Textbooks in Mathematics. Berlin: EMS Press, 2021.

[Fig08] Alessio Figalli. “Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients”. en. In: *Journal of Functional Analysis* 254.1 (Jan. 2008), pp. 109–153.

[FJM24] Mathis Fitoussi, Benjamin Jourdain, and Stéphane Menozzi. *Weak well-posedness and weak discretization error for stable-driven SDEs with Lebesgue drift*. arXiv:2405.08378 [math]. May 2024.

[FM97] Begona Fernandez and Sylvie Méléard. “A Hilbertian approach for fluctuations on the McKean-Vlasov model”. In: *Stochastic processes and their applications* 71.1 (1997), pp. 33–53.

[FPZ19] Franco Flandoli, Enrico Priola, and Giovanni Zanco. “A mean-field model with discontinuous coefficients for neurons with spatial interaction”. In: *Discrete and Continuous Dynamical Systems* 39.6 (2019), pp. 3037–3067.

[Fun84] Tadahisa Funaki. “A certain class of diffusion processes associated with nonlinear parabolic equations”. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 67.3 (1984), pp. 331–348.

[FZ05] Shizan Fang and Tusheng Zhang. “A study of a class of stochastic differential equations with non-Lipschitzian coefficients”. In: *Probability Theory and Related Fields* 132 (2005), pp. 356–390.

[Gär88] Jürgen Gärtner. “On the McKean-Vlasov limit for interacting diffusions”. In: *Mathematische Nachrichten* 137.1 (1988), pp. 197–248.

[GGZ22] Xuefeng Gao, Mert Gürbüzbalaban, and Lingjiong Zhu. “Global convergence of stochastic gradient hamiltonian monte carlo for nonconvex stochastic optimization: Nonasymptotic performance bounds and momentum-based acceleration”. In: *Operations Research* 70.5 (2022), pp. 2931–2947.

[GHM22] Lucio Galeati, Fabian A. Harang, and Avi Mayorcas. "Distribution dependent SDEs driven by additive continuous noise". In: *Electronic Journal of Probability* 27.none (Jan. 2022).

[GHM23] Lucio Galeati, Fabian A. Harang, and Avi Mayorcas. "Distribution dependent SDEs driven by additive fractional Brownian motion". en. In: *Probability Theory and Related Fields* 185.1-2 (Feb. 2023), pp. 251–309.

[GK83] E Gutkin and M Kac. "Propagation of chaos and the Burgers equation". In: *SIAM Journal on Applied Mathematics* 43.4 (1983), pp. 971–980.

[GM01] Istvan Gyongy and Teresa Martinez. "On Stochastic Differential Equations with Locally Unbounded Drift". en. In: *Czechoslovak Mathematical Journal* 51.4 (Dec. 2001), pp. 763–783.

[GM91] Saul B Gelfand and Sanjoy K Mitter. "Recursive stochastic algorithms for global optimization in $\mathbb{R}^d$". In: *SIAM Journal on Control and Optimization* 29.5 (1991), pp. 999–1018.

[GM93] Carl Graham and Sylvie Méléard. "Propagation of chaos for a fully connected loss network with alternate routing". In: *Stochastic processes and their applications* 44.1 (1993), pp. 159–180.

[GM94a] Carl Graham and Sylvie Méléard. "Chaos hypothesis for a system interacting through shared resources". In: *Probability Theory and Related Fields* 100 (1994), pp. 157–174.

[GM94b] Carl Graham and Sylvie Méléard. "Fluctuations for a fully connected loss network with alternate routing". In: *Stochastic Processes and their Applications* 53.1 (1994), pp. 97–115.

[GM97] Carl Graham and Sylvie Méléard. "Stochastic particle approximations for generalized Boltzmann models and convergence estimates". In: *The Annals of probability* 25.1 (1997), pp. 115–132.

[Gra92] Carl Graham. "McKean-Vlasov Itô–Skorohod equations, and nonlinear diffusions with discrete jump sets". In: *Stochastic processes and their applications* 40.1 (1992), pp. 69–82.

[Gru23] Sebastian Grube. "Strong solutions to McKean–Vlasov SDEs with coefficients of Nemytskii-type". In: *Electronic Communications in Probability* 28.none (Jan. 2023).

