Linear and Nonlinear Fokker–Planck equations: Analysis & Probabilistic Counterparts

Marco Rehmeier

TU Berlin

Winter 2024/2025

last update: 12.02.2025

Contents

1 Linear Fokker–Planck equations

1.1 Probabilistic basics and motivation

Set $\mathbb{R}_+ := [0,\infty)$, $\mathbb{N} := \{1,2,3,\dots\}$ and $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$. The distribution of a random map X is denoted by $\mathcal{L}_X$.

We begin by repeating the definition of solutions to stochastic differential equations on $\mathbb{R}^d$

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dB_t, \tag{SDE}$$

where the drift- and diffusion-coefficients

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

are assumed to be product-measurable w.r.t. the usual Borel σ-algebras.

Definition 1.1.1. (i) A *(probabilistically weak) solution* to (SDE) is a triple consisting of a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$, a d-dimensional standard $(\mathcal{F}_t)$-Brownian motion B and an $(\mathcal{F}_t)$-adapted $\mathbb{R}^d$-valued stochastic process $X = (X_t)_{t\geqslant 0}$ on Ω such that

$$\mathbb{E}\Big[\int_0^T |b(t,X_t)| + |\sigma(t,X_t)|^2\, dt\Big] < \infty, \quad \forall T > 0$$

and $\mathbb{P}$-a.s.

$$X_t = X_0 + \int_0^t b(s,X_s)ds + \int_0^t \sigma(s,X_s)dB_s, \quad \forall t > 0.$$

(ii) If $\mathcal{L}_{X_0} = \mu$, the weak solution has *initial value* μ.

We often simply say "X is the weak solution". Note that the definition implies that the paths $t \mapsto X_t(\omega)$ are continuous $\mathbb{P}$-a.s.

To consider an initial time $s > 0$, replace 0 in the above definition by s. One then says X has initial condition (s,μ).

The *path law*, or simply *law*, of a stochastic process X with continuous paths on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is its distribution on $C_+\mathbb{R}^d := C(\mathbb{R}_+, \mathbb{R}^d)$, i.e. the image measure (on path space) $\mathcal{L}_X = \mathbb{P} \circ X^{-1}$ of $X : \Omega \to C_+\mathbb{R}^d$.

Remark 1.1.2. *(i) More generally, for any $m \in \mathbb{N}$, one may consider σ with values in $\mathbb{R}^{d\times m}$ and m-dimensional Brownian motions. We will, however, restrict to the case $\sigma \in \mathbb{R}^{d\times d}$.*

(ii) *The law of a weak solution solves the martingale problem associated with* b *and* $\frac{1}{2}\sigma\sigma^T$ *and, vice versa, for every solution* P *of the latter, there is a weak solution to* (SDE) *with law* P*. Thus, we will often identify weak solutions, their laws and solutions to the associated martingale problem.*[1]

From a probabilistic point of view, the following proposition is one main motivation to study Fokker–Planck equations. Set $a = (a_{ij})_{i,j\leqslant d}$, $a_{ij} := \frac{1}{2}(\sigma\sigma^T)_{ij}$. The matrix $a(t,x)$ is symmetric and nonnegative definite for all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$.

Proposition 1.1.3. *Let* X *be a weak solution to* (SDE)*. Then its probability measure-valued weakly continuous curve of* one-dimensional time marginals

$$t \mapsto \mathcal{L}_{X_t} =: \mu_t, \quad t \geqslant 0,$$

satisfies (using Einstein summation convention and denoting by ∂_i *and* ∂^2_{ij} *first and second order spatial partial derivatives)*

$$\int_{\mathbb{R}^d} \varphi(x)\, d\mu_t(x) = \int_{\mathbb{R}^d} \varphi(x)\, d\mu_0(x) + \int_0^t \int_{\mathbb{R}^d} a_{ij}(s,x)\partial^2_{ij}\varphi(x) + b_i(s,x)\partial_i\varphi(x)\, d\mu_s(x) ds$$

for all $t \geqslant 0$ *and* $\varphi \in C_c^\infty(\mathbb{R}^d)$ *(the latter denotes the space of smooth real-valued functions on* $\mathbb{R}^d$ *with compact support).*

Proof. Exercise 1.1. □

The distributional formulation of the previous equality is

$$\partial_t\mu_t = \partial^2_{ij}(a_{ij}\mu_t) - \partial_i(b_i\mu_t),$$

which, as we shall see, is a Fokker–Planck equation for Borel (probability) measures on $\mathbb{R}^d$. If

$$\mu_t = \varrho_t(x)dx$$

and $\varrho : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ and a_{ij}, b_i are sufficiently regular, then

$$\partial_t\varrho_t = \partial^2_{ij}(a_{ij}\varrho_t) - \partial_i(b_i\varrho_t)$$

holds pointwise, i.e. in the classical, strong sense.

Hence: Marginals of SDE-solution solve a deterministic PDE for measures!

Spaces of measures, vague and weak topology For a topological space Y, $\mathcal{M}_b^+(Y)$ denotes the set of nonnegative finite Borel measures on Y. We write $\mathcal{M}_b^+ := \mathcal{M}_b^+(\mathbb{R}^d)$ when no confusion about the dimension d can occur. Denote by $C_c(Y)$ and $C_b(Y)$ the spaces of continuous functions $g : Y \to \mathbb{R}$ which are compactly supported and bounded, respectively. Let now Y be a metric space.

[1] We briefly review the definition and basic theory of martingale problems later on.

Definition 1.1.4. (i) The *vague*, respectively *weak topology* on $\mathcal{M}_b^+(Y)$ is the initial topology of the maps $\mu \mapsto \int_Y f\,d\mu$ for all $f \in C_c(Y)$, respectively $f \in C_b(Y)$, i.e. the coarsest topology τ on $\mathcal{M}_b^+(Y)$ such that each of these maps is continuous between $(\mathcal{M}_b^+(Y), \tau)$ and $\mathbb{R}$.

(ii) $(\mu_n)_{n\in\mathbb{N}}$ *converges vaguely (weakly) to* μ in $\mathcal{M}_b^+(Y)$, if it converges in the vague (weak) topology, i.e. if $\int_Y f\,d\mu_n \xrightarrow{n\to\infty} \int_Y f\,d\mu$ for all $f \in C_c(Y)$ ($f \in C_b(Y)$).

Remark 1.1.5. *(i)* $\mu_n \xrightarrow{n\to\infty} \mu$ *weakly* $\implies$ $\mu_n(Y) \xrightarrow{n\to\infty} \mu(Y)$.

(ii) Wrong for vague *convergence:* $Y = \mathbb{R}$, $\mu_n = \delta_n, \mu = 0$ *(the trivial measure), then* $\mu_n \xrightarrow{n\to\infty} \mu$ *vaguely,* $\mu_n(\mathbb{R}) = 1$ *for all* $n \in \mathbb{N}$ *and* $\mu(\mathbb{R}) = 0$.

(iii) Let $Y = \mathbb{R}^d$*. The set of subprobability measures*

$$\mathcal{SP} := \mathcal{M}_b^+ \cap \{\mu : \mu(\mathbb{R}^d) \leqslant 1\}$$

is the positive hemisphere of the unit ball in $\overline{C_c(\mathbb{R}^d)}'$ *(the closure of* $C_c(\mathbb{R}^d)$ *w.r.t. the topology of uniform convergence), which is weak-∗-sequentially compact. In particular:* ***Every sequence of subprobability measures has a vaguely convergent subsequence****. This is not true when "weakly" replaces "vaguely".*

(iv) $\mathcal{M}_b^+$ *and* $\mathcal{P}$ *(the set of Borel probability measures on* $\mathbb{R}^d$*) with the weak topology and* $\mathcal{SP}$ *with the vague topology are Polish spaces.*

1.2 Definition, existence, uniqueness

Let $d \in \mathbb{N}$ and consider Borel coefficients

$$b = (b_i)_{i\leqslant d}, a = (a_{ij})_{i,j\leqslant d}, c, \quad b_i, a_{ij}, c : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}, \quad i,j \leqslant d.$$

We always assume that $a(t,x)$ is symmetric nonnegative definite for all (t,x). The class of linear Fokker–Planck equation (FPE) we intend to study is

$$\partial_t \mu = \partial_{ij}^2\big(a_{ij}(t,x)\mu\big) - \partial_i\big(b_i(t,x)\mu\big) + c(t,x)\mu, \tag{FPE}$$

often with $c = 0$. These are linear equations, since the coefficients do not depend on the solution. Setting

$$L_{a,b,c} : \varphi \mapsto L_{a,b,c}\varphi(t,x) := a_{ij}(t,x)\partial_{ij}^2\varphi(x) + b_i(t,x)\partial_i\varphi(x) + c(t,x)\varphi(x),$$

a shorter way of writing (FPE) is

$$\partial_t \mu = L_{a,b,c}^*\mu,$$

where L^* denotes the formal dual of L.

Definition 1.2.1. A locally finite Borel measure μ on $(0,\infty)\times\mathbb{R}^d$ satisfies (FPE) if $a_{ij}, b_i, c \in L^1_{\text{loc}}\big((0,\infty)\times\mathbb{R}^d;\mu\big)$ and for every $\varphi \in C_c^\infty((0,\infty)\times\mathbb{R}^d)$ we have

$$\int_{(0,\infty)\times\mathbb{R}^d} \partial_t\varphi + L_{a,b,c}\varphi \, d\mu = 0.$$

We always restrict to nonnegative measures μ given by a family of nonnegative locally finite Borel measures $(\mu_t)_{t>0}$ on $\mathbb{R}^d$ via $\mu = \mu_t dt$, i.e.

$$\int_{(0,\infty)\times\mathbb{R}^d} f(t,x)\, d\mu(t,x) = \int_0^\infty \Big(\int_{\mathbb{R}^d} f(t,x)\, d\mu_t(x)\Big) dt, \tag{1.2.1}$$

and we usually say $(\mu_t)_{t>0}$ is the solution. In order for the integral on the right hand-side to make sense, the measures $(\mu_t)_{t>0}$ need to be a *Borel curve*, i.e. $t\mapsto\mu_t(A)$ has to be Borel for every $A\in\mathcal{B}(\mathbb{R}^d)$.

Remark 1.2.2. *Note that not every locally finite nonnegative Borel measure μ on $(0,\infty)\times\mathbb{R}^d$ can be expressed as in (1.2.1). Indeed, by the disintegration theorem, one may always represent*

$$\int_{(0,\infty)\times\mathbb{R}^d} f(t,x)\, d\mu(t,x) = \int_0^\infty \Big(\int_{\mathbb{R}^d} f(t,x)\, d\mu_t(x)\Big) d\eta_1(t), \tag{1.2.2}$$

where $\eta_1 = \mu\circ(ev_1)^{-1}$ ($ev_1 : (0,\infty)\times\mathbb{R}^d \to (0,\infty)$, $ev_1(t,x) = t$ denotes the projection on the first component). Clearly, this can be rewritten as in (1.2.1) if and only if $\eta_1 \ll dt$.

We call μ a *(sub-)probability solution*, respectively a *solution with constant mass*, if every μ_t is a (sub-)probability measure or if $\mu_t(\mathbb{R}^d) = \mu_s(\mathbb{R}^d)$ for all $t,s>0$, respectively. Depending on context, these conditions may also be understood for dt-almost surely.

Remark 1.2.3. *If $(\mu_t)_{t>0}$ solves (FPE) and $(\tilde\mu_t)_{t>0}$ is a Borel curve of locally finite Borel measures such that $\mu_t = \tilde\mu_t$ for dt-a.a. $t>0$, then $(\tilde\mu_t)_{t>0}$ also satisfies (FPE). Hence solutions are only determined dt-a.s., and a natural question is whether the dt-equivalence class of a solution contains a vaguely or weakly continuous representative. As we shall see, this is true under very broad assumptions.*

Definition 1.2.4. A solution $(\mu_t)_{t>0}$ to (FPE) *has initial value* $\nu\in\mathcal{M}_b^+$, if for every $\varphi\in C_c^\infty(\mathbb{R}^d)$ there is a set of full dt-measure $O_\varphi\subseteq(0,\infty)$ such that

$$\int_{\mathbb{R}^d}\varphi\, d\nu = \lim_{t\to0, t\in O_\varphi}\int_{\mathbb{R}^d}\varphi\, d\mu_t. \tag{1.2.3}$$

In this case, one sets $\mu_0 := \nu$ and considers $(\mu_t)_{t\geqslant0}$ instead of $(\mu_t)_{t>0}$. The pair (FPE)+(1.2.3) is the *Cauchy problem* associated with the FPE.

Clearly, the initial value is unique (Exercise 1.2). Equation (1.2.3) does not imply vague convergence $\mu_t \xrightarrow{t\to0} \nu$, but it does, if $O_\varphi^c = \emptyset$ for all φ.

For the case $\mu = (\mu_t)_{t>0}$, the following equivalent definition of solution to (FPE) is very useful.

Definition 1.2.5. A Borel curve of locally finite Borel measures $(\mu_t)_{t>0}$ *is a solution to* (FPE) *with initial value* ν, if $a_{ij}, b_i, c \in L^1_{\text{loc}}\big((0,\infty)\times\mathbb{R}^d;\mu_t dt\big)$ and for every $\varphi \in C_c^\infty(\mathbb{R}^d)$ there is a set of full dt-measure $J_\varphi \subseteq (0,\infty)$ such that for all $t \in J_\varphi$

$$\int_{\mathbb{R}^d} \varphi\, d\mu_t = \int_{\mathbb{R}^d} \varphi\, d\nu + \lim_{\tau\to 0+} \int_\tau^t \int_{\mathbb{R}^d} L_{a,b,c}\varphi\, d\mu_s ds. \tag{1.2.4}$$

Lemma 1.2.6. *(i) If $a_{ij}, b_i, c \in L^1_{\text{loc}}\big([0,\infty)\times\mathbb{R}^d;\mu_t dt\big)$, then*

$$\lim_{\tau\to 0+} \int_\tau^t \int_{\mathbb{R}^d} L_{a,b,c}\varphi\, d\mu_s ds = \int_0^t \int_{\mathbb{R}^d} L_{a,b,c}\varphi\, d\mu_s ds. \tag{1.2.5}$$

In this case: $J_\varphi^c = \emptyset$ for all φ if and only if $t \mapsto \mu_t$ is vaguely continuous on $[0,\infty)$.

(ii) If $c = 0$, $t \mapsto \mu_t$ is vaguely continuous and the first assumption in (i) is strengthened to $a_{ij}, b_i \in L^1\big([0,T]\times\mathbb{R}^d;\mu_t dt\big)$ for all $T > 0$, then $(\mu_t)_{t\geqslant 0}$ has constant mass. Moreover, in this case (1.2.5) holds for all $\varphi \in C_b^2(\mathbb{R}^d)$, the space of real-valued bounded continuous functions on $\mathbb{R}^d$ with uniformly bounded first- and second-order derivatives.

Proof. (i) The first assertion holds, since the compact support of φ implies $[t \mapsto \int_{\mathbb{R}^d} L_{a,b,c}\varphi\, d\mu_t] \in L^1_{\text{loc}}([0,\infty);dt)$, which yields the claim. The second assertion follows from the continuity of the map $t \mapsto \int_0^t f(t,x)\, d\mu_t(x)dt$ for every f such that $[t \mapsto \int f(t,\cdot)d\mu_t] \in L^1_{\text{loc}}\big([0,\infty);dt\big)$.

(ii) The first part is Exercise 1.3, the second part follows by a standard approximation.

□

The proof of the following result can be found on p.243 in [9].

Proposition 1.2.7. *μ given by (1.2.1) satisfies (FPE) with initial value ν in the sense of Definition 1.2.1 and 1.2.4 if and only if $(\mu_t)_{t>0}$ satisfies Definition 1.2.5.*

We may now reformulate Proposition 1.1.3 by saying that the one-dimensional time marginals $\mu_t := \mathcal{L}_{X_t}$ of a weak solution X to SDE are a weakly continuous probability solution to the Fokker–Planck equation FPE, with $c = 0$, $a = \frac{1}{2}\sigma\sigma^T$, b, and with initial value $\mathcal{L}_{X_0}$ (which may be prescribed on the level of the SDE). Definition 1.1.1 entails

$$\int_0^T \int_{\mathbb{R}^d} |a_{ij}(t,x)| + |b_i(t,x)| d\mu_t(x)dt = \mathbb{E}\Big[\int_0^T |a_{ij}(t,X_t)| + |b_i(t,X_t)|dt\Big] < \infty$$

for all $T > 0$ and $i,j \leqslant d$, i.e. all assertions of Lemma 1.2.6 hold.

This relation between SDEs and FPEs is one main reason why we will mostly be interested in the case $c = 0$ and in weakly continuous probability solutions in the general sense of the previous definitions.

Remark 1.2.8. *Several generalizations of and related equations to* (FPE) *have been studied in the literature, for instance equations for measures on more general state spaces, e.g. on open subsets $U \subseteq \mathbb{R}^d$, infinite-dimensional spaces and manifolds. A related class of equations are* elliptic FPEs

$$L^*_{a,b,c}\eta = 0.$$

Depending on time, we might briefly touch these aspects during the course of the lecture. Moreover, we will study nonlinear Fokker–Planck equations*. The term* nonlinear *refers to coefficients depending on the solution μ.*

Equation (FPE), Definitions 1.2.1,1.2.4,1.2.5 and all previous assertions can obviously be generalized to an initial time $s \geqslant 0$. In this case, the initial condition is the pair (s, ν). Also, it is obvious how to modify the definitions and previous results to the time interval $(0, T)$ instead of $(0, \infty)$ for some $T > 0$.

1.2.1 An existence result

There are many results on the existence of solutions to the Cauchy problem (FPE)+(1.2.3). Here, we present one result (Proposition 1.2.11 below) whose proof proceeds via standard arguments for the construction of solutions to PDEs with irregular coefficients: First, the coefficients and initial datum are approximated by more regular ones, for which existence of solutions is known or easy to obtain. Then, one proves uniform estimates of the corresponding solutions in order to extract a converging subsequence. Finally, one shows that its limit solves the original equation. We restrict to a finite time interval $[0, T]$, i.e. we consider Borel coefficients $a_{ij}, b_i, c : [0, T] \times \mathbb{R}^d \to \mathbb{R}$. The case $T = \infty$ can be obtained by a simple variation. We need the following two basic results. For their proofs, see [9, Ch.6.3, 6.6].

Lemma 1.2.9. *Assume there are numbers $0 < m < M$ such that $m\,\mathrm{Id} \leqslant a(t, x) \leqslant M\,\mathrm{Id}$ for all $(t, x) \in [0, T] \times \mathbb{R}^d$. Moreover, let a_{ij}, its first- and second-order derivatives, b_i and its first-order derivatives, and c be bounded and continuous on $(0, T) \times \mathbb{R}^d$ and Hölder continuous in x uniformly in t of some degree $\alpha \in (0, 1)$. Finally, suppose for some $C > 0$*

$$|a_{ij}(t, x) - a_{ij}(s, y)| \leqslant C(|x - y|^\alpha + |t - s|^{\frac{\alpha}{2}}), \quad \forall (t, x) \in (0, T) \times \mathbb{R}^d.$$

Then for every probability density $\varrho_0 \in C_b(\mathbb{R}^d)$ there is a subprobability solution $(\mu_t)_{t\in[0,T)}$ to (FPE) *with initial datum $\nu = \varrho_0 dx$ such that $\mu_t = \varrho_t dx$,*

$$[(t, x) \mapsto \varrho_t(x)] \in C^{1,2}((0, T) \times \mathbb{R}^d) \cap C([0, T) \times \mathbb{R}^d),$$

and for dt-a.e. $t \in (0, T)$

$$\mu_t(\mathbb{R}^d) \leqslant \nu(\mathbb{R}^d) + \int_0^t \int_{\mathbb{R}^d} c\, d\mu_s ds.$$

For the next two results, we denote by $U \subseteq \mathbb{R}^d$ an arbitrary open Euclidean ball and by J an arbitrary set of type $[t_0, T - t_0], t_0 > 0$.

Lemma 1.2.10. *Let $\mu = (\mu_t)_{t\in(0,T)}$ be a solution to (FPE), and assume on every $J\times U$ a is bounded, Hölder-continuous in x uniformly in t and there is $m(J,U) > 0$ such that $m(J,U)\,\mathrm{Id} \leqslant a(t,x)$ for all $(t,x) \in J\times U$.*

Then $\mu_t = \varrho_t(x)dx$, $\varrho \in L^{\frac{d+3}{d+2}}_{\mathrm{loc}}((0,T)\times\mathbb{R}^d)$, and for every $J\times U$ and every neighborhood W of $J\times U$ with compact closure in $(0,T)\times\mathbb{R}^d$ one has

$$|\varrho|_{L^{\frac{d+3}{d+2}}(J\times U)} \leqslant \bar{C},$$

where $\bar{C}$ depends on $d, \inf_W \det a, |a|_{L^\infty(W)}, |b|_{L^1(W;\mu)}, |c|_{L^1(W;\mu)}, J, U, W$ and $\mu_t dt(W)$.

The main result of this section is the following proposition.

Proposition 1.2.11. *Suppose $c \leqslant 0$, and that a_{ij}, b_i and c are bounded on each $[0,T]\times U$. Assume for every U there are numbers $0 < m(U) < M(U)$ such that*

$$m(U)\,\mathrm{Id} \leqslant a(t,x) \leqslant M(U)\,\mathrm{Id}, \quad \forall (t,x) \in [0,T]\times U.$$

Then, for every $\nu \in \mathcal{P}$, there is a subprobability solution $\mu = (\mu_t)_{t\in[0,T)}$ to (FPE) with initial datum ν such that $c \in L^1((0,T)\times\mathbb{R}^d; \mu_t dt)$ and for dt-a.e. $t \in (0,T)$

$$\mu_t(\mathbb{R}^d) \leqslant \nu(\mathbb{R}^d) + \int_0^t \int_{\mathbb{R}^d} c\, d\mu_t dt. \tag{1.2.6}$$

The result can be extended to the case $c \leqslant c_0$ for $c_0 > 0$.