[Gru24] Sebastian Grube. "Strong solutions to McKean–Vlasov SDEs with coefficients of Nemytskii type: the time-dependent case". en. In: *Journal of Evolution Equations* 24.2 (June 2024), p. 37.

[Han22] Yi Han. *Solving McKean-Vlasov SDEs via relative entropy.* arXiv:2204.05709 [math]. Oct. 2022.

[Hao23] Zimo Hao. "McKean-Vlasov SDEs with singular drifts". PhD thesis. Universität Bielefeld, 2023, 1051153 bytes.

[HHL24] Wei Hong, Shanshan Hu, and Wei Liu. "McKean–Vlasov SDE and SPDE with locally monotone coefficients". In: *The Annals of Applied Probability* 34.2 (Apr. 2024).

[HKS89] Richard A Holley, Shigeo Kusuoka, and Daniel W Stroock. "Asymptotics of the spectral gap with applications to the theory of simulated annealing". In: *J. Funct. Anal* 83.2 (1989), pp. 333–347.

[HLS22] Martin Herdegen, Gechun Liang, and Osian Shelley. "Vague and weak convergence of signed measures". In: *arXiv preprint arXiv:2205.13207* (2022).

[HM86] Masuyuki Hitsuda and Itaru Mitoma. "Tightness problem and stochastic evolution equation arising from fluctuation phenomena for interacting diffusions". In: *Journal of Multivariate Analysis* 19.2 (1986), pp. 311–328.

[HRZ21] Zimo Hao, Michael Röckner, and Xicheng Zhang. "Euler scheme for density dependent stochastic differential equations". In: *Journal of Differential Equations* 274 (2021), pp. 996–1014.

[HRZ24] Zimo Hao, Michael Röckner, and Xicheng Zhang. "Strong Convergence of Propagation of Chaos for McKean–Vlasov SDEs with Singular Interactions". en. In: *SIAM Journal on Mathematical Analysis* 56.2 (Apr. 2024), pp. 2661–2713.

[HS87] Richard Holley and Daniel Stroock. "Logarithmic Sobolev inequalities and stochastic Ising models". en. In: *Journal of Statistical Physics* 46.5-6 (Mar. 1987), pp. 1159–1194.

[HŠS21a] William R. P. Hammersley, David Šiška, and Łukasz Szpruch. "McKean–Vlasov SDEs under measure dependent Lyapunov conditions". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 57.2 (May 2021).

[HŠS21b] William R. P. Hammersley, David Šiška, and Łukasz Szpruch. "Weak existence and uniqueness for McKean–Vlasov SDEs with common noise". In: *The Annals of Probability* 49.2 (Mar. 2021).

[Hu+21] Kaitong Hu et al. "Mean-field Langevin dynamics and energy landscape of neural networks". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 57.4 (Nov. 2021).

[Hua21] Xing Huang. "Path-distribution dependent SDEs with singular coefficients". In: *Electronic Journal of Probability* 26.none (Jan. 2021).

[Hua23] Xing Huang. *McKean-Vlasov SDEs with Sigularity in Distribution Variable and Distribution Dependent Diffusion.* arXiv:2302.05845 [math]. May 2023.

[HW19] Xing Huang and Feng-Yu Wang. "Distribution dependent SDEs with singular coefficients". In: *Stochastic Processes and their Applications* 129.11 (2019), pp. 4747–4770.

[HW21a] Xing Huang and Feng-Yu Wang. "Derivative estimates on distributions of McKean-Vlasov SDEs". In: *Electronic Journal of Probability* 26.none (Jan. 2021).

[HW21b] Xing Huang and Feng-Yu Wang. "Mckean-Vlasov sdes with drifts discontinuous under wasserstein distance". en. In: *Discrete & Continuous Dynamical Systems* 41.4 (2021), pp. 1667–1679.

[HW22] Xing Huang and Feng-Yu Wang. "Singular McKean-Vlasov (reflecting) SDEs with distribution dependent noise". In: *Journal of Mathematical Analysis and Applications* 514.1 (2022), p. 126301.

[Hwa80] Chii-Ruey Hwang. "Laplace's method revisited: weak convergence of probability measures". In: *The Annals of Probability* (1980), pp. 1177–1182.

[Hwa81] Chii Ruey Hwang. "A generalization of Laplace's method". In: *Proceedings of the American Mathematical Society* 82.3 (1981), pp. 446–451.

[HY21] Xing Huang and Fen-Fen Yang. "Distribution-dependent SDEs with Hölder continuous drift and α-stable noise". en. In: *Numerical Algorithms* 86.2 (Feb. 2021), pp. 813–831.

[Ino91] Masaaki Inoue. "Derivation of a porous medium equation from many Markovian particles and the propagation of chaos". In: *Hiroshima Mathematical Journal* 21.1 (1991), pp. 85–110.

[IR23] Elena Issoglio and Francesco Russo. "McKean SDEs with singular coefficients". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 59.3 (Aug. 2023).

[Iss+24] Elena Issoglio et al. *Degenerate McKean-Vlasov equations with drift in anisotropic negative Besov spaces.* arXiv:2401.09165 [math]. Jan. 2024.

[Itô44] Kiyosi Itô. “Stochastic integral”. In: *Proceedings of the Japan Academy, Series A, Mathematical Sciences* 20.8 (Jan. 1944).

[Itô46] Kiyosi Itô. “On a stochastic integral equation”. In: *Proceedings of the Japan Academy, Series A, Mathematical Sciences* 22.2 (Jan. 1946).

[JIP23] Luis Mario Chaparro Jáquez, Elena Issoglio, and Jan Palczewski. *Convergence rate of numerical scheme for SDEs with a distributional drift in Besov space.* arXiv:2309.11396 [cs, math]. Sept. 2023.

[JKO98] Richard Jordan, David Kinderlehrer, and Felix Otto. “The variational formulation of the Fokker–Planck equation”. In: *SIAM journal on mathematical analysis* 29.1 (1998), pp. 1–17.

[JM08] Benjamin Jourdain and Florent Malrieu. “Propagation of chaos and Poincaré inequalities for a system of particles interacting through their cdf”. In: *The Annals of Applied Probability* 18.5 (Oct. 2008).

[JM24] Benjamin Jourdain and Stéphane Menozzi. “Convergence rate of the Euler–Maruyama scheme applied to diffusion processes with drift coefficient and additive noise”. In: *The Annals of Applied Probability* 34.1B (2024), pp. 1663–1697.

[JM98] Benjamin Jourdain and Sylvie Meleard. “Propagation of chaos and fluctuations for a moderate model with smooth initial data”. en. In: *Annales de l’Institut Henri Poincare (B) Probability and Statistics* 34.6 (1998), pp. 727–766.

[JMW05] Benjamin Jourdain, Sylvie Méléard, and Wojbor A. Woyczynski. “Probabilistic approximation and inviscid limits for one-dimensional fractional conservation laws”. In: *Bernoulli* 11.4 (Aug. 2005).

[Jou00a] B. Jourdain. “Probabilistic approximation for a porous medium equation”. en. In: *Stochastic Processes and their Applications* 89.1 (Sept. 2000), pp. 81–99.

[Jou00b] Benjamin Jourdain. “Diffusion processes associated with nonlinear evolution equations for signed measures”. In: *Methodology and computing in applied probability* 2 (2000), pp. 69–91.

[Jou02] Benjamin Jourdain. “Probabilistic characteristics method for a one-dimensional inviscid scalar conservation law”. In: *The Annals of Applied Probability* 12.1 (2002), pp. 334–360.

[Jou97] Benjamin Jourdain. “Diffusions with a nonlinear irregular drift coefficient and probabilistic interpretation of generalized Burgers’ equations”. In: *ESAIM: Probability and Statistics* 1 (1997), pp. 339–355.

[JP04] Robert Jarrow and Philip Protter. “A short history of stochastic integration and mathematical finance: the early years, 1880–1970”. English. In: *A Festschrift for Herman Rubin.* Beachwood, OH: IMS, Institute of Mathematical Statistics, 2004, pp. 75–91.

[JR13] Benjamin Jourdain and Julien Reygner. “Propagation of chaos for rank-based interacting diffusions and long time behaviour of a scalar quasilinear parabolic equation”. en. In: *Stochastic Partial Differential Equations: Analysis and Computations* 1.3 (Sept. 2013), pp. 455–506.

[JR15] Benjamin Jourdain and Julien Reygner. “Capital distribution and portfolio performance in the mean-field Atlas model”. In: *Annals of Finance* 11.2 (2015), pp. 151–198.

[Kac56] Mark Kac. “Foundations of kinetic theory”. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume III.* Berkeley: University of California Press, 1956, pp. 173–200.