Proof. We divide the proof into five steps.

1. Define $a_{ij}(t,x) = \delta_{ij}, b_i(t,x) = 0 = c(t,x)$ for $(t,x) \in [0,T]^c \times \mathbb{R}^d$, let $\omega : \mathbb{R}^{d+1} \to \mathbb{R}$ satisfy

$$\omega \in C_c^\infty(\mathbb{R}^{d+1}), \quad \omega \geqslant 0, \quad \int_{\mathbb{R}^{d+1}} \omega(t,x)\, dxdt = 1, \quad \omega(t,x) = 0 \text{ for } |x| > 1,$$

set $\omega_\varepsilon(t,x) := \varepsilon^{-d-1}\omega(x\varepsilon^{-1}, t\varepsilon^{-1})$ for $\varepsilon > 0$, and, for $n \in \mathbb{N}$,

$$a^n_{ij} := a_{ij} * \omega_{\frac{1}{n}} + n^{-1}\delta_{ij}, \quad b^n_i := b_i * \omega_{\frac{1}{n}}, \quad c^n := c * \omega_{\frac{1}{n}}.$$

For each n and l, the functions a^n_{ij}, b^n_i, c^n and its derivatives up to order l are uniformly bounded on $\mathbb{R}^{d+1}$. Moreover, $a^n(t,x) \geqslant n^{-1}\,\mathrm{Id}$ for all $(t,x) \in \mathbb{R}^{d+1}$. Each of the sequences $(a^n_{ij})_{n\in\mathbb{N}}, (b^n_i)_{n\in\mathbb{N}}, (c^n)_{n\in\mathbb{N}}$ converges in $L^p([0,T]\times U_k)$ for each $p \geqslant 1$ and $k \in \mathbb{N}$, where U_k denotes the ball centered at the origin of radius k, with limits a_{ij}, b_i and c, respectively.

Let $\nu_n = \eta_n dx$, $\eta_n \in C_c^\infty(\mathbb{R}^d)$, be a sequence of probability measures converging weakly to ν, and consider the Cauchy problems

$$\partial_t \mu^n = \partial^2_{ij}\big(a^n_{ij}\mu^n\big) - \partial_i\big(b^n_i\mu^n\big) + c^n\mu^n, \quad \mu^n_{|t=0} = \eta_n. \tag{1.2.7}$$

By Lemma 1.2.9, for each $n \in \mathbb{N}$, there is a subprobability solution $(\mu_t^n)_{t\in[0,T)}$ to (1.2.7) such that $[(t,x) \mapsto \varrho_t^n(x)] \in C^{1,2}((0,T)\times\mathbb{R}^d) \cap C([0,T)\times\mathbb{R}^d)$, and for dt-a.e. $t \in (0,T)$

$$\mu_t^n(\mathbb{R}^d) \leqslant \nu^n(\mathbb{R}^d) + \int_0^t \int_{\mathbb{R}^d} c^n \, d\mu_s^n ds. \tag{1.2.8}$$

In particular, (1.2.5) holds and (1.2.4) is satisfied for every $t \in (0,T)$, since $t \mapsto \varrho_t(x)dx$ is vaguely continuous.

2. By definition of a^n, we have, independently of n,

$$a^n(t,x) \geqslant m_{k+1}\,\mathrm{Id}, \quad \forall (t,x) \in [0,T]\times U_k,$$

where $m_{k+1} = m(U_{k+1})$ is the number from the hypotheses of the proposition corresponding to the ball U_{k+1}. In addition, for any k and n we have (the L^∞-spaces in the next lines are understood with respect to the measure $dtdx$)

$$|a_{ij}^n|_{L^\infty([0,T]\times U_k)} \leqslant |a|_{L^\infty([0,T]\times U_{k+1})} + 1,$$
$$|b_i^n|_{L^\infty([0,T]\times U_k)} \leqslant |b_i|_{L^\infty([0,T]\times U_{k+1})}, \quad |c^n|_{L^\infty([0,T]\times U_k)} \leqslant |c|_{L^\infty([0,T]\times U_{k+1})}.$$

Lemma 1.2.10 implies for every $k > 2$

$$\int_{[Tk^{-1},T(1-k^{-1})]\times U_{k-1}} \varrho_t^n(x)^{\frac{d+3}{d+2}} dxdt \leqslant C_k, \tag{1.2.9}$$

where C_k depends on m_{k+1} and on the right hand-sides of the previous three estimates, but not on n. Since $L^{\frac{d+3}{d+2}}\big([Tk^{-1},T(1-k^{-1})]\times U_{k-1}\big)$ is reflexive, the sequence $(\varrho^n)_{n\in\mathbb{N}}$ contains, for every $k > 2$, a weakly converging subsequence in that space. By a standard diagonal argument, we may consider a subsequence, still denoted $(\varrho^n)_{n\in\mathbb{N}}$, which converges weakly in $L^{\frac{d+3}{d+2}}\big([Tk^{-1},T(1-k^{-1})]\times U_{k-1}\big)$ for *every* $k > 2$, to a limit ϱ (which does not depend on k). Set, for $t \in (0,T)$,

$$\mu_t := \varrho(t,x)dx$$

(note that this defines $(\mu_t)_{t\in(0,T)}$ up to a set of dt-measure zero).

3. Let $\varphi \in C_c^\infty(\mathbb{R}^d)$. There is $C(\varphi) \geqslant 0$, independent of n, such that for all $0 \leqslant s \leqslant t < T$

$$\left| \int_{\mathbb{R}^d} \varphi \, d\mu_t^n - \int_{\mathbb{R}^d} \varphi \, d\mu_s^n \right| = \left| \int_s^t \int_{\mathbb{R}^d} L_{a^n,b^n,c^n}\varphi \, d\mu_r^n dr \right| \leqslant C(\varphi)|t-s|. \tag{1.2.10}$$

Consequently, for fixed φ, the functions

$$[0,T) \ni t \mapsto \int_{\mathbb{R}^d} \varphi \, d\mu_t^n =: f^n(t), \quad n \in \mathbb{N},$$

are uniformly bounded and equicontinuous, hence the Arzelá-Ascoli theorem implies that every subsequence of $(f^n)_{n\in\mathbb{N}}$ contains a locally uniformly on $[0,T)$ converging subsubsequence. Since f^n converges $L^{\frac{d+3}{d+2}}([Tk^{-1},T(1-k^{-1})];dt)$-weakly to

$$f(t) := \int_{\mathbb{R}^d} \varphi \, d\mu_t$$

for all $k > 2$ and since uniform and weak limits coincide, it follows that any two subsubsequence limits coincide dt-a.s., hence pointwise (since they are continuous) on $(0,T)$. Consequently, $(f^n)_{n\in\mathbb{N}}$ converges locally uniformly on $(0,T)$ to a limit equal to f dt-a.s. For $t=0$, the definition of ν_n entails $\mu_0^n \to \nu$ weakly in the sense of measures, i.e. $f^n(0)\to f(0)$. The dt-exceptional set depends on φ and is denoted by $\mathbb{T}(\varphi)$.

4. We are now going to prove that $(\mu_t)_{t\in[0,T)}$, $\mu_t = \varrho(t,x)dx$ for $t>0$ and $\mu_0=\nu$, is a solution as in the assertion. For $\varphi \in C_c^\infty(\mathbb{R}^d)$, we have, for $k=k(\varphi)$, $L_{a,b,c}\varphi \in L^\infty((0,T)\times U_k; dtdx)$,

$$\sup_n |L_{a^n,b^n,c^n}\varphi|_{L^\infty((0,T)\times U_k;dtdx)} \leqslant C_k,$$

and $L_{a^n,b^n,c^n}\varphi \xrightarrow{n\to\infty} L_{a,b,c}\varphi$ in $L^p((0,T)\times\mathbb{R}^d;dtdx)$ for every $p\geqslant 1$. Let $t\in\mathbb{T}(\varphi)^c$, i.e.

$$\int_{\mathbb{R}^d}\varphi\, d\mu_t^n \xrightarrow{n\to\infty} \int_{\mathbb{R}^d}\varphi\,\varrho(t,x)dx = \int_{\mathbb{R}^d}\varphi d\mu_t,$$

and let $0<s<t<T$. Then

$$\begin{aligned}\Big|\int_{\mathbb{R}^d}\varphi\, d\mu_t^n - \int_{\mathbb{R}^d}\varphi\,\nu_n - \int_s^t\int_{\mathbb{R}^d} L_{a^n,b^n,c^n}\varphi\, d\mu_r^n dr\Big| \\ = \Big|\int_{\mathbb{R}^d}\varphi\, d\mu_s^n - \int_{\mathbb{R}^d}\varphi\, d\nu_n\Big| \leqslant C(\varphi)s,\end{aligned} \tag{1.2.11}$$

where $C(\varphi)$ is as in (1.2.10). Since

$$\lim_n \int_s^t\int_{\mathbb{R}^d} L_{a^n,b^n,c^n}\varphi\, d\mu_r^n dr = \int_s^t\int_{\mathbb{R}^d} L_{a,b,c}\varphi\,\varrho(r,x)\,dxdr,$$

letting first $n\to\infty$ and then $s\to 0$ in (1.2.11) yields

$$\int_{\mathbb{R}^d}\varphi\,\varrho(t,x)\,dx = \int_{\mathbb{R}^d}\varphi\,d\nu + \int_0^t\int_{\mathbb{R}^d} L_{a,b,c}\varphi\,\varrho(r,x)\,dxdr.$$

Therefore, $\mu = (\mu_t)_{t\in[0,T)} = (\varrho(t,x)dx)_{t\in[0,T)}$ is a solution to the Cauchy problem (FPE)+(1.2.3) with initial datum ν.

5. It remains to prove the additional properties of μ claimed in the assertion. Since each ν_n is a probability measure and due to (1.2.8) and $c^n\leqslant 0$, we find, for every $\phi\in C_c^\infty(\mathbb{R}^d)$ with $0\leqslant\phi\leqslant 1$ and $t\in(0,T)$,

$$\int_{\mathbb{R}^d}\phi\, d\mu_t^n - \int_0^t\int_{\mathbb{R}^d}\phi c^n\, d\mu_s^n ds \leqslant 1. \tag{1.2.12}$$

Consider $\phi_N\in C_c^\infty(\mathbb{R}^d)$, $0\leqslant\phi_N\leqslant 1$ such that $\phi_N=1$ on U_N, and let $t\in\bigcap_N \mathbb{T}(\phi_N)^c$, i.e.

$$\int_{\mathbb{R}^d}\phi_N\, d\mu_t^n \xrightarrow{n\to\infty} \int_{\mathbb{R}^d}\phi_N\,\varrho(t,x)dx, \quad \forall N\in\mathbb{N}.$$

Considering such ϕ_N and t in (1.2.12) and letting $n \to \infty$ yields

$$\int_{\mathbb{R}^d} \phi_N \varrho(t,x)\,dx - \int_0^t \int_{\mathbb{R}^d} \phi_N c\,\varrho(s,x)dxds \leqslant 1$$

(precisely: first replace 0 by $\varepsilon > 0$ and use the local weak convergence of ϱ^n to ϱ, then let $\varepsilon \searrow 0$). Finally, letting $N \to \infty$, by Fatou's lemma we conclude, for dt-a.e. $t \in (0,T)$,

$$\int_{\mathbb{R}^d} \varrho(t,x)dx - \int_0^t \int_{\mathbb{R}^d} c\,\varrho(s,x)dxds \leqslant 1 = \nu(\mathbb{R}^d).$$

This proves all remaining assertions. □

1.2.2 Uniqueness of solutions

Next, we present some classical uniqueness results and an example of an ill-posed FPE with smooth coefficients. For the rather long proofs of these results, we refer to Chapter 9 of [9] and to the exercises. Let $a_{ij}, b_i, c : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ be Borel maps and $a = (a_{ij})_{i,j\leqslant}$ symmetric nonnegative definite for all (t,x).

First, assume $c = 0$. The following result is classical.

Proposition 1.2.12. *Assume a, b satisfy $\int_0^T |a(t)|_{C_b^2(\mathbb{R}^d)} + |b(t)|_{C_b^2(\mathbb{R}^d)} dt < \infty$. Then* (FPE) *has a* unique *weakly continuous solution $(\mu_t)_{t\in[0,T]}$ with constant mass for every initial datum $\nu \in \mathcal{M}_b^+$. In particular, for $\nu \in \mathcal{P}$, there is a unique weakly continuous probability solution with initial datum ν.*

Now let $c \leqslant 0$ and denote, for a subprobability measure ν, by $\mathcal{SP}_\nu$ the set of solutions $\mu = (\mu_t)_{t\in[0,T)}$ to (FPE) with initial value ν such that

$$c \in L^1((0,T) \times \mathbb{R}^d; \mu_t dt), \quad b \in L^2((0,T) \times U; \mu_t dt) \quad \forall \text{ balls } U \subseteq \mathbb{R}^d,$$

and such that (1.2.6) holds for dt-a.e. $t \in (0,T)$. In particular, for $\mu \in \mathcal{SP}_\nu$, dt-a.e. μ_t is a subprobability measure. The assumption on b is, for instance, fulfilled, if b is bounded on each $(0,T) \times U$.

We introduce the following assumptions on a.

(H1) For each ball $U \subseteq \mathbb{R}^d$ there is $m(U), M(U) > 0$ such that

$$a(t,x) \geqslant m(U)\,\mathrm{Id}, \quad |a(t,x)| \leqslant M(U), \quad \forall (t,x) \in (0,T) \times U. \tag{1.2.13}$$

(H2) For each ball $U \subseteq \mathbb{R}^d$ there is $\Lambda(U) > 0$ such that for all $i, j \leqslant d$

$$|a_{ij}(t,x) - a_{ij}(t,y)| \leqslant \Lambda(U)|x-y|, \quad \forall x, y \in U, t \in (0,T). \tag{1.2.14}$$

Proposition 1.2.13. *Suppose that (H1) and (H2) hold, $b \in L^p((0,T) \times \mathbb{R}^d)$, $c \in L^{\frac{p}{2}}(0,T) \times \mathbb{R}^d)$ for some $p > d+2$, and that there is a solution $\mu = (\mu_t)_{t\in[0,T)} \in \mathcal{SP}_\nu$ such that*

$$\frac{|a_{ij}|}{1+|x|^2} + \frac{|b_i|}{1+|x|} \in L^1((0,T) \times \mathbb{R}^d; \mu_t dt). \tag{1.2.15}$$

Then μ is the unique element in $\mathcal{SP}_\nu$.

Proposition 1.2.14. *Suppose that (H1) and (H2) hold, $b \in L^p((0,T) \times \mathbb{R}^d)$, $c \in L^{\frac{p}{2}}(0,T) \times \mathbb{R}^d)$ for some $p > d+2$. In addition, assume there is a positive function $V \in C^2(\mathbb{R}^d)$ with $V(x) \xrightarrow{|x|\to\infty} \infty$ such that*

$$L_{a,b,c}V(t,x) \leqslant C + CV(x), \quad \forall (t,x) \in (0,T) \times \mathbb{R}^d, \tag{1.2.16}$$

for some $C > 0$. Then $\mathcal{SP}_\nu$ contains at most one element.

The function V is called a *Lyapunov function.*

Remark 1.2.15. *(i) In both cases one can prove that the unique element in $\mathcal{SP}_\nu$ satisfies (1.2.6) with equality. Hence, if $c = 0$, it is a constant mass solution.*

(ii) If $c = 0$ and b is bounded on each $(0,T) \times U$, then the assertions of both propositions mean that for every probability initial value ν, there is exactly one, respectively at most one, probability solution to (FPE).

Another way to obtain uniqueness of probability solutions is via the corresponding martingale problem, i.e. via the already indicated relation of FPEs to probability theory. We will come back to this topic in due time.

Examples of nonuniqueness. Solutions to Fokker–Planck equations may be non-unique, even for regular coefficients. A simple example in dimension $d = 1$ is

$$a = 0, \quad b(x) = (3x)^{\frac{2}{3}}. \tag{1.2.17}$$

The ODE $\dot{y} = b(y), y(0) = 0$ has the smooth solutions $y^1(t) = 0$ and $y^2(t) = \frac{t^3}{3}$. It is straightforward to check that $(\mu_t^i)_{t\geqslant 0}$, $\mu_t^i = \delta_{y^i(t)}$, $i \in \{1,2\}$, are weakly continuous probability solutions with initial datum $\mu_{|t=0} = \delta_0$ to (FPE), which in this case is the *continuity equation*

$$\partial_t \mu = -\operatorname{div}(b\mu), \quad \mu_{|t=0} = \delta_0.$$

Here the source of non-uniqueness is insufficient regularity of b and the degeneracy of a. However, even for $a = \mathrm{Id}$ and for smooth b, examples of non-uniqueness exist. Indeed there is the following result.

Proposition 1.2.16. *There is $b = (b_1, \dots, b_4) \in C^\infty(\mathbb{R}^4, \mathbb{R}^4)$ such that the FPE on $(0,T) \times \mathbb{R}^4$ with $a = \mathrm{Id}_{4\times 4}$ and b has several probability solutions.*

Proof. See Section 9.2. in [9]. □

1.3 Superposition principle

In this chapter, we set $c = 0$ and consider the Fokker–Planck equation

$$\partial_t \mu_t = L^*_{a,b}\mu_t \big[= \partial^2_{ij}\big(a_{ij}(t,x)\mu_t\big) - \partial_i\big(b_i(t,x)\mu_t\big)\big]$$

(to which we simply refer as "the FPE"), where for Borel coefficients $a = (a_{ij})_{i,j\leqslant d}, b = (b_i)_{i\leqslant d}$ on $\mathbb{R}_+ \times \mathbb{R}^d$ we denote again

$$L_{a,b}\varphi := a_{ij}\partial^2_{ij}\varphi + b_i\partial_i\varphi.$$

On $C_+\mathbb{R}^d := C(\mathbb{R}_+, \mathbb{R}^d)$ with the topology of locally uniform convergence, we denote by π_t, $t \geqslant 0$, the canonical projections $\pi_t w := w(t)$. As usual, we assume a to be pointwise symmetric nonnegative definite.

The martingale problem.

Definition 1.3.1. A *solution to the martingale problem associated with a and b* is a Borel probability measure $P \in \mathcal{P}\big(C_+\mathbb{R}^d\big)$ such that

$$\int_{C_+\mathbb{R}^d}\int_0^T |a_{ij}(s,\pi_s)| + |b_i(s,\pi_s)|dsdP < \infty, \quad \forall i,j \leqslant d,\, T > 0,$$

and for every $\varphi \in C_c^\infty(\mathbb{R}^d)$ (equivalently: $\varphi \in C_b^2(\mathbb{R}^d)$) the real-valued stochastic process $M^\varphi = (M_t^\varphi)_{t\geqslant 0}$ on $C_+\mathbb{R}^d$,

$$M_t^\varphi := \varphi \circ \pi_t + \int_0^t (L_{a,b}\varphi)(s,\pi_s)ds,$$

is a P-martingale w.r.t. the filtration $\mathcal{F}_t := \sigma(\pi_r, 0 \leqslant r \leqslant t)$. The set of solutions with initial condition $P \circ \pi_0^{-1} = \nu$ is denoted by $MP_\nu(a,b)$.

With obvious modifications, the martingale problem can be posed on $[s,\infty)$ instead of $\mathbb{R}_+$. In this case, the initial datum is a pair $(s,\nu) \in \mathbb{R}_+ \times \mathcal{P}$, and martingale problem solutions are measures on $C([s,\infty),\mathbb{R}^d)$. The set of martingale solutions with initial condition (s,ν) is denoted by $MP_{s,\nu}(a,b)$. The results of this section hold for any initial time s. On a path space starting from time s, we denote for $t \geqslant s$ the canonical projection by π_t^s.

A particularly useful property of the martingale problem is the stability of its solutions w.r.t. to disintegration in the sense of the following lemma. For the proof in the case of bounded coefficients, see [20, Thm.6.2.1]. The generalization to unbounded coefficients follows by approximation.

Lemma 1.3.2. *(i) Let $\nu \in \mathcal{P}$, $s \geqslant 0$, $P \in MP_{s,\nu}(a,b)$ and let $(Q_x)_{x\in\mathbb{R}^d} \subseteq \mathcal{P}(C([s,\infty),\mathbb{R}^d))$ be the ν-a.s. unique family such that $x \mapsto Q_x(A)$ is measurable for all $A \in \mathcal{B}(C([s,\infty),\mathbb{R}^d))$,*

$$P(A) = \int_{\mathbb{R}^d} Q_x(A)\,\nu(dx),$$

and $Q_x(\{w : w(s) = x\}) = 1$ (i.e. $(Q_x)_{x\in\mathbb{R}^d}$ is the disintegration family of P w.r.t. π_s). Then $Q_x \in MP_{s,x}(a,b)$ for ν-a.e. $x \in \mathbb{R}^d$.

(ii) Let $t_i \geqslant s$, $Y = (\pi^s_{t_0}, \dots, \pi^s_{t_n}) : C([s,\infty), \mathbb{R}^d) \to (\mathbb{R}^d)^{n+1}$, $\mathcal{A} = \sigma(Y)$, $P \in MP_{s,\nu}(a,b)$ and $(Q_w)_{w \in C([s,\infty),\mathbb{R}^d)}$ be a regular conditional probability of P w.r.t. $\mathcal{A}$. Denote by $Q^{t_n}_w$ the restriction of Q_w to $\mathcal{B}(C([t_n,\infty),\mathbb{R}^d))$, i.e.

$$Q^{t_n}_w(A) := Q_w(u \in C([s,\infty), \mathbb{R}^d \,|\, u_{[t_n,\infty)} \in A),$$

for $A \in \mathcal{B}(C([t_n,\infty),\mathbb{R}^d))$, where $u_{[t_n,\infty)}$ denotes the restriction of $u \in C_+\mathbb{R}^d$ to $C([t_n,\infty),\mathbb{R}^d)$. Then there is a set $A \in \mathcal{A}$, $P(A) = 0$, such that $Q^{t_n}_w \in MP_{t_n,w(t_n)}(a,b)$ for all $w \in A^c$.