[Kan42] Leonid V Kantorovich. "On translation of mass". In: *Doklady Akademii Nauk SSSR* 37 (1942), pp. 227–229.

[Kan48] Leonid V Kantorovich. "On a problem of Monge". In: *Uspekhi Matematicheskikh Nauk* 3 (1948), pp. 225–226.

[KMP22] Alexander Kalinin, Thilo Meyer-Brandis, and Frank Proske. *Stability, uniqueness and existence of solutions to McKean-Vlasov SDEs in arbitrary moments.* arXiv:2205.02176 [math]. May 2022.

[KMP24] Alexander Kalinin, Thilo Meyer-Brandis, and Frank Proske. *Stability, uniqueness and existence of solutions to McKean-Vlasov SDEs: a multidimensional Yamada-Watanabe approach.* arXiv:2107.07838 [math]. June 2024.

[KNS16] Hamed Karimi, Julie Nutini, and Mark Schmidt. "Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition". In: *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part I 16.* Springer. 2016, pp. 795–811.

[KP92] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations.* Berlin, Heidelberg: Springer Berlin Heidelberg, 1992.

[KR05] Nicolai V Krylov and Michael Röckner. "Strong solutions of stochastic equations with singular time dependent drift". In: *Probability theory and related fields* 131 (2005), pp. 154–196.

[Kry08] Nikolaj Vladimirovič Krylov. *Controlled diffusion processes.* Vol. 14. Springer Science & Business Media, 2008.

[Kuo06] H.H. Kuo. *Introduction to Stochastic Integration.* Universitext. Springer New York, 2006.

[KW67] Hiroshi Kunita and Shinzo Watanabe. "On square integrable martingales". In: *Nagoya Mathematical Journal* 30 (1967), pp. 209–245.

[Lac23] Daniel Lacker. "Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions". en. In: *Probability and Mathematical Physics* 4.2 (May 2023), pp. 377–432.

[Le24a] Anh-Dung Le. *Convergence rate of Euler scheme for McKean-Vlasov SDEs with density-dependent drift.* 2024. under review.

[Le24b] Anh-Dung Le. *Well-posedness of McKean-Vlasov SDEs with density-dependent drift.* 2024. under review.

[Léo86] Christian Léonard. "Une loi des grands nombres pour des systèmes de diffusions avec interaction et à coefficients non bornés". In: *Annales de l'IHP Probabilités et statistiques.* Vol. 22. 2. 1986, pp. 237–262.

[LL22] Khoa Lê and Chengcheng Ling. *Taming singular stochastic differential equations: A numerical method.* arXiv:2110.01343 [cs, math]. June 2022.

[LM16] Juan Li and Hui Min. "Weak Solutions of Mean-Field Stochastic Differential Equations and Application to Zero-Sum Stochastic Differential Games". en. In: *SIAM Journal on Control and Optimization* 54.3 (Jan. 2016), pp. 1826–1858.

[LM22] Zhenxin Liu and Jun Ma. "Existence, uniqueness and exponential ergodicity under Lyapunov conditions for McKean-Vlasov SDEs with Markovian switching". en. In: *Journal of Differential Equations* 337 (Nov. 2022), pp. 138–167.

[LM23] Zhenxin Liu and Jun Ma. *Existence, uniqueness and ergodicity for McKean-Vlasov SDEs under distribution-dependent Lyapunov conditions.* arXiv:2309.05411 [math]. Sept. 2023.

[LM24] Zhenxin Liu and Jun Ma. *Continuous dependence for McKean-Vlasov SDEs under distribution-dependent Lyapunov conditions.* arXiv:2406.00309 [math]. June 2024.

[LMS23] Linshan Liu, Mateusz B. Majka, and Łukasz Szpruch. "Polyak–Łojasiewicz inequality on the space of measures and convergence of mean-field birth-death processes". en. In: *Applied Mathematics & Optimization* 87.3 (June 2023), p. 48.

[LS88] O.A. Ladyženskaja and V.A. Solonnikov. *Linear and Quasi-linear Equations of Parabolic Type*. Translations of mathematical monographs. American Mathematical Society, 1988.

[LSZ20] Daniel Lacker, Mykhaylo Shkolnikov, and Jiacheng Zhang. "Inverting the Markovian projection, with an application to local stochastic volatility models". In: *The Annals of Probability* 48.5 (Sept. 2020).