Moreover, if $P^1, P^2 \in MP_{s,\nu}(a,b)$ such that $P^1 = P^2$ on $\mathcal{A}$, then A can be chosen such that $P^1(A) = P^2(A) = 0$.

The following standard result is one reason why the martingale problem is popular in probability theory. For the proof, see [19].

Proposition 1.3.3. *If X is a weak solution to the SDE*

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dB_t, \quad t \geqslant 0, \tag{1.3.1}$$

where $\sigma \in \mathbb{R}^{d\times d}$ such that $a = \frac{1}{2}\sigma\sigma^T$, then $\mathcal{L}_X \in M_{\mathcal{L}_{X_0}}(a,b)$. Conversely, for every $\nu \in \mathcal{P}$ and $P \in MP_\nu(a,b)$ there is a weak solution X to this SDE such that $\mathcal{L}_X = P$.

Remark 1.3.4. *Recall that solutions to this SDE are said to be* weakly unique*, if for any two weak solutions X and Y it holds*

$$\mathcal{L}_{X_0} = \mathcal{L}_{Y_0} \implies \mathcal{L}_X = \mathcal{L}_Y.$$

Similarly, solutions are weakly unique for an initial datum $\nu \in \mathcal{P}$*, if the previous implication holds for all weak solutions with $\mathcal{L}_{X_0} = \mathcal{L}_{Y_0} = \nu$.*

For the rest of the chapter, we simply refer to (1.3.1) as "the SDE", and to the corresponding martingale problem as "the martingale problem". It is obvious how to generalize the initial time of the SDE to any $s \geqslant 0$.

There is a wide literature on the martingale problem and, in particular, its connection to Markov processes and probability theory, see for instance the classical reference [20]. In this lecture, we only use the martingale problem as a tool, via the previous proposition.

We have already seen in Section 1.1 that for every weak solution X to the SDE, $(\mathcal{L}_{X_t})_{t\geqslant 0}$ is a weakly continuous probability solution to the FPE. By Proposition 1.3.3, equivalently we have:

Corollary 1.3.5. *$P \in MP_\nu(a,b) \implies (P\circ\pi_t^{-1})_{t\geqslant 0}$ is a weakly continuous probability solution to the FPE with initial datum ν, and all assumptions of Lemma 1.2.6 are true.*

The superposition principle.
The main aim of this chapter is to prove the following theorem, the first cornerstone of the lecture.

Theorem 1.3.6 (Superposition principle)**.** *Let $\sigma_{ij}, b_i : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$, $i,j \leqslant d$, be Borel. For every weakly continuous probability solution $(\mu_t)_{t\geqslant 0}$ to the FPE with coefficients $a = (a_{ij})_{i,j\leqslant d} = \frac{1}{2}\sigma\sigma^T$ and b such that*

$$\int_0^T \int_{\mathbb{R}^d} |a_{ij}| + |b_i| \, d\mu_t dt < \infty, \quad \forall T > 0,\ i,j \leqslant d, \tag{1.3.2}$$

there is a weak solution X to the SDE such that $\mathcal{L}_{X_t} = \mu_t$ for all $t \geqslant 0$.

In particular, if $\mu_0 = \nu$, then $\mathcal{L}_{X_0} = \nu$.

This result is relatively new: It was iteratively proven by Ambrosio, Figalli (Fields-medalist!) and Trevisan between 2004 and 2016, see [1, 11, 21].

Due to the equivalence of the SDE and the martingale problem, we may instead prove that for $(\mu_t)_{t\geqslant 0}$ as in the assertion there is $P \in MP_{\mu_0}(a,b)$ such that $P\circ\pi_t^{-1} = \mu_t$ for all $t \geqslant 0$.

Remark 1.3.7. *(i) It should be remarked that there is no regularity assumption on a and b (except measurability).*

(ii) Assumption (1.3.2) can be generalized to

$$\int_0^T \int_{\mathbb{R}^d} \frac{|a_{ij}| + |\langle b, x\rangle|}{1+|x|^2} \, d\mu_t dt < \infty, \quad \forall T > 0,\ i,j \leqslant d, \tag{1.3.3}$$

see [10], which is essentially sharp ($\langle\cdot,\cdot\rangle$ denotes the standard Euclidean inner product).

For merely local in space integrability there are counterexamples to the assertion of Theorem 1.3.6. For instance, $|MP_\nu(a,b)| \leqslant 1$ for every $\nu \in \mathcal{P}$, if $a(t,x)$ is strictly elliptic for every x and continuous in x uniformly in $t \geqslant 0$, and a and b are locally bounded on $\mathbb{R}_+ \times \mathbb{R}^d$. But [8] contains an example of such coefficients for which the FPE has several probability solutions for every initial probability measure ν (which do not satisfy (1.3.3)).

(iii) Weak continuity and constant mass 1 of $(\mu_t)_{t\geqslant 0}$ is necessary, since the one-dimensional time marginals of any weak SDE solution are a weakly continuous curve of probability measures. However, due to the following proposition, the continuity assumption is no restriction.

(iv) Recall that, by Lemma 1.2.6 (ii), for any solution as in Theorem 1.3.6, (1.2.4) and (1.2.5) hold for all $\varphi \in C_b^2(\mathbb{R}^d)$.

Proposition 1.3.8. *Let $\mu = (\mu_t)_{t\geqslant 0}$ be a solution to $\partial_t \mu_t = L^*_{a,b}\mu_t$ with initial condition $\nu \in \mathcal{P}$ such that $a_{ij}, b_i \in L^1_{\mathrm{loc}}([0,\infty)\times\mathbb{R}^d; \mu_t dt)$, with each μ_t is a nonnegative finite Borel measure und $\operatorname{ess\,sup}_{t\geqslant 0} \mu_t(\mathbb{R}^d) < \infty$ ($\operatorname{ess\,sup}_{t\geqslant 0} \mu_t(\mathbb{R}^d)$ denotes the infimum of numbers $c > 0$ such that $\mu_t(\mathbb{R}^d) \leqslant c$ for all but dt-negligible many $t \geqslant 0$).*

(i) There is a unique vaguely continuous dt-version $(\tilde{\mu}_t)_{t\geqslant 0}$ (i.e. $\mu_t = \tilde{\mu}_t$ dt-a.s.), and $\tilde{\mu}$ also solves the FPE with initial datum ν.

(ii) *If, in addition,* $a_{ij}, b_i \in L^1([0,T] \times \mathbb{R}^d; \mu_t dt)$ *for all* $T > 0$*, then* $\tilde{\mu}$ *is a probability solution and, hence, weakly continuous.*

Proof. The second part of (i) follows from Remark 1.2.3 and the simple observation that any two solutions from the same dt-equivalence class have the same initial datum. (ii) follows from Exercise 1.3. First part of (i): Exercise 2.1. □

The superposition principle allows to prove uniqueness of FPE-probability solutions via weak uniqueness for the SDE:

Corollary 1.3.9. *Let* $s \geqslant 0, \nu \in \mathcal{P}$*. If solutions to the martingale problem (the SDE) with initial condition* (s, ν) *are (weakly) unique, then there is, up to* dt*-zero sets, at most one probability solution to the FPE with initial condition* (s, ν) *such that* (1.3.2) *holds (with* s *instead of* 0*).*

Proof. Without loss of generality, let $s = 0$. Suppose $\mu^i = (\mu^i)_{t\geqslant 0}$, $i \in \{1,2\}$, are two probability solutions to the FPE with initial datum ν, satisfying (1.3.2). By Proposition 1.3.8, there exist weakly continuous dt-versions $(\tilde{\mu}^i_t)_{t\geqslant 0}$ (with initial datum ν), and by the superposition principle, there exist weak SDE solutions $\tilde{X}^i$ such that $\mathcal{L}_{\tilde{X}^i_t} = \tilde{\mu}^i_t$ for all $t \geqslant 0$ and $i \in \{1,2\}$. By assumption, $\tilde{\mu}^1_t = \tilde{\mu}^2_t$ for all t follows. Hence also $\mu^1_t = \mu^2_t$ dt-a.s. □

The reverse uniqueness implication is not true, i.e. uniqueness of FPE-probability solutions for *one* initial datum ν does not imply weak uniqueness of SDE solutions with initial distribution ν. Instead, one needs FPE-uniqueness for *sufficiently many* initial times and measures:

Proposition 1.3.10. *Suppose weakly continuous probability solutions* $(\mu_t)_{t\geqslant s}$ *to the FPE satisfying* (1.3.2) *(with* s *instead of* 0*) are unique for every initial condition* $(s, \delta_x) \in \mathbb{R}_+ \times \mathcal{P}$, $x \in \mathbb{R}^d$*. Then solutions to the martingale problem (the SDE) are (weakly) unique for every initial datum* (s, δ_x), $s \geqslant 0$, $x \in \mathbb{R}^d$*.*

Proof. Let $x \in \mathbb{R}^d$. We have to prove

$$P^1, P^2 \in MP_x(a,b) \implies P^1 = P^2,$$

where the equality on the RHS is equivalent to

$$P^1 \circ (\pi_{t_0}, \dots, \pi_{t_n})^{-1} = P^2 \circ (\pi_{t_0}, \dots, \pi_{t_n})^{-1}, \quad \forall 0 \leqslant t_0 < \dots < t_n \tag{1.3.4}$$

for all $n \in \mathbb{N}_0$. Then the assertion follows, since the proof for $s \neq 0$ is the same. The assumption entails the previous equality for $n = 0$, since by Corollary 1.3.5 the curves $(P^1 \circ \pi_t^{-1})_{t\geqslant 0}, (P^2 \circ \pi_t^{-1})_{t\geqslant 0}$ are weakly continuous probability solutions to the FPE satisfying (1.3.2) with initial condition δ_x.

We proceed by induction. Assume (1.3.4) holds for $n-1 \in \mathbb{N}$. For arbitrary $f_i : \mathbb{R}^d \to \mathbb{R}$ measurable bounded, $i \in \{0, \dots, n\}$ and $0 \leqslant t_0 < \dots < t_n$, we have to show

$$\mathbb{E}_{P^1}\big[f_0(\pi_{t_0}) \cdots f_n(\pi_{t_n})\big] = \mathbb{E}_{P^2}\big[f_0(\pi_{t_0}) \cdots f_n(\pi_{t_n})\big]. \tag{1.3.5}$$

Let $(Q^i_w)_{w\in C_+\mathbb{R}^d}$ be a r.c.p. of P^i w.r.t. $\sigma(\pi_{t_0},\dots,\pi_{t_{n-1}})$. By Lemma 1.3.2 (ii) $Q^{i,t_{n-1}}_w \in MP_{t_{n-1},w(t_{n-1})}(a,b)$ for P_i-a.e. w. By the last part of the lemma and the induction assumption, the exceptional set A such that the previous inclusion holds for all $w \in A^c$ can be chosen independently of $i \in \{1,2\}$. Hence, by assumption, for all $w \in A^c \cap N_2^c$, where N_2, $P_2(N_2)=0$, is such that

$$\mathbb{E}_{Q^2_w}[f_n(\pi_{t_n})] = \mathbb{E}_{P_2}[f_n(\pi_{t_n})|\sigma(\pi_{t_0},\dots,\pi_{t_{n-1}})], \quad \forall w \in N_2^c,$$

we have

$$\mathbb{E}_{Q^{1,t_{n-1}}_w}\big[f_n(\pi^{t_{n-1}}_{t_n})\big] = \mathbb{E}_{Q^{2,t_{n-1}}_w}\big[f_n(\pi^{t_{n-1}}_{t_n})\big] = \mathbb{E}_{P_2}[f_n(\pi_{t_n})|\sigma(\pi_{t_0},\dots,\pi_{t_{n-1}})](w),$$

which implies that for

$$H: C_+\mathbb{R}^d \to \mathbb{R}, \quad H(w) = \mathbb{E}_{Q^{1,t_{n-1}}_w}\big[f_n(\pi^{t_{n-1}}_{t_n})\big],$$

which is bounded $\sigma(\pi_{t_0},\dots,\pi_{t_{n-1}})$-measurable, we have

$$H = \mathbb{E}_{P^i}\big[f_n(\pi_{t_n})|\sigma(\pi_{t_0},\dots,\pi_{t_{n-1}})\big],$$

both P^1- and P^2-a.s. Now we can conclude, since the LHS and RHS of (1.3.5) equal $\mathbb{E}_{P^i}\big[f_0(\pi_{t_0})\cdots f_{n-1}(\pi_{t_{n-1}})H\big]$, $i=1$ and $i=2$, respectively. But for $i=1$ and $i=2$, these integral values are the same by the induction assumption, since the integrand is $\sigma(\pi_{t_0},\dots,\pi_{t_{n-1}})$-measurable. □

A natural question is whether the previous proposition implies uniqueness for *all* initial data $(s,\nu) \in \mathbb{R}_+ \times \mathcal{P}$. The answer is positive:

Proposition 1.3.11. *Let $s \geqslant 0$ and assume $|MP_{s,x}(a,b)| \leqslant 1$ for all $x \in \mathbb{R}^d$. Then $|MP_{s,\nu}(a,b)| \leqslant 1$ for all $\nu \in \mathcal{P}$.*

Proof. Exercise 3.1. □

Corollary 1.3.12. *Under the assumption of Proposition 1.3.11, the FPE has at most one weakly continuous probability solution satisfying* (1.3.2) *(with s instead of 0) with initial condition (s,ν) for* every *probability measure ν and $s \geqslant 0$.*

Proof. Let $(s,\nu) \in \mathbb{R}_+ \times \mathcal{P}$. Proposition 1.3.11 yields $|M_{s,\nu}(a,b)| \leqslant 1$, and the assertion follows from Corollary 1.3.9. □

1.3.1 Deterministic special case

Here we consider the case $a=0$, i.e. the FPE becomes the *continuity equation*

$$\partial_t \mu_t = -\operatorname{div}\big(b\mu_t\big), \quad t \in (0,\infty). \tag{1.3.6}$$

In this case, we have the following characterization of solutions to the martingale problem:

$$P \in MP_\nu(0,b) \iff P \in \mathcal{P}(C_+\mathbb{R}^d) \text{ such that } P(C_{ac}(b)) = 1, P\circ\pi_0^{-1} = \nu,$$
$$\int_{C_+\mathbb{R}^d}\int_0^T |b(t,w(t))|dtdP(w) < \infty\, \forall T > 0,$$

where $C_{ab}(b)$ denotes the set of absolutely continuous maps $y : [0, \infty) \to \mathbb{R}^d$ such that $y'(t) = b(t, y(t))$ dt-a.s. (i.e. the set of integral solutions to this ODE).

In this case, the superposition principle asserts: For any weakly continuous probability solution $(\mu_t)_{t\geqslant 0}$ to (1.3.6) such that $\int_0^T \int_{\mathbb{R}^d} |b(t, x)| d\mu_t dt < \infty$ for all $T > 0$ there is a probability measure P on the set of integral solutions to the ODE corresponding to b such that $P \circ \pi_t^{-1} = \mu_t$, $t \geqslant 0$.

In particular, the existence of such a solution $(\mu_t)_{t\geqslant 0}$ with initial datum ν yields the existence of at least one integral solution to the ODE with initial datum x for ν-a.e. $x \in \mathbb{R}^d$. Conversely, if for ν-a.e. x there is at most one integral solution to the ODE with initial datum x, then there is at most one weakly continuous probability solution to the continuity equation with initial datum ν satisfying the previously mentioned integrability condition.

In general, there may be many ODE solutions from the same initial value x and the disintegration measures Q_x of P with support on those ODE solutions starting from x need not be Dirac measures, i.e. need not be concentrated on a single ODE solution. This is the reason for the name *superposition principle*: The path measure P may *superpose* many ODE solutions with the same initial value.

Analog statements hold for initial times $s > 0$.

1.3.2 Proof

We now prove Theorem 1.3.6, closely following Trevisan [21], restricting to the time interval $[0, 1]$ instead of $\mathbb{R}_+$. The latter case is a simple modification of the proof below (the definition of solution to the FPE and the martingale problem on $[0, 1]$ is the same as on $\mathbb{R}_+$, with the obvious modifications). The idea is the following.

(1) Approximate a and b by sufficiently regular coefficients a^n and b^n, consider the corresponding FPEs with solutions μ^n for which the assertion is already known, such that
$$\mu^n \to \mu \text{ and } (a^n, b^n) \to (a, b)$$
in a suitable sense. This yields the existence of $P^n \in MP_{\mu_0^n}(a^n, b^n)$ with $P^n \circ \pi_t^{-1} = \mu_t^n$.

(2) Prove tightness of $(P^n)_{n\in\mathbb{N}}$ in $\mathcal{P}(C_{[0,1]}\mathbb{R}^d)$ in order to extract a weak limit point P.

(3) Prove $P \in MP_{\mu_0}(a, b)$.

Remark 1.3.13. *If $\mu^n \to \mu$ weakly (which will be the case in the proof below), then $P \circ \pi_t^{-1} = \mu_t$ follows from the weak convergence $P^n \to P$ and $P^n \circ \pi_t^{-1} = \mu_t^n$.*

First assume, writing $a(t) = [x \mapsto a(t, x)]$ and likewise for b,

$$\text{(A1)} \qquad \int_0^1 |a(t)|_{C_b^2(\mathbb{R}^d)} + |b(t)|_{C_b^2(\mathbb{R}^d)} dt < \infty.$$

In this case, the superposition principle holds. Indeed, by the standard Picard–Lindelöf theorem, under (A1) the SDE has a unique weak solution for any initial condition $\nu \in \mathcal{P}$. On the other hand, by Proposition 1.2.12, for any initial probability measure ν, there is at most one weakly continuous probability solution to the FPE (by (A1), every such solution satisfies (1.3.2)). Hence, by Proposition 1.1.3, such a solution exists and is necessarily the one-dimensional time marginal curve of the unique SDE solution.

The generalization from this base case to the full assertion proceeds via several steps: We verify the assertion under each of the following increasingly general assumptions. Below we denote by $U \subseteq \mathbb{R}^d$ an arbitrary ball.

$$\text{(A2)} \qquad \int_0^1 \sup_x |a(t,x)| + \sup_x |b(t,x)| dt < \infty,$$

$$\text{(A3)} \qquad \int_0^1 |a(t)|_{L^\infty(U)} + |b(t)|_{L^\infty(U)} dt < \infty \ \ \forall U \qquad \text{and (1.3.2) holds,}$$

$$\text{(A4)} \qquad \int_0^1 \int_{\mathbb{R}^d} |a(t,x)| + |b(t,x)| d\mu_t dt < \infty.$$

Each step proceeds via (1)-(3), and the main task in each step is to choose a suitable approximations of the coefficients and the solution.

We first present the general ideas for (1)-(3) before applying them to each generalization step. Let $\mu = (\mu_t)_{t\in[0,1]}$ be the solution from the assertion.

(1) Approximation

(1).1 Image measures of smooth maps. Let $g = (g^1, \dots, g^d) \in C^2(\mathbb{R}^d, \mathbb{R}^d)$ have uniformly bounded first- and second-order derivatives, and set

$$\mu^g = (\mu_t^g)_{t\in[0,1]} := (\mu_t \circ g^{-1})_{t\in[0,1]}.$$

Note that $\varphi \circ g \in C_b^2(\mathbb{R}^d)$ for $\varphi \in C_b^2(\mathbb{R}^d)$ and

$$L_{a,b}(\varphi \circ g) = \sum_{k=1}^d L_{a,b}(g^k)[(\partial_k \varphi) \circ g] + \sum_{k,l=1}^d a_{ij} \partial_i g^k \partial_j g^l [(\partial^2_{kl} \varphi) \circ g].$$

For any $t \in [0,1]$ and $k, l \leqslant d$, let $a^g_{kl}(t), b^g_k(t) : \mathbb{R}^d \to \mathbb{R}$ be Borel maps such that

$$\mathbb{E}_{\mu_t}[a_{ij}(t) \partial_i g^k \partial_j g^l | \sigma(g)] = a^g_{kl}(t) \circ g, \quad \mathbb{E}_{\mu_t}[L_{a,b} g^k(t) | \sigma(g)] = b^g_k(t) \circ g, \quad \mu_t - \text{a.s.}$$

(Einstein summation convention is used for repeated indices). $a^g_{kl}(t)$ and $b^g_k(t)$ exist and are uniquely determined μ_t-a.s. by the factorization lemma. Note that $a^g_{kl}(t)$ and $b^g_k(t)$ are the density of $[(a_{ij}(t)\partial_i g^k \partial_j g^l)\mu_t] \circ g^{-1}$ and $[(L_{a,b}(g^k)\mu_t] \circ g^{-1}$ w.r.t. μ^g_t, respectively. The curve μ^g is a weakly continuous probability solution to

$$\partial_t \nu_t = L^*_{a^g, b^g} \nu_t, \quad t \in [0,1].$$

Moreover, by definition we have for all $t \in [0,1]$ and $p \in [1, \infty]$

$$|a^g_{kl}(t)|_{L^p(\mathbb{R}^d;\mu^g_t)} \leqslant C |a(t)|_{L^p(\mathbb{R}^d;\mu_t)}, \quad |b^g_k(t)|_{L^p(\mathbb{R}^d;\mu^g_t)} \leqslant C \big| |a(t)| + |b(t)| \big|_{L^p(\mathbb{R}^d;\mu_t)}$$

(1.3.7)

by the contraction property of conditional expectations, where $C > 0$ depends only on the L^∞-norm of the first- and second-order derivatives of g.

(1).2 Mollification by convolutions. Let $\varrho : \mathbb{R}^d \to \mathbb{R}$, $\varrho > 0$, be a smooth probability density (w.r.t. dx), and set $\mu * \varrho = (\mu_t * \varrho)_{t\in[0,1]}$, i.e. $\int_{\mathbb{R}^d} f\, d(\mu_t * \varrho) := \int_{\mathbb{R}^d} (f * \varrho)\, d\mu_t$. Since $\varphi * \varrho \in C_b^2(\mathbb{R}^d)$ for $\varphi \in C_b^2(\mathbb{R}^d)$ and

$$L_{a,b}(\varphi * \varrho) = b_i(\partial_i \varphi) * \varrho + a_{ij}(\partial_{ij}^2 f) * \varrho,$$

by defining

$$a_{ij}^\varrho(t,x) := \frac{d\big((a_{ij}(t)\mu_t) * \varrho\big)}{d(\mu_t * \varrho)}(x), \quad b_i^\varrho(t,x) := \frac{d\big((b_i(t)\mu_t) * \varrho\big)}{d(\mu_t * \varrho)}(x),$$

(well-defined by Lemma 1.3.14) we find that $\mu * \varrho$ is a weakly continuous curve of probability measures and solves the FPE

$$\partial_t \nu_t = L^*_{a^\varrho, b^\varrho} \nu_t, \quad t \geqslant 0.$$

The following lemma can be found as Lemma A.1 in [21]. Here $\mathcal{M}_b$ denotes the set of signed Borel measures on $\mathbb{R}^d$ with finite total variation, and we denote by $D^i\varrho$ the collection of i-th partial derivatives of ϱ (i.e. $D^1\varrho = \nabla\varrho$; $D^2\varrho$ is the Hessian of ϱ, etc.).

Lemma 1.3.14. *Let ϱ as above also satisfy $|D^i\varrho| \leqslant C\varrho$ pointwise for $i \in \{1,\dots,k\}$ for some $C \geqslant 0$ and $k \in \mathbb{N}$, and let $\eta^1 \in \mathcal{M}_b^+$, $\eta^2 \in \mathcal{M}_b$ with $\eta^2 = h\eta^1$, where $h : \mathbb{R}^d \to \mathbb{R}$. Then $\eta^2 * \varrho$ has a density h_ϱ w.r.t. $\eta^1 * \varrho$, h_ϱ has a C^k-version and*

$$|h_\varrho|_{L^p(\mathbb{R}^d;\eta^1 * \varrho)} \leqslant |h|_{L^p(\mathbb{R}^d;\eta^1)}, \quad \forall p \in [1,\infty].$$

Morevoer, for every convex map $\Theta : \mathbb{R} \to \mathbb{R}_+$

$$\int_{\mathbb{R}^d} \Theta(|h_\varrho|) d(\eta^1 * \varrho) \leqslant \int_{\mathbb{R}^d} \Theta(|h|)\, d\eta^1. \tag{1.3.8}$$

We will apply the lemma for $\eta^1 = \mu_t$, $\eta^2 = a_{ij}(t)\mu_t$, $h = a_{ij}(t)$ and $h_\varrho = a_{ij}^\varrho(t)$, and, similarly, for b_i and b_i^ϱ.

(2) Tightness
Recall: A sequence of Borel probability measures $(\mu_n)_{n\in\mathbb{N}}$ on a metric space S is called *tight*, if for every $\varepsilon > 0$, there is a compact set $K \subseteq S$ such that $\mu_n(K) \geqslant 1-\varepsilon$ for all $n \in \mathbb{N}$. If $(\mu_n)_{n\in\mathbb{N}}$ is tight, it contains a weakly converging subsequence. A sufficient criterion for tightness is the existence of a coercive function $f : S \to \mathbb{R}_+$ (i.e. $\{f \leqslant c\}$ is compact for all $c \geqslant 0$) such that

$$\sup_{n\in\mathbb{N}} \int_S f\, d\mu_n < \infty.$$

See also Exercise 4.1. For the rest of this section, for $P \in \mathcal{P}(C_{[0,1]}\mathbb{R}^d)$, we sometimes write $P_t := P \circ \pi_t^{-1}$. We need the following result, see [21, Thm.A.2, Cor.A.5].

Proposition 1.3.15. *Let $\theta, \Theta_1, \Theta_2 : \mathbb{R}_+ \to \mathbb{R}_+$ be functions such that Θ_i, $i \in \{1,2\}$, are convex with*

$$\lim_{x\to\infty} \theta(x) = \lim_{x\to\infty} \frac{\Theta_i(x)}{x} = \infty, \quad i \in \{1,2\}.$$

Then there exists a coercive map $\Psi : C_{[0,1]}\mathbb{R}^1 \to \mathbb{R}_+ \cup \{\infty\}$ such that for all Borel maps a_{ij}, b_i on $[0,1] \times \mathbb{R}^d$ such that $a = (a_{ij})_{i,j\leqslant d}$ is pointwise nonnegative definite and symmetric, and every $P \in MP_{P_0}(a,b)$, we have

$$\mathbb{E}_P[\Psi(f\circ\pi)] \leqslant \int_{\mathbb{R}^d} \theta(|f|)\, dP_0 + \int_0^1 \Theta_1\big(|L_{a,b}f|\big) + \Theta_2\big(a_{ij}\partial_i f \partial_j f\big) dP_t dt, \quad \forall f \in C_b^2(\mathbb{R}^d). \tag{1.3.9}$$

Here we use the notation $f\circ\pi : C_{[0,1]}\mathbb{R}^d \to C_{[0,1]}\mathbb{R}^1$, $f\circ\pi(w) = [t \mapsto f(\pi_t(w))]$.

Note that the functions θ, Θ_i, $i \in \{1,2\}$, and Ψ are independent of a and b. In principle, both sides of the inequality may equal $+\infty$. If for our sequence of FPE-solutions μ^n for approximate coefficients a^n, b^n, we can find θ, Θ_i, $i \in \{1,2\}$ such that the RHS of (1.3.9) is finite and bounded uniformly in n for the corresponding martingale solutions P^n, then (1.3.9) provides a criterion to prove tightness of $(P^n)_{n\in\mathbb{N}}$.

(3) Limit
Here we assume $(P^n)_{n\in\mathbb{N}}$ obtained in part (1) has a weak limit point P, and we prove $P \in MP_{P_0}(a,b)$. The latter holds if and only if: for all $s,t \in [0,1]$, $s \leqslant t$, $\varphi \in C_b^2$, $|\varphi|_{C_b^2} \leqslant 1$, and $h : C_{[0,1]}\mathbb{R}^d \to \mathbb{R}$ continuous, bounded and $\mathcal{F}_s$-measurable it holds

$$\int_{C_{[0,1]}\mathbb{R}^d} h\Big[\varphi\circ\pi_t - \varphi\circ\pi_s - \int_s^t L_{a,b}\varphi(r,\pi_r)dr\Big] dP = 0. \tag{1.3.10}$$

This identity holds for P^n, a^n and b^n instead of P, a and b, hence by the weak convergence $P^n \to P$ it remains to prove

$$\int_{C_{[0,1]}\mathbb{R}^d} h\Big[\int_s^t L_{a^n,b^n}\varphi(r,\pi_r)dr\Big] dP^n - \int_{C_{[0,1]}\mathbb{R}^d} h\Big[\int_s^t L_{a,b}\varphi(r,\pi_r)dr\Big] dP \xrightarrow{n\to\infty} 0. \tag{1.3.11}$$

We now prove this convergence for both types of approximations from step (1).

(3).1 Image measures For each $n \in \mathbb{N}$, let g_n satisfy the assumptions of g in (1).1 and, in addition, $g_n(x) = x$ on $B_n(0)$, $D^1 g_n \xrightarrow{n\to\infty} 1_{d\times d}$ (the unit matrix), $D^2 g_n \xrightarrow{n\to\infty} 0$ and $|D^i g_n(x)| \leqslant C$ for all $i \in \{1,2\}$, $n \in \mathbb{N}$ and $x \in \mathbb{R}^d$. Denote $\mu^n := \mu^{g_n}$, where the latter is defined as in (1).1 above.

Let $\bar{L}$ denote *any* operator of type

$$\bar{L} = \bar{a}_{ij}\partial_{ij}^2 + \bar{b}_i\partial_i$$

for continuous (in (t,x)) and compactly supported (in x) coefficients $\bar{a}_{ij}$ and $\bar{b}_i$, $i,j \leqslant d$. We subtract from the LHS of (1.3.11) the term

$$\int_{C_{[0,1]}\mathbb{R}^d} h\Big[\int_s^t \bar{L}\varphi(r,\pi_r)dr\Big]dP^n - \int_{C_{[0,1]}\mathbb{R}^d} h\Big[\int_s^t \bar{L}\varphi(r,\pi_r)dr\Big]dP, \tag{1.3.12}$$

and note that this difference vanishes as $n \to \infty$ by the weak convergence of $P^n \to P$. Consequently, the $\limsup_n$ of the absolute value of the LHS of (1.3.11) is not affected by first subtracting this difference term before taking absolute value and $\limsup_n$. So, we estimate the $\limsup_n$ of the absolute value of the LHS of (1.3.11), up to a multiplicative constant depending only on h, by

$$\limsup_n \int_s^t \int_{\mathbb{R}^d} |L_{a^n,b^n}\varphi - \bar{L}\varphi| d\mu_r^n dr + \int_s^t \int_{\mathbb{R}^d} |L_{a,b}\varphi - \bar{L}\varphi| d\mu_r dr. \tag{1.3.13}$$

By definition of a^n, b^n and μ^n, the first summand of the previous line is equal to

$$\begin{aligned}
&\int_s^t \int_{\mathbb{R}^d} |\mathbb{E}_{\mu_r}[L_{a,b}(\varphi\circ g_n)|\sigma(g_n)] - \bar{L}\varphi\circ g_n| d\mu_r dr \\
&\quad = \int_s^t \int_{\mathbb{R}^d} |\mathbb{E}_{\mu_r}[L_{a,b}(\varphi\circ g_n) - \bar{L}\varphi\circ g_n|\sigma(g_n)] d\mu_r dr \\
&\quad \leqslant \int_s^t \int_{\mathbb{R}^d} |L_{a,b}(\varphi\circ g_n) - \bar{L}\varphi\circ g_n| d\mu_r dr \\
&\quad \leqslant \int_s^t \int_{\mathbb{R}^d} \sum_{k,l=1}^d |a_{ij}\partial_i g_n^k \partial_j g_n^l - \bar{a}_{kl}\circ g_n| + \sum_{k=1}^d |L_{a,b}(g_n^k) - \bar{b}_k\circ g_n| d\mu_r dr,
\end{aligned}$$

where for the equality we used that $\bar{L}\varphi\circ g_n$ is $\sigma(g_n)$-measurable, the first inequality is due to the L^1-contraction property of conditional expectations, and the second inequality is obtained by writing the previous line explicitly and using the estimate $|\varphi|_{C_b^2} \leqslant 1$.

Using the convergence properties of g_n and its first- and second-order derivatives specified above and taking $\limsup_n$ of the RHS gives the estimate

$$\limsup_n \int_s^t \int_{\mathbb{R}^d} |L_{a^n,b^n}\varphi - \bar{L}\varphi| d\mu_r^n dr \leqslant \int_s^t \int_{\mathbb{R}^d} \sum_{i,j=1}^d |a_{ij} - \bar{a}_{ij}| + \sum_{i=1}^d |b_i - \bar{b}_i|\, d\mu_r dr.$$

Hence, taking into account (1.3.13), altogether the $\limsup_n$ of the LHS of (1.3.11) is bounded above by

$$2C \int_s^t \int_{\mathbb{R}^d} \sum_{i,j=1}^d |a_{ij} - \bar{a}_{ij}| + \sum_{i=1}^d |b_i - \bar{b}_i|\, d\mu_r dr.$$

Since $\bar{a}_{ij}$, $\bar{b}_{ij} : [0,1]\times\mathbb{R}^d \to \mathbb{R}$ were arbitrary continuous and compactly supported (in $x \in \mathbb{R}^d$) maps, and since the class of such maps is dense in $L^1((s,t)\times\mathbb{R}^d;\zeta)$ for every locally finite Borel measure ζ on $[0,1]\times\mathbb{R}^d$, we can make the previous sum arbitrary small and, hence, conclude (1.3.11) (here $\zeta = \mu_t dt$).

(3).2 Mollifications
Let $(\varrho_n)_{n\in\mathbb{N}}$ be a sequence of smooth probability densities w.r.t. dx such that $\varrho_n(x)dx \xrightarrow{n\to\infty} \delta_0$ weakly. Then the argument is similar to the previous case. Details are left to the reader, or see [21].

Proof of Theorem 1.3.6
Now we apply (1)-(3) to coefficients satisfying (A2) to deduce the assertion from the validity of the assertion for coefficients satisfying (A1). Then, one assumes (A3) and, via (1)-(3), proves the assertion in this case, relying on the validity in the case (A2) proven before. Finally, one proceeds similarly under the general assumption, i.e. (A4), by relying on(A3). In each step, one has to make fitting choices along (1)-(3).

Under assumption (A2).
(1). Let $\zeta(x) := C\exp(-\sqrt{1+|x|^2})$, where $C>0$ such that $|\zeta|_{L^1}=1$, and set $\varrho_n(x) := n^d\zeta(nx)$, $n\in\mathbb{N}$. Then $|D^i\varrho_n| \leqslant cn^2\varrho_n$, $i\in\{1,2\}$ for some $c>0$, and $\varrho_n(x)dx \xrightarrow{n\to\infty} \delta_0$ weakly. Set $\mu^n := \mu * \varrho_n$, and note $\mu^n_t \xrightarrow{n\to\infty} \mu_t$ weakly for all $t\in[0,1]$. By (1).2, μ^n is a weakly continuous probability solution for the FPE with coefficients $a^n := a^{\varrho_n}$ and $b^n := b^{\varrho_n}$, defined as in (1).2 with ϱ_n in place of ϱ. By Lemma 1.3.14, we have for all $p\geqslant 1$ and $t\in[0,1]$

$$|a^n_{ij}(t)|_{L^p(\mathbb{R}^d;\mu^n_t)} \leqslant |a(t)|_{L^p(\mathbb{R}^d;\mu_t)}.$$

An analog estimate holds for b^n_i, $i\leqslant d$.

Since a^n_{ij} and b^n_i satisfy (A1), there is a family $(P^n)_{n\in\mathbb{N}}$, $P^n\in MP_{\mu^n_0}(a^n,b^n)$, such that $P^n_t = \mu^n_t$ for all $t\in[0,1]$.

(2). Since the sequence $(\mu^n_0)_{n\in\mathbb{N}}$ converges weakly to μ_0, it is tight, and thus there is an increasing function $\theta:\mathbb{R}_+\to\mathbb{R}_+$ with $\lim_{x\to\infty}\theta(x)=\infty$ such that $\sup_n\int_{\mathbb{R}^d}\theta(|x|)\,d\mu^n_0 \leqslant 1$ (cf. Exercise 4.1). By (A2) and the de la Vallée Poussin criterion, there is a nondecreasing, convex map $\Theta:\mathbb{R}_+\to\mathbb{R}_+$ with $\lim_{x\to\infty}\frac{\Theta(x)}{x}=\infty$ such that

$$\int_0^1 \Theta\big(\sup_x|a(t,x)|\big) + \Theta\big(\sup_x|b(t,x)|\big)\,dt < \infty.$$

For $k\in\{1,\dots,d\}$, denote by $x_k:\mathbb{R}^d\to\mathbb{R}$ the map $x=(x_1,\dots,x_d)\mapsto x_k$. Apply Proposition (1.3.15) with θ, $\Theta_1=\Theta_2=\Theta$, $f=x_k\chi_R$, $k\leqslant d$, where $\chi_R:\mathbb{R}^d\to[0,1]$ denotes a standard cutoff function, equal to 1 on $B_R(0)$, $R>0$, to obtain the existence of a coercive (hence lower semicontinuous) map Ψ such that

$$\begin{aligned}&\mathbb{E}_{P^n}[\Psi(x_k\chi_R\circ\pi)]\\ &\quad\leqslant \int_{\mathbb{R}^d}\theta(|x_k\chi_R|)d\mu^n_0 + \int_0^1\Theta\big(|L_{a^n,b^n}x_k\chi_R|\big)+\Theta\big(a^n_{ij}\partial_i(x_k\chi_R)\partial_j(x_k\chi_R)\big)d\mu^n_t dt.\end{aligned}$$

Note $x_k\chi_R \xrightarrow{R\to\infty} x_k$, $|x_k\chi_R|\leqslant|x_k|$, $\partial_i(x_k\chi_R)$ is bounded uniformly in $R>0$ and $\partial^2_{ij}(x_k\chi_R)$ converges to 0 pointwise as $R\to\infty$. Hence, the lower semicontinuity of

Ψ, the monotonicity of θ, Fatou's lemma and dominated convergence imply

$$\mathbb{E}_{P^n}[\Psi(x_k \circ \pi)] \leqslant \int_{\mathbb{R}^d} \theta(|x_k|) d\mu_0^n + \int_0^1 \int_{\mathbb{R}^d} \Theta\big(b_k^n(t)\big) + \Theta\big(a_{kk}^n(t)\big)\, d\mu_t^n dt.$$

By construction of θ and (1.3.8) the RHS is bounded above by

$$1+\int_0^1 \int_{\mathbb{R}^d} \Theta\big(|b_k(t)|\big)+\Theta\big(|a_{kk}(t)|\big)\, d\mu_t dt \leqslant 1+\int_0^t \Theta\big(\sup_x |b(t,x)|\big)+\Theta\big(\sup_x |a(t,x)|\big) dt < \infty.$$

Since $w \mapsto \sum_{k=1}^d \Psi(x_k \circ w)$ is coercive on $C_{[0,1]}\mathbb{R}^d$ (Exercise 5.2), we obtain tightness of $(P^n)_{n\in\mathbb{N}}$.

(3). Follows from (3).2 above. Remark 1.3.13 concludes this part of the proof.

Under assumption (A3). Proceed similarly as in the previous case, but use image measures instead of mollifications to approximate a and b. Steps (2) and (3) follow similarly as in the previous case.

Under assumption (A4). Similarly to the previous cases, approximate a and b by convolutions. For the detailed arguments of the previous two cases, please see [21, p.38,39]. □

2 Connection to Markov processes

There is vast literature on the theory of Markov processes and their applications. A very short list of standard, mostly rather recent, references (in no particular order), including material on discrete time Markov processes (usually called Markov chains) is: [Stroock2014], [LeGall2016], [Liggett2010], [Eberle2010] (lecture notes), [Kirkwood2015], [Wentzell81], [GikhmanSkorokhod04].

2.1 Brief repetition of Markov processes

Let $(S, \mathcal{S})$ be a measurable space. A map $\Lambda : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathcal{M}_b^+(S) \to \mathcal{M}_b^+(S)$ has the *flow property*, if

$$\Lambda(s,t,\zeta) = \Lambda\big(r,t,\Lambda(s,r,\zeta)\big), \quad \forall 0 \leqslant s \leqslant r \leqslant t,\, \zeta \in \mathcal{M}_b^+(S). \tag{2.1.1}$$

Likewise, Λ has the flow property in $\mathcal{M} \subset \mathcal{M}_b^+(S)$, if $\Lambda(s,t,\mathcal{M}) \subseteq \mathcal{M}$ for all $0 \leqslant s \leqslant t$, and (2.1.1) holds for all $\zeta \in \mathcal{M}$.

Definition 2.1.1. A tuple $(\Omega, \mathcal{F}, (X_t)_{t\geqslant 0}, (P_x)_{x\in S})$, consisting of a measurable space $(\Omega, \mathcal{F})$, an S-valued stochastic process $X = (X_t)_{t\geqslant 0}$ on Ω and a family $(P_x)_{x\in S} \subseteq \mathcal{P}(\Omega)$ is a *Markov process*, if

(i) $x \mapsto P_x(\Gamma)$ is $\mathcal{S}-$measurable for all $\Gamma \in \mathcal{F}$,

(ii) there is a filtration $(\mathcal{F}_t)_{t\geqslant 0}$ on $(\Omega, \mathcal{F})$ such that each X_t is $\mathcal{F}_t$-measurable and

$$P_x(X_{t+s} \in B|\mathcal{F}_s) = P_{X_s}(X_t \in B) \;\; P_x-\text{a.s.} \quad \forall s,t \geqslant 0,\, B \in \mathcal{S}, x \in S. \tag{2.1.2}$$

A Markov process is called *normal*, if $P_x(X_0 = x) = 1$ for all $x \in S$.

Without further mentioning, we always consider normal Markov processes.

Remark 2.1.2. *If (ii) is true for $(\hat{\mathcal{F}}_t)_{t\geqslant 0}$, and $(\mathcal{F}_t)_{t\geqslant 0}$ is such that $\mathcal{F}_t \subseteq \hat{\mathcal{F}}_t$ for all $0 \leqslant t$, then (2.1.2) holds with $(\mathcal{F}_t)_{t\geqslant 0}$, if $(X_t)_{t\geqslant 0}$ is $(\mathcal{F}_t)_{t\geqslant 0}$-adapted.*

The generic framework for Markov processes with continuous sample paths is the *canonical model*:

Example 2.1.3 (Canonical model)**.** $\Omega = C(\mathbb{R}_+, S)$, $\pi_t : \Omega \to S$, $\pi_t(w) = w(t)$, $\mathcal{F} = \sigma(\pi_t, t \geqslant 0)$, $\mathcal{F}_t = \sigma(\pi_r, 0 \leqslant r \leqslant t), X_t = \pi_t$.

P_x is often given as a family of solution laws to an SDE (equivalently: as a family of solutions to the corresponding martingale problem), and then the Markov process is normal if and only if P_x has initial condition δ_x. Every Markov process of this type can be modeled on the canonical model.

(2.1.2) is the *Markov property*. An intuitive interpretation, in particular for normal Markov processes, is that $(P_x)_{x\in S}$ models a random memoryless evolution in time on S, and P_x is the law of the evolution trajectories originated from x. Another succinct description of (2.1.2) is:

"The past (of the process X with law P_x) is independent of the future given the present."