[LSZ22] Daniel Lacker, Mykhaylo Shkolnikov, and Jiacheng Zhang. "Superposition and mimicking theorems for conditional McKean–Vlasov equations". In: *Journal of the European Mathematical Society* 25.8 (Aug. 2022), pp. 3229–3288.

[Luo18] Dejun Luo. "The Itô SDEs and Fokker-Planck equations with Osgood and Sobolev coefficients". In: *Stochastics* 90.3 (2018), pp. 379–410.

[LW14] Guangqiang Lan and Jiang-Lun Wu. "New sufficient conditions of existence, moment estimations and non confluence for SDEs with non-Lipschitzian coefficients". en. In: *Stochastic Processes and their Applications* 124.12 (Dec. 2014), pp. 4030–4049.

[Mag23] Francesco Maggi. *Optimal Mass Transport on Euclidean Spaces.* Cambridge University Press, 2023.

[Mar23] Gabriela Marinoschi. *Dual variational approach to nonlinear diffusion equations.* eng. Progress in nonlinear differential equations and their applications PNLDE subseries in control volume 102. Birkhäuser, 2023.

[McC01] Robert J McCann. "Polar factorization of maps on Riemannian manifolds". In: *Geometric & Functional Analysis GAFA* 11.3 (2001), pp. 589–608.

[McK66] Henry P McKean Jr. "A class of Markov processes associated with nonlinear parabolic equations". In: *Proceedings of the National Academy of Sciences* 56.6 (1966), pp. 1907–1911.

[McK67] Henry P McKean. "Propagation of chaos for a class of non-linear parabolic equations". In: *Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967)* (1967), pp. 41–57.

[Med07] Péter Medvegyev. *Stochastic integration theory.* Vol. 14. Oxford University Press, 2007.

[Mél96] Sylvie Méléard. "Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models". In: *Probabilistic Models for Nonlinear Partial Differential Equations: Lectures given at the 1st Session of the Centro Internazionale Matematico Estivo (C.I.M.E.) held in Montecatini Terme, Italy, May 22–30, 1995.* Ed. by Denis Talay and Luciano Tubaro. Berlin, Heidelberg: Springer Berlin Heidelberg, 1996, pp. 42–95.

[Met+53] Nicholas Metropolis et al. "Equation of state calculations by fast computing machines". In: *The journal of chemical physics* 21.6 (1953), pp. 1087–1092.

[Mey09] Paul-André Meyer. "Stochastic processes from 1950 to the present". In: *Electronic Journal for History of Probability and Statistics* 5.1 (2009), pp. 1–42.

[Mey62] Paul-André Meyer. "A decomposition theorem for supermartingales". In: *Illinois Journal of Mathematics* 6.2 (1962), pp. 193–205.

[Mey67a] Paul-André Meyer. “Intégrales stochastiques I”. In: *Séminaire de Probabilités de Strasbourg*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1967, pp. 72–87.

[Mey67b] Paul-André Meyer. “Intégrales stochastiques II”. In: *Séminaire de Probabilités de Strasbourg*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1967, pp. 95–117.

[Mey67c] Paul-André Meyer. “Intégrales stochastiques III”. In: *Séminaire de Probabilités de Strasbourg*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1967, pp. 118–141.

[Mey67d] Paul-André Meyer. “Intégrales stochastiques IV”. In: *Séminaire de Probabilités de Strasbourg*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1967, pp. 142–162.

[Mic92] Laurent Miclo. “Recuit simulé sur $\mathbb{R}^n$. Étude de l’évolution de l’énergie libre”. In: *Annales de l’IHP Probabilités et statistiques*. Vol. 28. 2. 1992, pp. 235–266.

[MMN18] Song Mei, Andrea Montanari, and Phan-Minh Nguyen. “A mean field view of the landscape of two-layer neural networks”. In: *Proceedings of the National Academy of Sciences* 115.33 (2018), E7665–E7671.

[Mon81] Gaspard Monge. “Mémoire sur la théorie des déblais et des remblais”. In: *Histoires de l’Académie royale des sciences* (1781), pp. 666–704.

[MP14] Michel Métivier and Jean Pellaumail. *Stochastic integration*. Academic Press, 2014.

[MPZ21] Stéphane Menozzi, Antonello Pesce, and Xicheng Zhang. “Density and gradient estimates for non degenerate Brownian SDEs with unbounded measurable drift”. In: *Journal of Differential Equations* 272 (2021), pp. 330–369.