The "future" is the event $\{X_{t+s} \in B\}$, the past is $\mathcal{F}_s$, i.e. the information available at time s, and the present is the random state X_s at time s.

Markovian semigroups. A *Markovian transition function* on S is a family of measurable kernels $(p_t)_{t\geqslant 0}$, $p_t : S \times \mathcal{S} \to [0,1]$ such that

(i) $p_t(x, S) = 1, \quad \forall t \geqslant 0, x \in S,$

(ii) $p_t p_s = p_{t+s}$, which means

$$\int_S p_s(y, A)\, p_t(x, dy) = p_{t+s}(x, A), \quad \forall x \in S, A \in \mathcal{S}, t, s \geqslant 0. \tag{2.1.3}$$

(2.1.3) are the *Chapman–Kolmogorov equations.*

Lemma 2.1.4. *Let $(p_t)_{t\geqslant 0}$ be a Markovian transition function and define Λ via*

$$\Lambda(s,t,\zeta) := \int_S p_{t-s}(x, dy)\, \zeta(dx) \in \mathcal{M}_b^+, \text{ i.e. } \Lambda(s,t,\zeta)(A) = \int_S p_{t-s}(x, A)\, \zeta(dx).$$

Then Λ satisfies the flow property (2.1.1).

Proof. Simple exercise. □

In general, the converse is not true. We recall the following well-known results without proofs.

Proposition 2.1.5. *(i) Let $(\Omega, \mathcal{F}, (X_t)_{t\geqslant 0}, (P_x)_{x\in S})$ be a Markov process. Then*

$$(p_t)_{t\geqslant 0}, \; p_t(x, A) := P_x(X_t \in A),$$

is a Markovian transition function. Moreover, for all $f : S^{n+1} \to \mathbb{R}$ bounded and $\mathcal{S}^{n+1}$-measurable and all $0 \leqslant t_0 \leqslant \ldots \leqslant t_n$

$$\begin{aligned} &\mathbb{E}_x\big[f(X_{t_0}, \ldots, X_{t_n})\big] \qquad (2.1.4)\\ &= \int_S \cdots \int_S \Big(\int_S f(x_0, \ldots, x_n)\, p_{t_n - t_{n-1}}(x_{n-1}, dx_n) \Big) p_{t_{n-1}-t_{n-2}}(x_{n-2}, dx_{n-1}) \ldots p_{t_0}(x, dx_0). \end{aligned}$$

(ii) If $(S, \mathcal{S})$ is Polish, then for every Markovian transition function $(p_t)_{t\geqslant 0}$ there is a Markov process with (2.1.4).

For a normal Markov process, the corresponding Markovian transition function satisfies $p_0(x,\cdot) = \delta_x(\cdot)$ for all $x \in S$.

For a Markov process and any $\nu \in \mathcal{P}(S)$, one sets $P_\nu := \int_S P_x \, \nu(dx)$ and sometimes considers $(P_\nu)_{\nu\in\mathcal{P}(S)}$ instead of $(P_x)_{x\in S}$. It is straightforward to check

$$\begin{aligned}&\mathbb{E}_\nu\big[f(X_{t_0},\dots,X_{t_n})\big]\\&= \int_S\int_S \cdots \int_S \Big(\int_S f(x_0,\dots,x_n)\, p_{t_n-t_{n-1}}(x_{n-1},dx_n)\Big) p_{t_{n-1}-t_{n-2}}(x_{n-2},dx_{n-1})\dots\, p_{t_0}(x,dx_0)\,\nu(dx).\end{aligned}$$

The essence of the previous proposition is that the measures P_ν of a Markov process are uniquely determined by its transition function and initial datum. Succinctly written, the above formula reads

$$P_\nu \circ (\pi_{t_0},\dots,\pi_{t_n})^{-1} = \nu p_{t_0} p_{t_1-t_0} p_{t_2-t_1} p_{t_3-t_2} \cdots p_{t_n-t_{n-1}}.$$

Markovian (dual) semigroups and generator. Denote by $\mathcal{S}_b^+$ the set of bounded $\mathcal{S}$-measurable maps $g : S \to \mathbb{R}_+$. For a Markovian transition function $(p_t)_{t\geqslant 0}$, define

$$P_t : \mathcal{S}_b^+ \to \mathcal{S}_b^+, \quad (P_t f)(x) := \int_S f(y)\, p_t(x,dy).$$

$(P_t)_{t\geqslant 0}$ is called the *Markovian semigroup* associated with $(p_t)_{t\geqslant 0}$. P_t is simply the canonical extension from $\{1_A \,|\, A \in \mathcal{S}\}$ to $\mathcal{S}_b^+$ of the map $p_t : 1_A \mapsto [x \mapsto p_t(x,A)]$. Since we only consider normal Markov processes, we always have $P_0 = \mathrm{Id}$.

The *dual semigroup* $(P_t^*)_{t\geqslant 0}$ consists of the maps

$$P_t^* : \mathcal{P}(S) \to \mathcal{P}(S), \quad (P_t^*\nu)(A) := \int_S p_t(x,A)\,\nu(dx),$$

i.e. in particular $P_t^*\delta_x = p_t(x,\cdot)$. By Lemma 2.1.4, $(s,t,\zeta) \mapsto P_{t-s}^*\zeta$ has the flow property in $\mathcal{P}(S)$.

Definition 2.1.6. The *generator* of a normal Markov process with Markovian semigroup $(P_t)_{t\geqslant 0}$ is the linear, typically unbounded, operator $(A, D(A))$,

$$(Af)(x) := \lim_{h\to 0} \frac{P_h f(x) - f(x)}{h},$$

where the domain $D(A)$ consists of those measurable maps $f : S \to \mathbb{R}$ for which the limit on the RHS exists for every $x \in S$, possibly restricted to subspaces such as $C_b(S)$ or $L^p(S;\mu)$ for a measure μ on $\mathcal{S}$.

In other words, Af is the (pointwise in x) right-derivative of $t \mapsto P_t f$ in $t = 0$.

Time-inhomogeneous Markov processes. So far (with the exception of the flow property), in this chapter we considered the *time-homogeneous* setting: the measures P_x in Definition 2.1.1 do not depend on a time parameter s, considered as the "starting time" of the corresponding process, and the corresponding Markovian transition function $(p_t)_{t\geqslant 0}$ is a *one*-parameter family of kernels. Definition 2.1.1

can be extended to the time-inhomogeneous case. For the sake of simplicity, we only consider this generalization in the canonical model as follows.

Let, for $s \geqslant 0$, $\Omega_s = C([s,\infty), S)$, $\mathcal{F}_s = \sigma(\pi^s_r, s \leqslant r)$, $\mathcal{F}_{s,t} = \sigma(\pi^s_r, r \in [s,t])$, where $\pi^s_t : \Omega_s \to S$, $\pi^s_t(w) = w(t)$.

Definition 2.1.7. A family $(P_{s,x})_{s\in\mathbb{R}_+, x\in S}$ of Borel probability measures $P_{s,x} \in \mathcal{P}(\Omega_s)$ is a *time-inhomogeneous Markov process*, if

(i) $x \mapsto P_{s,x}(\Gamma)$ is $\mathcal{S}$-measurable for all $\Gamma \in \mathcal{F}_s$ and $s \geqslant 0$,

(ii) the time-inhomogeneous Markov property holds, i.e.

$$P_{s,x}\big(\pi^s_t \in B|\mathcal{F}_{s,r}\big) = P_{r,\pi^s_r}\big(\pi^r_t \in B\big) \;\; P_{s,x}-\text{a.s.}, \quad \forall 0 \leqslant s \leqslant r \leqslant t, x \in S, B \in S.$$

Again, we restrict to the *normal* case, i.e. $P_{s,x}(\pi^s_s = x) = 1$. The assertions of Proposition 2.1.5 have time-inhomogeneous analogs.

Similar to the time-homogeneous case, a family of measurable probability kernels $(p_{s,t})_{s\leqslant t}$, $p_{s,t} : S \times \mathcal{S} \to [0,1]$ such that $p_{s,t} = p_{s,r}p_{r,t}$ for all $0 \leqslant s \leqslant r \leqslant t$ is called *time-inhomogeneous Markovian transition function.* For a time-inhomogeneous normal Markov process $(P_{s,x})_{s\in\mathbb{R}_+, x\in S}$, we have $p_{s,s}(x,\cdot) = \delta_x(\cdot)$ and the family $(p_{s,t})_{s\leqslant t}$, $p_{s,t}(x,A) := P_{s,x}(\pi^s_t \in A)$, is a time-inhomogeneous Markovian transition function, which extends to the Markovian semigroup

$$P_{s,t} : \mathcal{S}^+_b \to \mathcal{S}^+_b, \; (P_{s,t}f)(x) := \int_S f(y)\, p_{s,t}(x,dy).$$

The dual semigroup is $(P^*_{s,t})_{s\leqslant t}$,

$$P^*_{s,t} : \mathcal{P}(\Omega_s) \to \mathcal{P}(\Omega_s), \quad (P^*_{s,t}\nu)(A) := \int_S p_{s,t}(x,A)\,\nu(dx).$$

A time-inhomogeneous normal Markov process has the generators A_s, defined by

$$(A_s f)(x) := \lim_{h\to 0} \frac{P_{s,s+h}f(x) - f(x)}{h},$$

with domain (which may depend on s) $D(A_s)$, consisting of those functions $f : S \to \mathbb{R}$ for which the limit on the RHS is defined for every x (with the same possible restrictions as in the time-inhomogeneous case).

It is left as a simple exercise to prove: Definition 2.1.7 extends Definition 2.1.1, and the time-inhomogeneous version of Lemma 2.1.4 is true for $(p_{s,t})_{s\leqslant t}, p_{s,t}(x,A) = P_{s,x}(\pi^s_t \in A)$, as well, i.e. $(s,t,\zeta) \mapsto P^*_{s,t}\zeta$ has the flow property in $\mathcal{P}(S)$.

2.2 Fokker–Planck equations and Markov processes

Let $S = \mathbb{R}^d$. We now briefly explore the relation between solutions to Fokker–Planck equations and Markov processes. Consider locally bounded Borel coefficients $a = (a_{ij})_{i,j\leqslant d}, b = (b_i)_{i\leqslant d}$ on $\mathbb{R}_+ \times \mathbb{R}^d$ such that $a(t,x)$ is symmetric and nonnegative definite for all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$, and the associated Fokker–Planck equation (FPE).

For the following result, we omit details on the assumptions on the coefficients.

Proposition 2.2.1. *Suppose $(P_{s,x})_{s\in\mathbb{R}_+,x\in\mathbb{R}^d}$ is a time-inhomogeneous Markov process with generator*

$$(A_s f)(x) = a_{ij}(s,x)\partial^2_{ij} f(x) + b_i(s,x)\partial_i f(x), \quad C_c^\infty(\mathbb{R}^d) \subseteq D(A_s)\ \forall s \in \mathbb{R}_+.$$

Assume sufficient regularity for a and b. Then $t \mapsto \mu_t^{s,x} := P^*_{s,t}\delta_x$ is a weakly continuous probability solution to the FPE

$$\partial_t \mu_t = L^*_{a,b}\mu_t$$

on (s,∞) with initial condition $\mu_s = \delta_x$.

In the context of Markov processes, the FPE is also called *Kolmogorov forward equation.*

Sketch of proof. Without loss of generality let $s = 0$ and set $P^*_{0,t}\delta_x =: \mu_t^x$. For $f \in C_c^\infty(\mathbb{R}^d)$ and $t \geqslant 0$, we have

$$\frac{1}{h}\Big(\int_{\mathbb{R}^d} f\, d\mu^x_{t+h} - \int_{\mathbb{R}^d} d\mu^x_t\Big) = P_{0,t}\Big(\frac{1}{h}\big(P_{t,t+h}f(x) - f(x)\big)\Big). \tag{2.2.1}$$

Thus, for $h \to 0$ we have

$$\lim_{h\to 0}\frac{1}{h}\Big(\int_{\mathbb{R}^d} f\, d\mu^x_{t+h} - \int_{\mathbb{R}^d} d\mu^x_t\Big) = P_{0,t}(A_t f)(x) = \int_{\mathbb{R}^d} L_{a,b} f(t,y)\, d\mu^x_t(y), \tag{2.2.2}$$

i.e. $\frac{d^+}{dt}\int_{\mathbb{R}^d} f\, d\mu^x_t = \int_{\mathbb{R}^d} L_{a,b} f(t,y)\, d\mu^x_t(y)$. It remains to justify that also the left derivative of $\int_{\mathbb{R}^d} f\, d\mu^x_t$ exists and coincides with the RHS dt-a.s. Finally, integrating over any interval $[0,T]$ gives the result. ☐

Now assume the FPE is well-posed among probability solutions with global spatial integrability. More precisely, assume:

(A1) For every $(s,x) \in \mathbb{R}_+ \times \mathbb{R}^d$, there is a unique weakly continuous probability solution $\mu^{s,x} = (\mu_t^{s,x})_{t\geqslant s}$ to (FPE) with initial condition $\mu_s^{s,x} = \delta_x$ such that $a_{ij}, b_i \in L^1\big([0,T]\times\mathbb{R}^d; \mu_t^{s,x}dt\big)$.

Theorem 2.2.2. *Under assumption (A1), there is a unique time-inhomogeneous Markov process $(P_{s,x})_{s\geqslant 0,x\in\mathbb{R}^d}$ with Markovian transition function $p_{s,t}(x,A) = \mu_t^{s,x}(A)$, and $P_{s,x}$ is the law of the unique weak solution to the SDE*

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dB_t, \quad X_s = x, \quad t \geqslant s. \tag{2.2.3}$$

As usual, σ in (2.2.3) is defined by $a = \frac{1}{2}\sigma\sigma^T$, and B is a d-dimensional standard Brownian motion.

Proof. The uniqueness of the Markov process is clear, consider for instance the time-inhomogeneous version of (2.1.4). (A1) allows to apply Theorem 1.3.6 in order to obtain a family $(P_{s,x})_{s\in\mathbb{R}_+,x\in\mathbb{R}^d}$ such that $P_{s,x}$ is the law of a weak solution to (2.2.3) with one-dimensional time marginals $(\mu_t^{s,x})_{t\geqslant s}$. By Proposition 1.3.10, each $P_{s,x}$ is the unique solution law with initial datum (s, x). To prove the Markov property, i.e. Definition 2.1.7 (ii), first note that for $\omega \in \Omega_s$ the path measure $P_{r,\pi^s_r(\omega)}$ is the unique element in $MP_{r,\delta_{\pi^s_r(\omega)}}(a, b)$. Denote by $(Q^{(s,x),r}_\omega)_{\omega\in\Omega_s}$ a r.c.p. of $P_{s,x}$ w.r.t. $\mathcal{F}_{s,r}$. We use the fact that the assertion of Lemma 1.3.2 (ii) remains true for *any* choice $\mathcal{A} = \sigma(\pi^s_u,\ s \leqslant u \leqslant t)$. We choose $\mathcal{A} = \mathcal{F}_{s,r}$ and obtain that the restriction $Q^{(s,x),r}_{\omega,\geqslant r}$ of $Q^{(s,x),r}_\omega$ to $\mathcal{B}(\Omega_r)$ is an element of $MP_{r,\delta_{\pi^s_r(\omega)}}(a, b)$, for $P_{s,x}$-a.e. ω. Thus

$$Q^{(s,x),r}_{\omega,\geqslant r} = P_{r,\pi^s_r(\omega)}, \quad P_{s,x}\text{-a.s.}$$

Since $P_{s,x}(C|\mathcal{F}_{s,r})(\omega) = Q^{(s,x),r}_\omega(C)$, $P_{s,x}$-a.s., for each $C \in \mathcal{B}(\Omega_s)$ (with zero set depending on C), we obtain, letting $C = \{\pi^s_t \in B\}$ for any $t \geqslant r \geqslant s$ and $B \in \mathcal{B}(\mathbb{R}^d)$:

$$P_{s,x}(\pi^s_t \in B|\mathcal{F}_{s,r})(\omega) = Q^{(s,x),r}_\omega(\pi^s_t \in B) = Q^{(s,x),r}_{\omega,\geqslant r}(\pi^r_t \in B) = P_{r,\pi^s_r(\omega)}(\pi^r_t \in B),$$

$P_{s,x}$-a.s. □

Remark 2.2.3. *If a and b are continuous in x and continuous in t locally uniformly in x, then for the generator $(A_s)_{s\geqslant 0}$ of the time-inhomogeneous Markov process of the previous proposition one has $C^2_c(\mathbb{R}^d) \subseteq D(A_s)$ for all $s \geqslant 0$ and*

$$A_s f(x) = a_{ij}(s, x)\partial^2_{ij} f(x) + b_i(s, x)\partial_i f(x), \quad \forall f \in C^2_c(\mathbb{R}^d).$$

The assertion can be generalized to less regular coefficients, but the proof is more involved.

3 Nonlinear Fokker–Planck equations

In this chapter we study *nonlinear* Fokker–Planck equations which, in contrast to linear ones, consist of coefficients depending on the solution itself. This renders the theory of existence and uniqueness of such equations considerably more difficult. On the other hand, the nonlinearity allows to cover large classes of very important nonlinear PDEs. Also the connection to probability theory gains a new component, namely the theory of interacting particle systems. Nonlinear Fokker–Planck equations belong to the most widely used equations in statistical mechanics and physics, see for instance [12]. A standard reference for the nonlinear case is [9] and the references therein. For more recent results, some references will be mentioned throughout the chapter.

3.1 Definition, existence, uniqueness

Let $a_{ij}, b_i : \mathbb{R}_+ \times \mathcal{M} \times \mathbb{R}^d \to \mathbb{R}$, $i, j \leqslant d$, such that $a(t, \zeta, x)$ is pointwise nonnegative definite and symmetric for all $(t, \zeta, x) \in \mathbb{R}_+ \times \mathcal{M}_b^+ \times \mathbb{R}^d$, where $\mathcal{M}$ is a subset of $\mathcal{M}_b^+$ (for instance, the set of measures absolutely continuous w.r.t. dx). We consider *nonlinear Fokker–Planck equations* of type

$$\partial_t \mu_t = \partial^2_{ij}(a_{ij}(t, \mu_t, x)\mu_t) - \partial_i(b_i(t, \mu_t, x)\mu_t), \quad t \geqslant 0 \tag{3.1.1}$$

(simply considered as "the NLFPE" in the sequel). For $\mu \in \mathcal{M}$, we set, for $\varphi \in C^2(\mathbb{R}^d)$,

$$L_{a,b,\mu}\varphi(t, x) = a_{ij}(t, \mu, x)\partial^2_{ij}\varphi(x) + b_i(t, \mu, x)\partial_i\varphi(x).$$

As before, in general solutions are measure-valued curves $t \mapsto \mu_t$. One can consider cases where $a(t)$ and $b(t)$ depend on $\big((\mu_t)_{t\geqslant 0}, x\big)$ instead of (μ_t, x); also the case of locally finite signed measures can be considered. We will, however, restrict ourselves to the case presented above.

Examples. *Global dependence.* The prototype of nonlinear coefficients with global measure dependence is

$$b(t, \mu, x) = \int_{\mathbb{R}^d} K(t, x, y)d\mu(y), \quad K : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d, \tag{3.1.2}$$

and likewise for a. Specifically, a common case is $K(t, x, y) = \nabla k(t, x - y)$ for $k : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$. k is called a *potential.*

Local dependence. A very important class is given by coefficients of type

$$a_{ij}(t, \mu, x) = \tilde{a}_{ij}\Big(t, \frac{d\mu}{dx}(x), x\Big), \quad b_i(t, \mu, x) = \tilde{b}_i\Big(t, \frac{d\mu}{dx}(x), x\Big), \quad i, j \leqslant d, \tag{3.1.3}$$

where $\tilde{a}_{ij}, \tilde{b}_i : \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ and $\frac{d\mu}{dx}(x)$ denotes the density of μ w.r.t. dx, evaluated at x. Without further mentioning, we always consider the version of $\frac{d\mu}{dx}$ which is 0 on those $x \in \mathbb{R}^d$ for which $\lim_{r\to 0} dx(B_r(0))^{-1}\mu(B_r(x))$ does not exist in $\mathbb{R}$. By Lebesgue's differentiation theorem, the set of such x is a dx-zero set. Then $(\mu, y) \mapsto \frac{d\mu}{dx}(y)$ is $\mathcal{B}(\mathcal{M}^+_{b,\ll}) \otimes \mathcal{B}(\mathbb{R}^d)$-measurable by [13, Sect.4.2.], where $\mathcal{M}^+_{b,\ll}$ denotes the subset of $\mathcal{M}^+_b$ of measures absolutely continuous w.r.t. dx, equipped with the topology of weak convergence of measures. The coefficients are defined on $\mathbb{R}_+ \times \mathcal{M}^+_{b,\ll} \times \mathbb{R}^d$. This case is often called *Nemytskii-case*, and a and b as in (3.1.3) are *of Nemytskii-type.*

In the Nemytskii-case, the NLFPE is often posed in *density form*

$$\partial_t u(t,x) = \partial^2_{ij}\big(\tilde{a}_{ij}(t, u(t,x), x)u(t)\big) - \operatorname{div}\big(\tilde{b}(t, u(t,x), x)u(t)\big),$$

i.e. in comparison with the general measure-valued formulation $\mu_t = u(t,x)dx$ this equation is a (function-valued) PDE for the density $(t,x) \mapsto u(t,x)$.

Note that even if $r \mapsto \tilde{a}_{ij}(t,r,x)$ is continuous for fixed (t,x), the map $\mu \mapsto a_{ij}(t,\mu,x) = \tilde{a}_{ij}(t, \frac{d\mu}{dx}(x), x)$ is not continuous w.r.t. the weak or vague topology (or, as a matter of fact, any other reasonable topology on $\mathcal{M}^+_b$), since $\mu \mapsto \frac{d\mu}{dx}(x)$ is not continuous between any of these topologies and $\mathbb{R}$. Hence, in the Nemytskii-case, one faces irregular coefficients.

We give a few important examples of NLFPEs of Nemytskii type.

(i) The classical Porous Media Equation (PME)

$$\partial_t u(t) = \Delta(u(t)^m), \quad (t,x) \in (0,\infty) \times \mathbb{R}^d,$$

$m > 0$, for the class of nonnegative solutions $u \geqslant 0$ can be written as

$$\partial_t u(t) = \partial^2_{ij}(a_{ij}(u(t,x))u(t))$$

with $a_{ij}(r) = \delta_{ij} r^{m-1}$. Hence, the PME is a Nemytskii-type NLFPE in density form. The cases $m > 1$ and $m < 1$ are called *slow* and *fast diffusion* case, respectively ($m = 1$ gives the heat equation). The reason for these names is that if $u(x) \to 0$, for $m > 1$ and $m < 1$ the diffusion coefficient u^{m-1} degenerates and explodes, respectively.

(ii) More generally, consider the *generalized PME*

$$\partial_t u(t) = \Delta\beta(u) - \operatorname{div}\big(DB(u(t))u(t)\big), \quad (t,x) \in (0,\infty) \times \mathbb{R}^d,$$

where $\beta \in C^1(\mathbb{R})$, $\beta(0) = 0$, $D : \mathbb{R}^d \to \mathbb{R}^d$, $B : \mathbb{R}_+ \to \mathbb{R}$, which can be written in NLFPE density form as

$$\partial_t u(t) = \partial^2_{ij}(a_{ij}(u(t,x))u(t)) - \operatorname{div}\big(b(x, u(t,x))u(t)\big),$$

with $a_{ij}(r) = \frac{\beta(r)}{r}$, $b(x,r) = D(x)B(r)$, where $\frac{\beta(0)}{0} := \beta'(0)$.

(iii) Consider the p-Laplace equation

$$\partial_t u(t) = \operatorname{div}(|\nabla u(t)|^{p-2}\nabla u(t)), \quad (t,x) \in (0,\infty) \times \mathbb{R}^d.$$

For a suitable subclass of solutions, it is equivalent to

$$\partial_t u(t) = \Delta\big(|\nabla u(t)|^{p-2}u\big) - \operatorname{div}\big(\nabla(|\nabla u(t)|^{p-2})u(t)\big),$$

which is a Nemytskii-type NLFPE in density form with coefficients

$$a_{ij}(u,x) = \delta_{ij}|\nabla u|^{p-2}(x), \quad b_i(u,x) = \partial_i|\nabla u|^{p-2}(x).$$

(iv) The $2D$ Navier–Stokes equations in vorticity form can written as a Nemytskii-type NLFPE in density form.

The definition of solutions in the nonlinear case is analogous to Definitions 1.2.1, 1.2.4, 1.2.5. We explicitly only state the following notion.

Definition 3.1.1. A Borel curve $(\mu_t)_{t>0} \subseteq \mathcal{M}$ solves (3.1.1) with initial value $\nu \in \mathcal{M}_b^+$, if $(t,x) \mapsto a_{ij}(t,\mu_t,x), b_i(t,\mu,x)$ are Borel maps in $L^1_{\text{loc}}((0,\infty)\times\mathbb{R}^d;\mu_t dt)$, and for every $\varphi \in C_c^\infty(\mathbb{R}^d)$ there is a set $J_\varphi \subseteq (0,\infty)$ of full dt-measure such that for all $t \in J_\varphi$

$$\int_{\mathbb{R}^d} \varphi\, d\mu_t = \int_{\mathbb{R}^d} \varphi\, d\nu + \lim_{\tau\to 0+}\int_\tau^t \int_{\mathbb{R}^d} L_{a,b,\mu_s}\varphi\, d\mu_s ds. \tag{3.1.4}$$

Compared to the linear case, here we omit a zero-order coefficient c (in general also dependent on the solution). A bit more generally, one may require $\mu_t \in \mathcal{M}$ only dt-a.s. and the existence of a Borel curve dt-version $\tilde{\mu}$ of μ such that $\tilde{\mu}_t \in \mathcal{M}$ for all $t > 0$ such that

$$\int_{\mathbb{R}^d} \varphi\, d\mu_t = \int_{\mathbb{R}^d} \varphi\, d\nu + \lim_{\tau\to 0+}\int_\tau^t \int_{\mathbb{R}^d} L_{a,b,\tilde{\mu}_s}\varphi\, d\mu_s ds.$$

Remark 3.1.2. *One could require the coefficients to be $\mathcal{B}(\mathbb{R}_+)\otimes\mathcal{B}(\mathcal{M}_b^+)\otimes\mathcal{B}(\mathbb{R}^d)$-measurable (where $\mathcal{B}(\mathcal{M}_b^+)$ denotes the Borel σ-algebra w.r.t. either the weak or vague topology). Then it follows that $(t,x)\mapsto a(t,\mu_t,x)$ and $(t,x)\mapsto b(t,\mu_t,x)$ are product measurable on $\mathbb{R}_+\times\mathbb{R}^d$ for every Borel curve $(\mu_t)_{t\geqslant 0}$ (Exercise 6.1). We follow a slightly different approach by not requiring such a property, but instead require a solution $(\mu_t)_{t>0}$ to render $(t,x)\mapsto a(t,\mu_t,x)$ and $(t,x)\mapsto b(t,\mu_t,x)$ measurable. Conceptually, the latter is a weaker assumptions on the coefficients without narrowing the notion of solution.*

Linearized equations. A very important object related to the nonlinear FPE is the family of associated *linearized FPEs*, obtained as follows: For any Borel curve $t\mapsto\mu_t\in\mathcal{M}_b^+$, consider the *linear* FPE

$$\partial_t\nu_t = \partial^2_{ij}\big(a_{ij}(\mu_t)\nu_t\big) - \partial_i\big(b_i(\mu_t)\nu_t\big), \quad t>0, \tag{μ-ℓFPE}$$

where by $a_{ij}(\mu_t)$ and $b_i(\mu_t)$ we abbreviate the maps $x\mapsto a_{ij}(t,\mu_t,x)$ and $x\mapsto b_i(t,\mu_t,x)$, respectively. For given $\mu = (\mu_t)_{t>0}$, we denote this linear equation by (μ-ℓFPE).

Remark 3.1.3. *(i) A solution $(\mu_t)_{t>0}$ to the nonlinear FPE in the sense of Definition 3.1.1 with initial datum ν also solves (μ-ℓFPE), i.e. "any nonlinear FPE-solution also solves its own linearized FPE".*

(ii) The coefficients of the linearized FPEs are time-dependent, even if the nonlinear coefficients itself are time-independent.

Taking into account (i) of the previous remark, many results for solutions to linear equations can be proven for solutions to nonlinear equations as well. For instance, we have

Lemma 3.1.4. *The results of Lemma 1.2.6 hold analogously for solutions to (3.1.1).*

Since (3.1.4) is invariant under changing $(L_{a,b,\mu_t})_{t>0}$ to $(L_{a,b,\tilde{\mu}_t})_{t>0}$ for a Borel curve dt-version $\tilde{\mu}$ of μ, the consideration of the linearized equations yields the following analog of Proposition 1.3.8:

Lemma 3.1.5. *Let $\mu = (\mu_t)_{t>0} \subseteq \mathcal{M}$ be a solution to the NLFPE with initial value $\nu \in \mathcal{M}_b^+$ such that $\operatorname{ess\,sup}_{t>0} \mu_t(\mathbb{R}^d) < \infty$ and*

$$[(t,x) \mapsto a_{ij}(t,\mu_t,x)],\ [(t,x) \mapsto b_i(t,\mu_t,x)] \in L^1_{\text{loc}}([0,\infty) \times \mathbb{R}^d; \mu_t dt). \tag{3.1.5}$$

Then there is a unique vaguely continuous dt-version $\tilde{\mu}$ of μ, and $\tilde{\mu}$ also solves the NLFPE with initial datum ν.

If in addition the maps from (3.1.5) are in $L^1([0,T] \times \mathbb{R}^d; \mu_t dt)$ for all $T > 0$, then $\tilde{\mu}(\mathbb{R}^d) = \nu(\mathbb{R}^d)$ for all $t > 0$ and $t \mapsto \tilde{\mu}_t$ is weakly continuous.

Proof. Consider $(\mu_t)_{t>0}$ as a solution to (μ-ℓFPE). By (3.1.5), Proposition 1.3.8 applies and yields a unique vaguely continuous version $(\tilde{\mu}_t)_{t\geqslant 0}$ with $\tilde{\mu}_0 = \nu$, solving (μ-ℓFPE). Hence $(\tilde{\mu}_t)_{t\geqslant 0}$ solves the NLFPE. The second part follows from the second part of Proposition 1.3.8. □

3.2 An existence result via a fixed point argument

The content of this subsection is taken from [14].

Let $T > 0$ and $\mathcal{M}_b([0,T]\times\mathbb{R}^d)$ be the linear space of *signed* measures with finite total variation. For Borel curves $(\mu_t)_{t\in[0,T]} \subseteq \mathcal{M}_b(\mathbb{R}^d)$ such that $\operatorname{ess\,sup}_{t\in[0,T]} |\mu_t|(\mathbb{R}^d) < \infty$, we identify $(\mu_t)_{t\in[0,T]}$ with $\mu = \mu_t dt \in \mathcal{M}_b([0,T] \times \mathbb{R}^d)$.

Recall that $\mathcal{M}_b([0,T] \times \mathbb{R}^d)$ is a normed space with the Kantorovich-Rubinstein norm

$$||\mu|| := \sup_{f\in \text{Lip}_1} \int f d\mu,$$

where Lip_1 denotes the set of Lipschitz functions from $\mathbb{R}^d$ to $\mathbb{R}$ with Lipschitz constant less or equal to 1 which are also uniformly bounded by 1. Moreover, the topology generated by this norm on the nonnegative halfspace $\mathcal{M}_b^+([0,T] \times \mathbb{R}^d)$ is the topology of weak convergence of measures.

We will use the following fixed point theorem by Schauder.

Theorem 3.2.1. *Let X be a normed space, $K \subseteq X$ a compact convex subset, and $F : K \to K$ continuous. Then there is $k \in K$ with $F(k) = k$.*

For $V : \mathbb{R}^d \to \mathbb{R}_+$, $T_0 \leqslant T$ and $g \in C^+([0,T])$ (the space of continuous maps from $[0,T]$ to $\mathbb{R}_+$), define $M_{T_0,g}(V)$ as the set of nonnegative measures $\mu = (\mu_t)_{t\in[0,T_0]}$ in $\mathcal{M}_b([0,T_0] \times \mathbb{R}^d))$ such that

$$\int_{\mathbb{R}^d} V \, d\mu_t \leqslant g(t), \quad \forall t \in [0,T_0].$$

Let a_{ij}, b_i be defined on $[0,T] \times \mathcal{M}_b^+ \times \mathbb{R}^d$. We will prove an existence result for the NLFPE (3.1.1) under the following assumptions.

(H1). There is $V \in C^2(\mathbb{R}^d, \mathbb{R}_+)$, $V > 0$, $\lim_{|x|\to\infty} V(x) \to \infty$, and maps $\Lambda_1, \Lambda_2 : C^+([0,T]) \to C^+([0,T])$ such that for all $T_0 \in (0,T]$ and $g \in C^+([0,T])$: For all $(t,\nu,x) \in [0,T_0] \times M_{T_0,g}(V) \times \mathbb{R}^d$

$$L_{a,b,\nu} V(t,x) \leqslant \Lambda_1[g](t) + \Lambda_2[g](t) V(x).$$

From now on, we fix V (but not T_0 or g) and write $M_{T_0,g}$ instead of $M_{T_0,g}(V)$.

Definition 3.2.2. We say a sequence $\mu^n = (\mu_t^n)_{t\in[0,T_0]}$ in $M_{T_0,g}$ is V-*convergent* to $\mu = (\mu_t)_{t\in[0,T_0]}$ in $M_{T_0,g}$ if

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} f \, d\mu_t^n = \int_{\mathbb{R}^d} f \, d\mu_t, \quad \forall t \in [0,T_0]$$

for all $f \in C(\mathbb{R}^d)$ such that $\lim_{|x|\to\infty} \frac{f(x)}{V(x)} = 0$. In particular, V-convergence implies weak convergence.

(H2). For all $T_0 \in (0,T]$, $g \in C^+([0,T])$, $\nu \in M_{T_0,g}$, the maps

$$t \mapsto a_{ij}(t,\nu_t,x), \quad t \mapsto b_i(t,\nu_t,x)$$

are Borel on $[0,T_0]$ for each fixed x, locally bounded in x uniformly in $(t,\nu) \in [0,T_0] \times M_{T_0,g}$, and x-locally equicontinuous in (t,ν). Moreover, if μ^n V-converges to μ in $M_{T_0,g}$, then

$$a_{ij}(t,\mu_t^n,x) \to a_{ij}(t,\mu_t,x), \quad b_i(t,\mu_t^n,x) \to b_i(t,\mu_t,x), \quad \forall (t,x) \in [0,T_0] \times \mathbb{R}^d.$$

(H3). For all $T_0 \in (0,T]$, $g \in C^+([0,T])$ and $\nu \in M_{T_0,g}$, $a(t,\nu_t,x)$ is symmetric and nonnegative definite for all $(t,x) \in [0,T_0] \times \mathbb{R}^d$.

Theorem 3.2.3. *Suppose (H1)-(H3) are satisfied, and let $\mu_0 \in \mathcal{P}$ such that $V \in L^1(\mathbb{R}^d, \mu_0)$.*

(i) There is $T_0 \leqslant T$ such that the NLFPE has a weakly continuous probability solution on $[0,T_0]$ with initial datum μ_0.

(ii) If Λ_1, Λ_2 from (H1) are constant from $C^+([0,T])$ to $C^+([0,T])$, then $T_0 = T$.

In both cases, this solution $(\mu_t)_{t\in[0,T_0]}$ *satisfies*

$$\sup_{t\in[0,T_0]}\int_{\mathbb{R}^d} V\,d\mu_t < \infty \tag{3.2.1}$$

and

$$[(t,x)\mapsto a_{ij}(t,\mu_t,x)],[(t,x)\mapsto b_i(t,\mu_t,x)]\in L^1_{\mathrm{loc}}([0,T_0]\times\mathbb{R}^d;\mu_t dt).$$

The proof proceeds via several steps:

(a) Case of a nondegenerate and sufficiently smooth diffusion matrix a;

(b) Degenerate and sufficiently smooth case;

(c) General case (i.e. only (H1)-(H3) are assumed).

Due to time constraints, we only give details regarding (a). The remaining parts can be found in [14].

3.2.1 Proof of Theorem 3.2.3.

For part (a), we replace (H3) by the following stronger assumption.

(H3'). (H3) holds, and in addition for each $T_0\in(0,T]$, $g\in C^+([0,T])$, $\nu\in M_{T_0,g}$ and compact $U\subseteq\mathbb{R}^d$, there is $\lambda=\lambda(\nu,U)>0$ such that $a(t,\nu_t,x)>0$ for all $(t,x)\in[0,T_0]\times\mathbb{R}^d$ and

$$|a(t,\nu_t,x)-a(t,\nu_t,y)|\leqslant\lambda(\nu,U)|x-y|,\quad\forall x,y\in U,t\in[0,T_0].$$

Moreover, assume there are finite constants $C_i=C_i(\nu)$ such that

$$|\sqrt{a(t,\nu_t,x)}\nabla V(x)|\leqslant C_1+C_2V(x),\quad\forall(t,x)\in[0,T_0]\times\mathbb{R}^d.$$

Let $T_0\leqslant T$, $g\in C^+([0,T])$ and $\nu\in M_{T_0,g}$. Then by [22, Thm.3.1], assumptions (H1),(H2),(H3') imply the existence of a unique weakly continuous probability solution $\zeta=\zeta(\nu)$ to (ν-ℓFPE) on $[0,T_0]$ with initial datum μ_0 such that

$$[(t,x)\mapsto a_{ij}(t,\nu_t,x)],[(t,x)\mapsto b_i(t,\nu_t,x)]\in L^1_{\mathrm{loc}}([0,T_0]\times\mathbb{R}^d;\zeta_t dt).$$

Hence we may consider the well-defined map

$$Q:M_{T_0,g}\to\mathcal{M}_b([0,T_0]\times\mathbb{R}^d),\quad Q(\nu):=\zeta(\nu).$$

Note that Q depends on T_0,g (and V).

Remark 3.2.4. *Suppose there is* $T_0\leqslant T,g\in C^+([0,T])$ *such that*

(I) $M_{T_0,g}\subseteq\mathcal{M}_b([0,T_0]\times\mathbb{R}^d)$ *is convex and compact;*

(II) Q *is continuous on* $M_{T_0,g}$ *and* $Q(M_{T_0,g})\subseteq M_{T_0,g}$.

Then, by Schauder's fixed point theorem, there is a fixed point of Q in $M_{T_0,g}$. This fixed point is a weakly continuous solution to the NLFPE with initial datum μ_0 and satisfies the final assertion of Theorem 3.2.3.

We will prove (I)+(II) for a subset $N_{T_0,g} \subset M_{T_0,g}$, which is obviously sufficient.

Indeed, define $N_{T_0,g}$ as the subset of $M_{T_0,g}$ consisting of those $(\mu_t)_{t\in[0,T_0]}$ such that for all $\varphi \in C_c^\infty(\mathbb{R}^d)$

$$\left| \int_{\mathbb{R}^d} \varphi \, d\mu_t - \int_{\mathbb{R}^d} \varphi \, d\mu_s \right| \leqslant \Lambda(T_0, g, \varphi)|t-s|, \quad \forall t,s \in [0,T_0], \tag{3.2.2}$$

where $\Lambda(T_0, g, \varphi) := \sup_{(t,\nu,x)\in[0,T_0]\times M_{T_0,g}\times\mathbb{R}^d}\{|L_{a,b,\nu}\varphi(t,x)|\}$. This value is finite due to (H2).

Lemma 3.2.5. *Every sequence $\mu^n = (\mu^n_t)_{t\in[0,T_0]}$ in $N_{T_0,g}$ has a weakly convergent subsequence (μ^{n_k}) with limit $\mu \in N_{T_0,g}$. Moreover, for each $t \in [0,T_0]$, $\mu^{n_k}_t$ weakly converges to μ_t.*

The first part of the assertion just means that $N_{T_0,g} \subseteq \mathcal{M}_b([0,T_0]\times\mathbb{R}^d)$ is sequentially compact.

Proof. Exercise 8.1. □

Corollary 3.2.6. *$N_{T_0,g} \subseteq \mathcal{M}_b([0,T_0]\times\mathbb{R}^d)$ is convex and compact.*

Proof. For convexity, note that (3.2.2) is stable w.r.t. convex combinations and that $M_{T_0,g}$ is convex by definition. Since the topology of $\mathcal{M}_b([0,T]\times\mathbb{R}^d)$ is induced by a norm, a subset $M \subseteq \mathcal{M}_b([0,T_0]\times\mathbb{R}^d)$ is compact if and only if it is sequentially compact. The latter holds for $N_{T_0,g}$ by the previous lemma. □

Lemma 3.2.7. *If a sequence μ^n weakly converges to μ in $N_{T_0,g}$, then μ^n V-converges to μ.*

Proof. First note that μ^n_t weakly converges to μ_t for all $t \in [0,T_0]$. Indeed, let $t \in [0,T_0]$. By Lemma 3.2.5, each subsequence μ^{n_l} has a further subsequence $\mu^{n_{l_k}}$ such that $\mu^{n_{l_k}}_t$ weakly converges to μ_t. Hence, μ^n_t weakly converges to μ_t.

Also note: Since g is bounded on $[0,T]$ and there is $\sigma > 0$ such that $V(x) > \sigma$ for all $x \in \mathbb{R}^d$, it follows that $\sup_{\mu\in M_{T_0,g}}\{\mu_t(\mathbb{R}^d), t \in [0,T_0]\} \leqslant c_0 < \infty$, with $c_0 := |g|_\infty[\inf_{x\in\mathbb{R}^d} V(x)]^{-1}$

Let now $f \in C(\mathbb{R}^d)$ such that $\lim_{|x|\to\infty}\frac{f(x)}{V(x)} = 0$. Set $h(x) := \frac{f(x)}{V(x)}$, i.e. $h \in C_0(\mathbb{R}^d)$ (the set of continuous functions vanishing at infinity). Hence, for $\varepsilon > 0$, there is $\psi \in C_c(\mathbb{R}^d)$ with $|h-\psi|_\infty < \varepsilon$. Then

$$\begin{aligned}\left| \int_{\mathbb{R}^d} f\, d\mu^n_t - \int_{\mathbb{R}^d} f\, d\mu_t \right| &= \left| \int_{\mathbb{R}^d} hV\, d\mu^n_t - \int_{\mathbb{R}^d} hV\, d\mu_t \right| \\ &\leqslant \left| \int_{\mathbb{R}^d} \psi V\, d\mu^n_t - \int_{\mathbb{R}^d} \psi V\, d\mu_t \right| + 2\varepsilon|g|_\infty.\end{aligned}$$

Since $\psi V \in C_b(\mathbb{R}^d)$, $\mu^n_t \to \mu_t$ weakly and $\varepsilon > 0$ was arbitrary, the claim follows. □

Lemma 3.2.8. *If $Q(N_{T_0,g}) \subseteq N_{T_0,g}$ for some $T_0 \leqslant T$, $g \in C^+([0,T])$, then Q is continuous on $N_{T_0,g}$.*

Proof. Since the topology on $N_{T_0,g} \subseteq \mathcal{M}_b([0,T_0]\times\mathbb{R}^d)$ is induced by a norm, it suffices to prove sequential continuity. So, let $\mu^n, \mu \in N_{T_0,g}$ such that $\mu^n \to \mu$ weakly, and set $\zeta^n := Q(\mu^n)$. Since $\zeta^n \in N_{T_0,g}$, for any subsequence of ζ^n, Lemma 3.2.5 yields a further subsequence $\zeta^{n_{kl}}$ with limit $\zeta \in N_{T_0,g}$. A priori, this limit depends on $\{n_{kl}\}$, but we will show $\zeta = Q(\mu)$, which then implies that ζ^n weakly converges to $Q(\mu)$. We now denote $\zeta^{n_{kl}}$ by ζ^n. Lemma 3.2.5 also implies the weak convergence $\zeta^n_t \to \zeta_t$ for all $t \in [0,T_0]$. Moreover, Lemma 3.2.7 implies V-convergence of μ^n to μ.