[MR14] V. Mandrekar and B. Rüdiger. *Stochastic Integration in Banach Spaces: Theory and Applications*. Probability Theory and Stochastic Modelling. Springer International Publishing, 2014.

[MR87] Sylvie Meleard and Sylvie Roelly-Coppoletta. “A propagation of chaos result for a system of particles with moderate interaction”. en. In: *Stochastic Processes and their Applications* 26 (1987), pp. 317–332.

[MR88] Sylvie Méléard and Sylvie Roelly-Coppoletta. “Systèmes de particules et mesures-martingales: un théorème de propagation du chaos”. In: *Séminaire de probabilités de Strasbourg* 22 (1988), pp. 438–448.

[MS19] Sima Mehri and Wilhelm Stannat. “Weak solutions to Vlasov–McKean equations under Lyapunov-type conditions”. en. In: *Stochastics and Dynamics* 19.06 (Dec. 2019), p. 1950042.

[MT14] Adrian Muntean and Federico Toschi. *Collective dynamics from bacteria to crowds: an excursion through modeling, analysis and simulation*. Vol. 553. Springer, 2014.

[MV20] Yuliya Mishura and Alexander Veretennikov. “Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations”. In: *Theory of Probability and Mathematical Statistics* 103 (2020), pp. 59–101.

[NPT10] Giovanni Naldi, Lorenzo Pareschi, and Giuseppe Toscani. *Mathematical modeling of collective behavior in socio-economic and life sciences*. Springer Science & Business Media, 2010.

[NT85] Masao Nagasawa and Hiroshi Tanaka. “A diffusion process in a singular mean-drift-field”. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 68.3 (1985), pp. 247–269.

[Oel84] Karl Oelschläger. “A martingale approach to the law of large numbers for weakly interacting stochastic processes”. In: *The Annals of Probability* (1984), pp. 458–479.

[Oel85] Karl Oelschläger. “A law of large numbers for moderately interacting diffusion processes”. en. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 69.2 (Mar. 1985), pp. 279–322.

[OK85] Hirofumi Osada and Shinichi Kotani. “Propagation of chaos for the Burgers equation”. In: *Journal of the Mathematical Society of Japan* 37.2 (1985), pp. 275–294.

[Ott01] Felix Otto. “The Geometry of Dissipative Evolution Equations: The Porous Medium Equation”. In: *Communications in Partial Differential Equations* 26.1-2 (2001), pp. 101–174.

[Ott96] Felix Otto. “Double degenerate diffusion equations as steepest descent”. 1996.

[Pie82] Michel Pierre. “Uniqueness of the solutions of $u_t - \Delta\varphi(u) = 0$ with initial datum a measure”. In: *Nonlinear Analysis: Theory, Methods & Applications* 6.2 (1982), pp. 175–187.

[Pro13] P. Protter. *Stochastic Integration and Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2013.

[PRS21] Sahani Pathiraja, Sebastian Reich, and Wilhelm Stannat. “McKean-Vlasov SDEs in nonlinear filtering”. In: *SIAM Journal on Control and Optimization* 59.6 (2021), pp. 4188–4215.

[Reh23a] Marco Rehmeier. “Linearization and a superposition principle for deterministic and stochastic nonlinear Fokker-Planck-Kolmogorov equations”. In: *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze* (Sept. 2023), pp. 1705–1739.

[Reh23b] Marco Rehmeier. *Weighted L^1-semigroup approach for nonlinear Fokker–Planck equations and generalized Ornstein–Uhlenbeck processes*. arXiv:2308.09420 [math]. Aug. 2023.

[Ren23] Panpan Ren. “Singular McKean–Vlasov SDEs: Well-posedness, regularities and Wang’s Harnack inequality”. en. In: *Stochastic Processes and their Applications* 156 (Feb. 2023), pp. 291–311.

[RJM22] P.-E. Chaudru de Raynal, J.-F. Jabir, and S. Menozzi. *Multidimensional Stable driven McKean-Vlasov SDEs with distributional interaction kernel - a regularization by noise perspective*. arXiv:2205.11866 [math]. May 2022.

[RJM23] P.-E. Chaudru de Raynal, J.-F. Jabir, and S. Menozzi. *Multidimensional Stable Driven McKean-Vlasov SDEs with Distributional Interaction Kernel: Critical Thresholds and Related Models*. arXiv:2302.09900 [math]. Feb. 2023.

[RR23] Marco Rehmeier and Michael Röckner. *Nonlinear Fokker-Planck-Kolmogorov equations as gradient flows on the space of probability measures*. arXiv:2306.09530 [math]. June 2023.

[RRT17] Maxim Raginsky, Alexander Rakhlin, and Matus Telgarsky. “Non-convex learning via stochastic gradient langevin dynamics: a nonasymptotic analysis”. In: *Conference on Learning Theory*. PMLR. 2017, pp. 1674–1703.

[RRW22] Panpan Ren, Michael Röckner, and Feng-Yu Wang. “Linearization of nonlinear Fokker-Planck equations and applications”. en. In: *Journal of Differential Equations* 322 (June 2022), pp. 1–37.

[Rud87] Walter Rudin. *Real and complex analysis*. 3rd. New York: McGraw-Hill, 1987.

[RV22] Grant Rotskoff and Eric Vanden-Eijnden. “Trainability and accuracy of artificial neural networks: An interacting particle system approach”. In: *Communications on Pure and Applied Mathematics* 75.9 (2022), pp. 1889–1935.

[RV95] B. Roynette and P. Vallois. “Instability of Certain Nonlinear Stochastic Differential Equations”. en. In: *Journal of Functional Analysis* 130.2 (June 1995), pp. 477–523.

[RXZ20] Michael Röckner, Longjie Xie, and Xicheng Zhang. “Superposition principle for non-local Fokker–Planck–Kolmogorov operators”. en. In: *Probability Theory and Related Fields* 178.3-4 (Dec. 2020), pp. 699–733.

[RZ21] Michael Röckner and Xicheng Zhang. “Well-posedness of distribution dependent SDEs with singular drifts”. In: *Bernoulli* 27.2 (May 2021).

[San15] Filippo Santambrogio. “Optimal transport for applied mathematicians”. In: *Birkäuser, NY* 55.58-63 (2015), p. 94.

[Sch21] René L. Schilling. *Brownian Motion: A Guide to Random Processes and Stochastic Calculus*. De Gruyter, Sept. 2021.

[Sch86] Michael Scheutzow. “Periodic behavior of the stochastic Brusselator in the mean-field limit”. In: *Probability theory and related fields* 72.3 (1986), pp. 425–462.

[Sch87] Michael Scheutzow. “Uniqueness and non-uniqueness of solutions of Vlasov-McKean equations”. In: *Journal of the Australian Mathematical Society* 43.2 (1987), pp. 246–256.

[SH24] Ke Song and Zimo Hao. *Convergence rate of the Euler-Maruyama scheme to density dependent SDEs driven by α-stable additive noise*. arXiv:2405.20840 [math]. May 2024.

[Shk12] Mykhaylo Shkolnikov. “Large systems of diffusions interacting through their ranks”. en. In: *Stochastic Processes and their Applications* 122.4 (Apr. 2012), pp. 1730–1747.

[SO18] Marielle Simon and Christian Olivera. “Non-local Conservation Law from Stochastic Particle Systems”. en. In: *Journal of Dynamics and Differential Equations* 30.4 (Dec. 2018), pp. 1661–1682.

[Sok12] Alexander Sokol. *An introduction to stochastic integration with respect to continuous semimartingales*. 2012.

[Sok13] Alexander Sokol. *An introduction to stochastic integration with respect to general semimartingales*. 2013.

[Son+21] Yang Song et al. “Score-Based Generative Modeling through Stochastic Differential Equations”. In: *International Conference on Learning Representations*. 2021.

[SS20a] Justin Sirignano and Konstantinos Spiliopoulos. “Mean field analysis of neural networks: A central limit theorem”. en. In: *Stochastic Processes and their Applications* 130.3 (Mar. 2020), pp. 1820–1852.

[SS20b] Justin Sirignano and Konstantinos Spiliopoulos. “Mean field analysis of neural networks: A law of large numbers”. In: *SIAM Journal on Applied Mathematics* 80.2 (2020), pp. 725–752.

[Szn86] A. S. Sznitman. “A propagation of chaos result for Burgers’ equation”. en. In: *Probability Theory and Related Fields* 71.4 (Dec. 1986), pp. 581–613.