Let $t \in [0,T_0]$. By (H2), the maps $x \mapsto a_{ij}(t,\mu^n_t,x)$ converge pointwise to $a_{ij}(t,\mu_t,x)$, are locally in x uniformly in n bounded and locally in x uniformly in n equicontinuous. Hence, by the Arzela-Ascoli theorem, they converge locally uniformly. The same is true for the convergence of $b_i(t,\mu^n_t,x)$ to $b_i(t,\mu_t,x)$.

Next we show $\zeta = Q(\mu)$. Let $\varphi \in C_c^\infty(\mathbb{R}^d)$. Then, since $Q(\mu^n) = \zeta^n$, we have

$$\int_{\mathbb{R}^d} \varphi \, d\zeta^n_t - \int_{\mathbb{R}^d} \varphi \, d\mu_0 = \int_0^t \int_{\mathbb{R}^d} L_{a,b,\mu^n}\varphi \, d\zeta^n_s ds, \quad t \in [0,T_0].$$

We have

$$\int_{\mathbb{R}^d} L_{a,b,\mu^n}\varphi \, d\zeta^n_s = \int_{\mathbb{R}^d} \left(L_{a,b,\mu^n}\varphi - L_{a,b,\mu}\varphi\right) d\zeta^n_s + \int_{\mathbb{R}^d} L_{a,b,\mu}\varphi \, d\zeta^n_s,$$

where the first summand on the RHS converges to 0 as $n \to \infty$ and the second one converges to $\int_{\mathbb{R}^d} L_{a,b,\mu}\varphi \, d\zeta_s$. Since $|L_{a,b,\mu^n}\varphi(t,x)| \leqslant \Lambda(T_0,g,\varphi) < \infty$, we can apply Lebesgue's dominated convergence theorem to obtain

$$\lim_{n\to\infty} \int_0^t \int_{\mathbb{R}^d} L_{a,b,\mu^n}\varphi \, d\zeta^n_s ds = \int_0^t \int_{\mathbb{R}^d} L_{a,b,\mu}\varphi \, d\zeta ds.$$

Now the weak convergence $\zeta^n_t \to \zeta_t$ for all $t \in [0,T_0]$ yields the claim. □

The next lemma is the final preliminary step for finding suitable T_0 and g to apply the previous lemma.

Lemma 3.2.9. *Suppose $\nu \in N_{T_0,g}$, $\zeta = Q(\nu)$. Then, for all $t \in [0,T_0]$,*

$$\int_{\mathbb{R}^d} V \, d\zeta_t \leqslant S[g](t) + R[g](t) \int_{\mathbb{R}^d} V \, d\mu_0,$$

where

$$R[g](t) := \exp\left(\int_0^t \Lambda_2[g](s) ds\right), \quad S[g](t) := R[g](t) \int_0^t \Lambda_1[g](s) ds.$$

Proof. Let ν be as in the assertion and $\eta_m \in C^\infty(\mathbb{R}_+)$ such that $0 \leqslant \eta_m'(x) \leqslant 1$, $\eta_m'' \leqslant 0, \eta_m(x) = x$ if $x \leqslant m-1$, $\eta_m(x) = m$ if $x > m$. Recall that by definition ζ satisfies for all $\varphi \in C_c^2(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d} \varphi \, d\zeta_t - \int_{R^d} \varphi \, d\mu_0 = \int_0^t \int_{\mathbb{R}^d} L_{a,b,\nu}\varphi \, d\zeta_s ds, \quad \forall t \in [0,T_0].$$

Choose $\varphi(x) := \eta_m \circ V(x) - m$, and note

$$L_{a,b,\nu}\varphi(t,x) = \eta_m'(V(x))L_{a,b,\nu}V(t,x) + \eta_m''(V(x))a(t,\nu_t,x)\nabla V(x)\cdot\nabla V(x).$$

Therefore

$$\int_{|V|\leqslant m-1} V\,d\zeta_t \leqslant \int_{\mathbb{R}^d} \eta_m(V)\,d\zeta_t \leqslant \int_{\mathbb{R}^d} V\,d\mu_0 + \int_0^t \int_{|V|\leqslant m} \eta_m'(V(x))L_{a,b,\nu}V(s,x)\,d\zeta_s(x)ds.$$

Since $\eta_m' \leqslant 1$ and since (H1) entails

$$\int_0^t \int_{|V|\leqslant m} L_{a,b,\nu}V(s,x)\,d\zeta_s ds \leqslant \int_0^t \Big(\Lambda_1[g](s) + \Lambda_2[g](s)\int_{|V|\leqslant m} V\,d\zeta_s\Big)ds,$$

we arrive, by letting $m \to \infty$, at

$$\int_{\mathbb{R}^d} V\,d\zeta_t \leqslant \int_{\mathbb{R}^d} V\,d\mu_0 + \int_0^t \Big(\Lambda_1[g](s) + \Lambda_2[g](s)\int_{\mathbb{R}^d} V\,d\zeta_s\Big)ds.$$

Now Gronwall's lemma yields

$$\int_{\mathbb{R}^d} V\,d\zeta_t \leqslant \Big[\int_{\mathbb{R}^d} V\,d\mu_0 + \int_0^t \Lambda_1[g](s)ds\Big]\exp\Big(\int_0^t \Lambda_2[g](s)ds\Big),$$

which is the claim. □

Finally, for both parts (i) and (ii) of the theorem, we find $T_0 \leqslant T$ and $g \in C^+([0,T])$ such that $Q(N_{T_0,g}) \subseteq N_{T_0,g}$:

Corollary 3.2.10. *There is $T_0 \leqslant T$ and $g \in C^+([0,T])$ constant and strictly positive such that $Q(N_{T_0,g}) \subseteq N_{T_0,g}$. Moreover, if the mappings Λ_1 and Λ_2 are constant, then one can choose $T_0 = T$.*

Proof. By the previous lemma, we have for any $\nu \in N_{T_0,g}$, $\zeta = Q(\nu)$, $T_0 \leqslant T, g \in C^+([0,T])$

$$\int_{\mathbb{R}^d} V\,d\zeta_t \leqslant S[g](t) + R[g](t)\int_{\mathbb{R}^d} V\,d\mu_0.$$

For any choice of g, note that $S[g](t) \to 0$ and $R[g](t) \to 1$ as $t \to 0$. Set $g := 2\int_{\mathbb{R}^d} V\,d\mu_0 + 1$ and choose $T_0 = T_0(g)$ such that $S[g](t) \leqslant 1$ and $R[g](t) \leqslant 2$ for all $t \in [0,T_0]$. Then

$$\int_{\mathbb{R}^d} V\,d\zeta_t \leqslant g(t), \quad \forall t \in [0,T_0].$$

So, $Q(N_{T_0,g}) \subseteq M_{T_0,g}$, and the claim follows, since (3.2.2) is fulfilled for every element in the range of Q.

For the second part, first note that S and R do not depend on g, since they are functions of Λ_1, Λ_2, which are now independent of g by assumption. Set

$$g(t) := \max_{r\in[0,T]}\Big(S(r) + R(r)\int_{\mathbb{R}^d} V\,d\mu_0\Big), \quad \forall t \in [0,T].$$

Then, obviously $\int_{\mathbb{R}^d} V\,d\zeta_t \leqslant g(t)$ for all $t \in [0,T]$. Hence, as above, we conclude $Q(N_{T,g}) \subseteq N_{T,g}$. □

We can now complete the proof of Theorem 3.2.3 as follows:

For (i) and (ii) of the assertion, consider T_0 and g as in the previous corollary, respectively, such that $Q(N_{T_0,g}) \subseteq N_{T_0,g}$. By Lemma 3.2.8 Q is continuous on $N_{T_0,g}$. Since Corollary 3.2.6 implies that $N_{T_0,g}$ is a convex and compact subset of the normed space $\mathcal{M}_b([0,T_0]\times\mathbb{R}^d)$, we may apply Schauder's fixed point theorem to obtain a fixed point $\mu = (\mu_t)_{t\in[0,T_0]} \in N_{T_0,g}$ of Q, i.e. $Q(\mu) = \mu$. As explained in Remark 3.2.4, μ is the solution from the assertion. $\mu \in N_{T_0,g}$ yields (3.2.1). □

3.3 McKean–Vlasov SDEs

Consider coefficients a_{ij}, b_i as in the beginning of Subsection 3.1, let $\sigma : \mathbb{R}_+ \times \mathcal{P}\times\mathbb{R}^d \to \mathbb{R}^{d\times d}$ be such that $\frac{1}{2}\sigma\sigma^T = a$ pointwise, and let B denote a standard d-dimensional Brownian motion.

In this section, we consider the following SDEs related to the measure-dependent coefficients b and σ

$$dX_t = b(t, X_t, \mathcal{L}_{X_t})dt + \sigma(t, X_t, \mathcal{L}_{X_t})dB_t, \quad t \geqslant 0. \tag{3.3.1}$$

Such equations are called *McKean–Vlasov SDEs* or *distribution-dependent SDEs*, short *DDSDEs*. In contrast to the classical "linear" case, here the drift vector and diffusion matrix depend not only on the current position, but also on the distribution of the solution. For a partial literature overview on DDSDEs, see Exercise sheet 8.

The following definition is completely analogous to the non-distribution dependent case.

Definition 3.3.1. A *weak solution* to (3.3.1) is a triple, consisting of a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$, a d-dimensional standard $(\mathcal{F}_t)$-Brownian motion and an $(\mathcal{F}_t)$-adapted $\mathbb{R}^d$-valued stochastic process $X = (X_t)_{t\geqslant 0}$ on Ω such that $(t,\omega)\mapsto b(t, X_t(\omega), \mathcal{L}_{X_t})$ and $(t,\omega)\mapsto \sigma(t, X_t(\omega), \mathcal{L}_{X_t})$ are $\mathcal{B}(\mathbb{R}_+)\otimes\mathcal{F}$-measurable,

$$\mathbb{E}\Big[\int_0^T |b(t, X_t, \mathcal{L}_{X_t})| + |\sigma(t, X_t, \mathcal{L}_{X_t})|^2\, dt\Big] < \infty, \quad \forall T > 0,$$

and $\mathbb{P}$-a.s.

$$X_t = X_0 + \int_0^t b(s, X_s, \mathcal{L}_{X_s})ds + \int_0^t \sigma(s, X_s, \mathcal{L}_{X_s})dB_s, \quad \forall t > 0.$$

As in the non-distribution dependent case, we call $\mathcal{L}_{X_0}$ the *initial condition* (or *datum*) of X.

Solutions are *weakly unique* for initial condition μ_0, if $\mathcal{L}_{X_0} = \mu_0 = \mathcal{L}_{Y_0}$ implies $\mathcal{L}_X = \mathcal{L}_Y$ for any weak solutions X, Y.

It is obvious how to extend the previous definition to initial times $s \geqslant 0$.

As for nonlinear Fokker–Planck equations, one can also consider *linearized* DDSDEs, i.e. one first fixes a curve $t\mapsto \nu_t$ (not necessarily related to any solution) of probability measures in the coefficients and then studies the non-distribution dependent SDE with coefficients $(t,x)\mapsto b(t,\nu_t,x)$, $\sigma(t,\nu_t,x)$. We denote this SDE by (ν-ℓSDE).

However, the name *linearized DDSDE* can be misleading, as the coefficients are typically nonlinear in x. Equation (ν-ℓSDE) is equivalent to the system

$$\begin{cases} dX_t &= b(t, X_t, \mathcal{L}_{X_t})dt + \sigma(t, X_t, \mathcal{L}_{X_t})dB_t, \\ \mathcal{L}_{X_t} &= \nu_t, \quad \forall t \geqslant 0. \end{cases}$$

Remark 3.3.2. *It is straightforward to check that any weak DDSDE-solution X is a weak solution to (ν-ℓSDE) with $\nu_t := \mathcal{L}_{X_t}$.*

One can also consider the distribution-dependent martingale problem (also called *nonlinear martingale problem*) associated with the DDSDE (3.3.1), and one has the same equivalence of existence and uniqueness of weak solutions to (3.3.1) and solutions to this nonlinear martingale problem as in the ”linear” case (Exercise 8.1).

From DDSDEs to NLFPEs. As might be expected, the relation from (3.3.1) to the NLFPE (3.1.1) is similar to the ”linear” case.

Proposition 3.3.3. *Let X be a weak solution to* (3.3.1)*. Then,*

$$\mu = (\mu_t)_{t\geqslant 0},\ \mu_t := \mathcal{L}_{X_t}$$

is a weakly continuous probability solution to the NLFPE with coefficients b and a, where $a = \frac{1}{2}\sigma\sigma^T$. Moreover, $a_{ij}(t, \mu_t, x)$, $, b_i(t, \mu_t, x) \in L^1([0, T] \times \mathbb{R}^d; \mu_t dt)$ for all $T > 0$.

Proof. Exercise 8.2. ☐

In particular: One method to construct weakly continuous probability solutions to NLFPEs is to first solve the corresponding DDSDE and then consider the curve of one-dimensional time marginals of the solution of the latter. There is a list of methods and results on existence and uniqueness for DDSDEs, but we are only going to briefly consider one of these results here. See Exercise 8.3 for more literature on such results.

Well-posedness under Wasserstein-Lipschitz- and monotonicity assumptions. Consider for $p \in [1, \infty)$ the *p-Wasserstein space*

$$\mathcal{P}_p := \left\{ \zeta \in \mathcal{P} : \int_{\mathbb{R}^d} |x|^p \, d\zeta(x) < \infty \right\}$$

and, for $\zeta, \nu \in \mathcal{P}_p$, the *p-Wasserstein distance*

$$\mathcal{W}_p(\zeta, \nu) := \inf_{\Lambda \in C(\zeta,\nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \, d\Lambda(x, y) \right)^{\frac{1}{p}},$$

where $C(\zeta, \nu)$ is the set of all *couplings* between ζ and ν. A coupling between ζ and ν is any Borel probability measure Λ on $\mathbb{R}^d \times \mathbb{R}^d$ such that $\Lambda \circ (\pi^1)^{-1} = \zeta$ and $\Lambda \circ \pi^2)^{-1} = \nu$. $C(\zeta, \nu)$ is non-empty, since $\zeta \otimes \nu \in C(\zeta, \nu)$. Here we denote by $\pi^i : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ the projection on the i-th component.

The spaces $(\mathcal{P}_p, \mathcal{W}_p)$ are complete (!) metric (!) spaces and are used frequently in the study of DDSDEs and other aspects of stochastic analysis.

Consider $p \geqslant 1$, product-measurable coefficients σ_{ij}, b_i defined on $\mathbb{R}_+ \times \mathcal{P} \times \mathbb{R}^d$ with the following assumptions. $\mathcal{P}_p$ is always equipped with the topology induced by $\mathcal{W}_p$ (which is strictly stronger than the weak topology on $\mathcal{P}_p$).

(A0) $b(t, \cdot, \cdot)$ is continuous on $\mathcal{P}_p \times \mathbb{R}^d$ for all $t \geqslant 0$.

(A1) $\exists\, K_1, K_2 \in C(\mathbb{R}_+, \mathbb{R}_+)$ non-decreasing such that for all $t \geqslant 0, \zeta, \nu \in \mathcal{P}_p, x, y \in \mathbb{R}^d$

$$|\sigma(t, \zeta, x) - \sigma(t, \nu, y)|^2 \leqslant K_1(t)|x-y|^2 + K_2(t)\mathcal{W}_p(\zeta, \nu)^2.$$

(A2) $2(b(t, \zeta, x) - b(t, \nu, y)) \cdot (x - y) \leqslant K_1(t)|x-y|^2 + K_2(t)\mathcal{W}_p(\zeta, \nu)|x-y|.$

(A3) b is bounded on bounded sets in $\mathbb{R}_+ \times \mathcal{P}_p \times \mathbb{R}^d$, and

$$|b(t, \zeta, 0)|^p \leqslant K_1(t)\big(1 + \zeta(|\cdot|^p)\big),$$

where $\zeta(|\cdot|^p) = \int_{\mathbb{R}^d} |x|^p \, d\zeta(x)$.

Theorem 3.3.4 (Thm.2.1 from [23])**.** *Assume there is $p \geqslant 1$ such that (A0)-(A3) are satisfied. If $p < 2$, additionally assume $K_2 = 0$. Then for every initial datum $\mu_0 \in \mathcal{P}_p$, the DDSDE has a unique weak solution $X(\mu_0)$ with $\mathcal{L}_{X_t} \in \mathcal{P}_p$ for all $t \geqslant 0$.*

Moreover, if $\mu_0 \in \mathcal{P}_q$ for $q \geqslant p$, then

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |X(\mu_0)_t|^q\Big] < \infty, \quad \forall T > 0.$$

Finally, there is $\psi \in C(\mathbb{R}_+, \mathbb{R}_+)$ non-decreasing such that

$$\mathcal{W}_p\big(\mathcal{L}_{X(\zeta)_t}, \mathcal{L}_{X(\nu)_t}\big)^p \leqslant \mathcal{W}_p\big(\zeta, \nu\big)^p e^{\int_0^t \psi(r)dr}, \quad \forall t > 0.$$

Remark 3.3.5. *Under assumptions (A0)-(A3) one can actually prove that solutions are probabilistically strong and strongly (i.e. pathwise) unique, see [23] for details.*

3.4 Superposition principle: nonlinear case

Unless stated otherwise, the results of this section hold for any initial time $s \geqslant 0$ instead of 0. Analogous to the linear case, we have the following superposition principle-result for nonlinear FPEs.

We refer to (3.1.1) and (3.3.1) as "the NLFPE" and "the DDSDE", respectively. For the following result, see [5, 6]

Theorem 3.4.1 (Superposition principle: nonlinear case)**.** *Let $\mu = (\mu_t)_{t\geqslant 0}$ be a weakly continuous probability solution to the NLFPE (3.1.1) in the sense of Definition 3.1.1 such that*

$$[(t,x) \mapsto b_i(t, \mu_t, x)],\ [(t,x) \mapsto a_{ij}(t, \mu_t, x)] \in L^1([0,T]\times\mathbb{R}^d; \mu_t dt), \quad \forall T > 0. \tag{3.4.1}$$

Then there is a weak solution X to the corresponding DDSDE (3.3.1) such that $\mathcal{L}_{X_t} = \mu_t$ for all $t \geqslant 0$. In particular, X and μ have the same initial condition.

Remark 3.4.2. *(i) Note that there is no regularity assumption on the coefficients, neither in their space- or measure-argument. In particular, the theorem applies to coefficients of Nemytskii-type.*

(ii) The integrability assumption can be weakened to

$$\big[(t,x) \mapsto \frac{|a_{ij}(t,\mu_t,x)| + |b(t,\mu_t,x)\cdot x|}{1+|x|^2}\big] \in L^1([0,T]\times\mathbb{R}^d;\mu_t dt), \quad \forall T>0.$$

Proof of Theorem 3.4.1. $(\mu_t)_{t\geqslant 0}$ is a solution to (μ-ℓFPE) and, by assumption, satisfies (1.3.2) with coefficients $(t,x)\mapsto a_{ij}(t,\mu_t,x)$ and $(t,x)\mapsto b_i(t,\mu_t,x)$. Hence by Theorem 1.3.6, there is a weak solution to (μ-ℓSDE) X with $\mathcal{L}_{X_t}=\mu_t$, $t\geqslant 0$. Therefore, X solves the DDSDE, which yields the claim. □

As in the linear case, the dual statement gives a uniqueness criterion for the NLFPE:

Corollary 3.4.3. *If there is at most one weak solution to the DDSDE with initial datum ζ, then there is at most one weakly continuous probability solution μ to the associated NLFPE with initial condition ζ satisfying*

$$\int_0^T\int_{\mathbb{R}^d} a_{ij}(t,\mu_t,x)| + |b_i(t,\mu_t,x)|\, d\mu_t dt < \infty, \quad \forall T>0.$$

Proof. By Theorem 3.4.1, any two such NLFPE-solutions can be lifted to a weak solution to the associated DDSDE. By assumption, in particular the one-dimensional time marginals of these solutions coincide, which yields the claim. □

It is left as an exercise to write down explicitly the corresponding DDSDEs for the NLFPE-examples from Section 3.1.

It is a natural question whether Proposition 1.3.10 extends to the nonlinear case. This is the content of the next result which shows again the importance of the linearized equation associated with a NLFPE.

Proposition 3.4.4. *Let $\mu_0\in\mathcal{P}$. Assume:*

(i) The NLFPE has a unique weakly continuous probability solution μ with initial condition μ_0.

(ii) The linear FPE (μ-ℓFPE) has a unique weakly continuous probability solution for every initial condition (s,δ_x).

Then weak solutions for the DDSDE with initial condition μ_0 are unique.