[Szn91] Alain-Sol Sznitman. “Topics in propagation of chaos”. In: *Ecole d’été de probabilités de Saint-Flour XIX—1989* 1464 (1991), pp. 165–251.

[Tan84] Hiroshi Tanaka. “Limit theorems for certain diffusion processes with interaction”. In: *North-Holland Mathematical Library*. Vol. 32. Elsevier, 1984, pp. 469–488.

[Tao11] Terence Tao. *An introduction to measure theory*. Vol. 126. American Mathematical Society, 2011.

[TH81] Hiroshi Tanaka and Masuyuki Hitsuda. “Central limit theorem for a simple diffusion model of interacting particles”. In: *Hiroshima Mathematical Journal* 11.2 (1981), pp. 415–423.

[Tho23] Matthew Thorpe. *Introduction to optimal transport*. 2023.

[Tre16] Dario Trevisan. “Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients”. In: *Electronic Journal of Probability* 21.none (Jan. 2016).

[Uch00] Kôhei Uchiyama. "Pressure in Classical Statistical Mechanics and Interacting Brownian Particles in Multi-dimensions:" en. In: *Annales Henri Poincaré* 1.6 (Dec. 2000), pp. 1159–1202.

[Van+87] Peter JM Van Laarhoven et al. *Simulated annealing*. Springer, 1987.

[Var91] S. R. S. Varadhan. "Scaling limits for interacting diffusions". en. In: *Communications in Mathematical Physics* 135.2 (Jan. 1991), pp. 313–353.

[Vaz06] Juan Luis Vazquez. *The Porous Medium Equation: Mathematical Theory*. Oxford University Press, Oct. 2006.

[Vil03] Cédric Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003.

[Vil09] Cédric Villani. *Optimal transport: old and new*. Vol. 338. Springer, 2009.

[VK76] A Ju Veretennikov and Nikolai Vladimirovich Krylov. "On explicit formulas for solutions of stochastic equations". In: *Mathematics of the USSR-Sbornik* 29.2 (1976), p. 239.

[Wan18] Feng-Yu Wang. "Distribution dependent SDEs for Landau type equations". en. In: *Stochastic Processes and their Applications* 128.2 (Feb. 2018), pp. 595–621.

[Wan23a] Feng-Yu Wang. "Derivative Formula for Singular McKean-Vlasov SDEs". In: *Communications on Pure and Applied Analysis* 22.6 (2023), pp. 1866–1898.

[Wan23b] Feng-Yu Wang. "Distribution dependent reflecting stochastic differential equations". en. In: *Science China Mathematics* 66.11 (Nov. 2023), pp. 2411–2456.

[Wan23c] Feng-Yu Wang. "Singular density dependent stochastic differential equations". In: *Journal of Differential Equations* 361 (2023), pp. 562–589.

[WH23] Mingyan Wu and Zimo Hao. "Well-posedness of density dependent SDE driven by α-stable process with Hölder drifts". en. In: *Stochastic Processes and their Applications* 164 (Oct. 2023), pp. 416–442.

[WR24] Feng-Yu Wang and Panpan Ren. *Distribution dependent stochastic differential equations*. en. Vol. 05. World Scientific Series on Probability Theory and Its Applications. World Scientific, Nov. 2024.

[Xia+20] Pengcheng Xia et al. "Lq (Lp)-theory of stochastic differential equations". In: *Stochastic Processes and their Applications* 130.8 (2020), pp. 5188–5211.

[Xu+18] Pan Xu et al. "Global convergence of Langevin dynamics based algorithms for nonconvex optimization". In: *Advances in Neural Information Processing Systems* 31 (2018).

[XZ20] Longjie Xie and Xicheng Zhang. "Ergodicity of stochastic differential equations with jumps and singular coefficients". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 56.1 (2020), pp. 175–229.

[Zha22] Xicheng Zhang. *Second order McKean-Vlasov SDEs and kinetic Fokker-Planck-Kolmogorov equations*. arXiv:2109.01273 [math]. Jan. 2022.

[Zha24] Guohuan Zhao. "Existence and Uniqueness for McKean-Vlasov Equations with Singular Interactions". en. In: *Potential Analysis* (June 2024).

[ZY21] Shao-Qin Zhang and Chenggui Yuan. "A Zvonkin's transformation for stochastic differential equations with singular drift and applications". In: *Journal of Differential Equations* 297 (2021), pp. 277–319.