Proof. Let X and Y be weak solutions to the DDSDE with initial condition μ_0. By Proposition 3.3.3, $\mu^1 := (\mathcal{L}_{X_t})_{t\geqslant 0}$ and $\mu^2 := (\mathcal{L}_{Y_t})_{t\geqslant 0}$ are weakly continuous probability solutions to the NLFPE with initial condition μ_0. Hence, the assumption implies $\mu^i=\mu$, $i\in\{1,2\}$, where μ is the solution from (i). So, X and Y are weak solutions to (μ-ℓSDE). So, by Propositions 1.3.10 and 1.3.11, the claim follows. □

Remark 3.4.5. *Let X be a weak solution to the DDSDE with initial condition μ_0 and denote by $(Q_x)_{x\in\mathbb{R}^d}$ the disintegration family of $\mathcal{L}_X$ w.r.t. π_0. In contrast to the linear case (see Lemma 1.3.2), it is not true in general that for μ_0-a.e. x the measure Q_x is a solution law to the same equation than X. In fact, considering X as a solution to its own linearized SDE, it follows from Lemma 1.3.2 (i) that μ_0-a.e. Q_x is a solution law to this linearized SDE. Since in general $\mathcal{L}_{Q_x(t)} \neq \mathcal{L}_{X_t}$ (unless μ_0 is a Dirac measure), the latter equation is not the same as the original DDSDE.*

Therefore, the uniqueness of weak solutions to the DDSDE for all Dirac initial data does not imply weak uniqueness for all *initial data. Note that in Proposition 1.3.11 this was proven in the linear case.*

DDSDEs and Markov processes. In Theorem 2.2.2, we particularly proved the following: If a "linear" (i.e. non-distribution dependent) time-homogeneous SDE has a unique weak solution law P_x for all initial data δ_x, $x \in \mathbb{R}^d$, then $(P_x)_{x\in\mathbb{R}^d}$ is a Markov process (in the canonical model). For the proof, we heavily used the stability of the associated linear martingale problem w.r.t. disintegration, i.e. Lemma 1.3.2, which – as said in the previous remark – fails in the case of a distribution-dependent SDE/a nonlinear martingale problem. As a consequence, we have:

Fact. The family of weak solution laws $(P_x)_{x\in\mathbb{R}^d}$ of a weakly well-posed DDSDE is, in general, not a Markov process.

One possible way to resolve this issue is to assume that for every $\nu^x = (\nu_t^x)_{t\geqslant 0}$, $\nu_t^x := P_x \circ \pi_t^{-1}$, the SDE ($\nu^x$-$\ell$SDE) is weakly well posed. Then, by Theorem 2.2.2, there is a family of Markov processes $(P_y^x)_{y\in\mathbb{R}^d}$, where P_y^x denotes the unique weak solution law to (ν^x-ℓSDE) with initial condition δ_y. This way, each P_x is a member of a Markov process, namely $P_x = P_x^x$. The issue with this ansatz is the additional assumption on the well-posedness of the family (!) of linearized SDEs and, even more, the fact that the family of families $(P_y^x)_{y\in\mathbb{R}^d}$, $x \in \mathbb{R}^d$, contains a lot of irrelevant "information" with regard to $(P_x)_{x\in\mathbb{R}^d}$.

We will present a different, more suitable, method later on.

4 Flow selections for nonlinear Fokker–Planck equations

Let $a_{ij}, b_i : \mathbb{R}_+ \times \mathcal{P} \times \mathbb{R}^d \to \mathbb{R}$ be such that $a = (a_{ij})_{i,j \leqslant d}$ is pointwise symmetry and nonnegative definite, and consider our usual NLFPE

$$\partial_t \mu_t = \partial^2_{ij}\big(a_{ij}(t,\mu_t,x)\mu_t\big) - \partial_i\big(b_i(t,\mu_t,x)\mu_t\big). \tag{4.0.1}$$

For $(s,\zeta) \in \mathbb{R}_+ \times \mathcal{P}$, denote by S_s and $S_{s,\zeta}$ the sets of weakly continuous probability solutions from time s and its subset of solutions with initial datum (s,ζ), respectively.

In this chapter, we address the following question: *Assume $|S_{s,\zeta}| \geqslant 1$ for all $(s,\zeta) \in \mathbb{R}_+ \times \mathcal{P}$. Is there $\mu^{s,\zeta} \in S_{s,\zeta}$ such that $(\mu^{s,\zeta})_{s\in\mathbb{R}_+,\zeta\in\mathcal{P}}$ has the flow property, i.e.*

$$\mu^{s,\zeta}_t = \mu^{r,\mu^{s,\zeta}_r}_t, \quad \forall 0 \leqslant s \leqslant r \leqslant t, \zeta \in \mathcal{P}?$$

This is the same notion of *flow* as in (2.1.1). We call such a family a *flow selection* for the NLFPE.

We will ask the same question for an a priori chosen subset of initial data $\mathcal{P}_0 \subseteq \mathcal{P}$. In this case, one also has to check that the flow leaves $\mathcal{P}_0$ invariant.

Remark 4.0.1. *(i) If $|S_{s,\zeta}| = 1$, the family of unique elements $\mu^{s,\zeta} \in S_{s,\zeta}$ has the flow property (Exercise 9.1). Note that this is* not *true if we consider the case of 'non-Markovian' coefficients, i.e. when $a(t)$ and $b(t)$ depend not only on (μ_t, x), but on $((\mu_r)_{r\leqslant t}, x)$.*

(ii) The importance of flow selections will become apparent in the next chapter.

Here we present two very different methods to give positive answers to this question: In Section 4.1, we *construct* a family of solutions with the flow property; in Section 4.2, we *select* solutions $\mu^{s,\zeta} \in S_{s,\zeta}$ such that this selected family has the flow property.

4.1 Crandall-Liggett semigroup-method

An excellent reference for the contents of this section is the monograph [2].

4.1.1 Accretive and dissipative operators in Banach spaces

Let X be a Banach space with norm $|\cdot|_X$. We simply write $|\cdot|$, if no confusion with the standard Euclidean norm on $\mathbb{R}$ can occur. By I we denote the identity operator, $I : X \to X$, $Ix = x$.

Definition 4.1.1. (i) An operator $(A, D(A))$, $A : D(A) \subseteq X \to X$, is called *accretive*, if

$$|x - y| \leqslant |x - y + \lambda(Ax - Ay)|, \quad \forall \lambda > 0, x, y \in D(A). \tag{4.1.1}$$

(ii) An accretive operator is called *m-accretive*, if $R(I + \lambda A) = X$ for all $\lambda > 0$, where $R(I + \lambda A)$ denotes the range of $I + \lambda A : D(A) \subseteq X \to X$.

(iii) $(A, D(A))$ is called *quasi m-accretive*, if there is $\omega \in \mathbb{R}$ such that $(A+\omega I, D(A))$ is m-accretive.

(iv) $(A, D(A))$ is called *dissipative, m-dissipative, quasi m-dissipative*, if $(-A, D(A))$ is accretive, m-accretive, quasi m-accretive, respectively.

'*accretive*' = dt. '*wachsend, zunehmend*'.

In fact, one can show that $(A, D(A))$ is accretive if and only if it satisfies the inequality from (4.1.1) for *some* $\lambda > 0$, and m-accretive if and only if it is accretive and $R(I + \lambda A) = X$ for *some* $\lambda > 0$.

Remark 4.1.2. *We write Ax for $A(x)$, $x \in D(A)$, but $(A, D(A))$ is NOT assumed to be linear. In fact, considering nonlinear accretive operators will be essential in the sequel.*

4.1.2 Differential equations in Banach spaces

Let $(A, D(A))$ be an operator on X, $T > 0$, and consider the Cauchy problem

$$y'(t) = Ay(t), \quad y(0) = y_0, \tag{4.1.2}$$

where $y_0 \in X$.

The equality is understood in X. This raises two immediate questions: What is the meaning of $y'(t)$? Second, to solve this equation pointwise, one needs $y(t) \in D(A)$, which is hard (think for instance of $X = L^2(\mathbb{R}^d)$ and A being a differential operator). There is a theory of *strong solutions* to such Cauchy problems, where both questions are taken into account. We will, however, focus on a different notion of solution.

Definition 4.1.3. Let $T > 0$, $\varepsilon > 0$.

(i) An ε-*discretization* of $[0, T]$ is any partition $p^\varepsilon(t_0, \dots, t_N)$, given by $0 = t_0 \leqslant t_1 \leqslant \dots \leqslant t_N \leqslant T$ such that $T - t_N \leqslant \varepsilon$ and $t_i - t_{i-1} \leqslant \varepsilon$, $i \in \{1, \dots, N\}$.

(ii) A $p^\varepsilon(t_0, \dots, t_N)$-solution to (4.1.2) on $[0, T]$ is a piecewise constant function $z : [0, t_N] \to X$ whose values z_i on $(t_{i-1}, t_i]$ satisfy the implicit difference scheme

$$z_i = (t_i - t_{i-1})Az_i + z_{i-1},$$

for all $i \in \{1, \dots, N\}$, and $z(0) := z_0 := y_0$.

(iii) For $\varepsilon > 0$, an *ε-approximate solution to the Cauchy problem* (4.1.2) on $[0, T]$ is any $p^\varepsilon(t_0, \dots, t_N)$-solution for any ε-discretization $p^\varepsilon(t_0, \dots, t_N)$.

Definition 4.1.4. A *mild solution* to the Cauchy problem (4.1.2) on $[0, \infty)$ is a function $y \in C([0, \infty), X)$ such that for each $\varepsilon > 0$ and $T > 0$ there is an ε-approximate solution z_ε to (4.1.2) on $[0, T]$ such that $\sup_{t \leqslant T} |y(t) - z_\varepsilon(t)| \leqslant \varepsilon$.

The usefulness of this solution notion stems from the famous *Crandall–Liggett nonlinear semigroup* result:

Theorem 4.1.5 (Crandall-Liggett nonlinear semigroup theorem, cf. Thm.4.1 of [2]). *Let $(A, D(A))$ be quasi m-dissipative and $y_0 \in \overline{D(A)}$ (the closure of $D(A)$ in X). Then the Cauchy problem (4.1.2) has a unique mild solution $y = y(y_0)$ on $\mathbb{R}_+$, and it is given by*

$$y(t) = \lim_{n \to \infty} \left(I - \frac{t}{n} A \right)^{-n} y_0, \quad t > 0 \tag{4.1.3}$$

where the convergence holds locally uniformly in t on $\mathbb{R}_+$.

Remark 4.1.6. *The exponential formula (4.1.3) justifies to also write $y(y_0)(t) = \exp(tA)(y_0)$, and it is readily seen that $S(t, y_0) := y(y_0)(t)$ has the (time-homogeneous) flow property $S(t + s, y_0) = S(t, S(s, y_0))$, $\forall t, s \geqslant 0, y_0 \in \overline{D(A)}$.*

Application to NLFPEs. Consider, for instance, the generalized PME

$$\partial_t u = \Delta\beta(u) - \operatorname{div}\big(DB(u)u\big), \quad (t, x) \in (0, \infty) \times \mathbb{R}^d \tag{4.1.4}$$

(see Example (ii) in Section 3.1) under suitable assumptions for β, D, B. In particular: $\beta \in C^2(\mathbb{R})$, D, B bounded. To treat this equation via the Crandall–Liggett method, consider the operator $(A_0, D(A_0))$ on $L^1(\mathbb{R}^d)$, defined by

$$A_0 : D(A_0) \subseteq L^1(\mathbb{R}^d) \to L^1(\mathbb{R}^d), \quad A_0 y := \Delta\beta(y) - \operatorname{div}(DB(y)y)$$

with domain

$$D(A_0) := \{y \in L^1(\mathbb{R}^d) : \beta(y) \in L^1_{\text{loc}}(\mathbb{R}^d), \Delta\beta(y) - \operatorname{div}(DB(y)y) \in L^1(\mathbb{R}^d)\}.$$

$\Delta\beta(y)$ and $\operatorname{div}(DB(y)y)$ are taken in the sense of distributions [which requires only $\beta(y), DB(y)y \in L^1_{\text{loc}}(\mathbb{R}^d)$], and it is only assumed that their sum is in $L^1(\mathbb{R}^d)$. One can show (cf. [4])

(i) $R(I - \lambda A_0) = L^1(\mathbb{R}^d), \quad \forall \lambda > 0$;

(ii) There is a restriction $(A, D(A))$ of $(A_0, D(A_0))$, i.e. $D(A) \subseteq D(A_0)$ and $A = A_0$ on $D(A)$, such that (i) also holds for A, and $(A, D(A))$ is dissipative on $L^1(\mathbb{R}^d)$;

(iii) $\overline{D(A)} = L^1(\mathbb{R}^d)$, where the closure is taken in $L^1(\mathbb{R}^d)$.

So, Theorem 4.1.5 implies the existence of a unique mild solution $u = u(u_0)$ for

$$u'(t) = Au(t), \quad y(0) = u_0 \tag{4.1.5}$$

on $(0,\infty)$ for all $u_0 \in L^1(\mathbb{R}^d)$. In particular, $u \in C(\mathbb{R}_+, L^1(\mathbb{R}^d))$. One can also show: $t \mapsto u(t,x)dx$ is a weakly continuous solution to the Nemytskii-type NLFPE (4.1.4) in the sense of Definition 3.1.1; $u \geqslant 0$ if $u_0 \geqslant 0$; $|u(t)|_{L^1} = |u_0|_{L^1}$ for all $t > 0$. By Remark 4.1.6, $\{u(u_0)\}_{u_0 \in L^1}$ has the flow property in $L^1(\mathbb{R}^d)$.

Conclusion: Posing (4.1.4) as a nonlinear evolution equation in $L^1(\mathbb{R}^d)$, the Crandall–Liggett semigroup approach yields a family of weakly continuous (probability) solutions for every $L^1(\cap\mathcal{P})$-valued initial datum, and this family has the flow property in $L^1(\cap\mathcal{P})$.

Remark 4.1.7. *Note that $u(u_0)$ is not necessarily the unique L^1-mild solution to* (4.1.4)*, since we considered a restriction A of A_0. So, we only obtain mild uniqueness for* (4.1.5)*, which is not equivalent to* (4.1.4)*. Under stronger assumptions on the coefficients, one can prove m-dissipativity of $(A_0, D(A_0))$, and in this case – without passing via a restriction – mild uniqueness for* (4.1.4) *follows.*

4.2 Flow selections

The reference for this section is [17].

We denote by SP_s the set of vaguely continuous subprobability measure-valued solutions μ to the NLFPE such that

$$[(t,x) \mapsto a_{ij}(t,\mu_t,x)], [(t,x) \mapsto b_i(t,\mu_t,x)] \in L^1_{\mathrm{loc}}([0,\infty) \times \mathbb{R}^d; \mu_t dt),$$

and $SP_{s,\zeta}$ its subset of solutions with initial datum $\zeta \in \mathcal{SP}$.

Definition 4.2.1. A family $\{\mathcal{A}_{s,\zeta}\}_{s \geqslant 0, \zeta \in \mathcal{SP}}$, $\mathcal{A}_{s,\zeta} \subseteq SP_{s,\zeta}$, is *flow-admissible*, if

(i) $(\mu_t)_{t \geqslant s} \in \mathcal{A}_{s,\zeta} \implies (\mu_t)_{t \geqslant r} \subseteq \mathcal{A}_{r,\mu_r}, \quad \forall r \geqslant s \geqslant 0, \zeta \in \mathcal{SP}$;

(ii) $(\mu_t)_{t \geqslant s} \in \mathcal{A}_{s,\zeta}$ and $(\eta_t)_{t \geqslant r} \in \mathcal{A}_{r,\mu_r}$ implies $\mu \circ_r \eta \in \mathcal{A}_{s,\zeta}$, where

$$(\mu \circ_r \eta)_t := \begin{cases} \mu_t, & \text{if } t \leqslant r \\ \eta_t, & \text{if } t \geqslant r. \end{cases}$$

For each $s \geqslant 0$, we denote by $A_s \subseteq \mathcal{SP}$ the set of ζ for which $\mathcal{A}_{s,\zeta} \neq \emptyset$. We say (s,ζ) is *admissible*, if $\zeta \in A_s$.

A family $\mu^{s,\zeta}$, $s \geqslant 0, \zeta \in A_s$, is a *solution flow to the NLFPE in* $\{\mathcal{A}_{s,\zeta}\}$, if $\mu^{s,\zeta} \in \mathcal{A}_{s,\zeta}$ and

$$\mu_t^{s,\zeta} = \mu_t^{r,\mu_r^{s,\zeta}}, \quad \forall t \geqslant r \geqslant s, \zeta \in A_s. \tag{4.2.1}$$

Example 4.2.2. The families $\mathcal{A}_{s,\zeta} = SP_{s,\zeta}$ and

$$\mathcal{A}_{s,\zeta} = \begin{cases} SP^1_{s,\zeta}, & \text{if } \zeta \in \mathcal{P} \\ \emptyset, & \text{if } \zeta \notin \mathcal{P} \end{cases}$$

are both flow-admissible, where $SP^1_{s,\zeta}$ is the subset of $SP_{s,\zeta}$ consisting of *probability* solutions. For a third example, denote by $SP^{\ll}_{s,\zeta}$ the subset of $SP_{s,\zeta}$ of curves consisting of dx-absolutely continuous subprobability measures for all $t > s$. Then, for each $\mathcal{SP}^{\ll} \subseteq \mathfrak{M} \subseteq \mathcal{SP}$ the family

$$\mathcal{A}_{s,\zeta} := \begin{cases} SP^{\ll}_{s,\zeta}, & \zeta \in \mathfrak{M} \\ \emptyset, & \zeta \notin \mathfrak{M} \end{cases}$$

is flow-admissible. This case appears for Nemytskii-type equations and in cases in which it is known that for each $\zeta \in \mathfrak{M}$ solutions from initial datum ζ are function-valued at each positive time (also called L^1-regularization).

We denote by τ_v the topology of vague convergence on $\mathcal{SP}$. Recall that a topological space X is *Hausdorff*, if for any pair of points $x, y \in X, x \neq y$, there exist disjoint open sets $A, B \subseteq X$ with $x \in A, y \in B$. In particular, every metric space is Hausdorff, but not every Hausdorff space is metrizable.

The main result of this section is the following theorem.

Theorem 4.2.3. *Let (H, τ) be a Hausdorff topological space with $H \subseteq \mathcal{SP}$, $\tau \supseteq \tau_v$, and let $\{\mathcal{A}_{s,\zeta}\}_{s\geqslant 0,\zeta\in\mathcal{SP}}$ be flow-admissible. If each $\mathcal{A}_{s,\zeta}$ is a compact subset of $C([s,\infty), H)$ w.r.t. the topology of pointwise (!) convergence, then there exists a solution flow to the NLFPE in $\{\mathcal{A}_{s,\zeta}\}$.*

We abbreviate $C_sH := C([s,\infty), H)$.

Remark 4.2.4. *(i) Note that the topology of pointwise convergence on C_sH, denoted τ_{pt} (suppressing the dependence on H and s in the notation), is a rather coarse topology. For instance, if H is a metric space, then $\tau_{pt} \subset \tau_{lu}$ on C_sH, where τ_{lu} denotes the topology of locally uniform convergence. Recall that for ordered topologies $\tau_1 \subseteq \tau_2$ on a set X any τ_2-compact subset is also τ_1-compact, Thus, the compactness-criterion in the previous theorem is relatively simple to check.*

(ii) Typical choices for H are $H = \mathcal{SP}$, $H = \mathcal{P}$ with $\tau = \tau_v$. Another choice is to take $\mathcal{A}_{s,\zeta}$ as a subset of L^2-valued L^2-weakly continuous curves and $(H, \tau) = (L^2 \cap \mathcal{SP}, \tau_{2,w})$, where $\tau_{2,w}$ denotes the weak topology on L^2. This space is not *metrizable, but Hausdorff.*

Regarding the proof of Theorem 4.2.3, we need the following definition. Set $\mathbb{Q}_s := \mathbb{Q} \cap [s,\infty)$.

Definition 4.2.5. (i) We call any bijective map $\xi : \mathbb{N} \times \mathbb{Q}_0 \to \mathbb{N}_0$ an *enumeration.* For such ξ and $k \in \mathbb{N}_0$, we write $(n_k, q_k) := \xi^{-1}(k)$.

(ii) For $s \geqslant 0$, denote by $(m^s_k)_{k\in\mathbb{N}_0} \subseteq \mathbb{N}_0$ the enumerating sequence of $\mathbb{N} \times \mathbb{Q}_s$ with respect to a prescribed enumeration ξ, i.e. there exist exactly k elements (n, q) in $\mathbb{N} \times \mathbb{Q}_s$ with $\xi(n, q) < m^s_k$.

Note that for $0 \leqslant s < r$, the sequence $(m^r_k)_{l\in\mathbb{N}_0}$ is a subsequence of $(m^s_k)_{l\in\mathbb{N}_0}$. Moreover, (C_sH, τ_{pt}) is Hausdorff, since so is H. A family $\{f_i\}_{i\in I}$ of bounded measurable functions $f_i : \mathbb{R}^d \to \mathbb{R}$ is called *measure-separating*, if

$$\mu^1 \neq \mu^2 \iff \exists i \in I : \int_{\mathbb{R}^d} f_i \, d\mu^1 \neq \int_{\mathbb{R}^d} f_i \, d\mu^2$$

for all $\mu^1, \mu^2 \in \mathcal{M}_b^+$. There exists a countable measure-separating family in $C_c(\mathbb{R}^d)$ (Exercise 10.1).

Proof of Theorem 4.2.3. Let $\mathcal{H} = \{h_n, n \in \mathbb{N}\} \subseteq C_c(\mathbb{R}^d)$ be measure-separating, ξ be an enumeration, $(s,\zeta) \in [0,\infty) \times \mathcal{SP}$ be admissible and consider

$$\begin{aligned} G_0^{s,\zeta} : C_sH \to \mathbb{R},\ \mu = (\mu_t)_{t\geqslant s} &\mapsto \int_{\mathbb{R}^d} h_{n_{m_0^s}} d\mu_{q_{m_0^s}}, \\ u_0^{s,\zeta} &:= \sup_{\mu \in \mathcal{A}_{s,\zeta}} G_0^{s,\zeta}(\mu), \\ M_0^{s,\zeta} &:= \big((G_0^{s,\zeta})^{-1}(u_0^{s,\zeta})\big) \cap \mathcal{A}_{s,\zeta}. \end{aligned}$$

Since $\tau_v \subseteq \tau$ and $\mathcal{H} \subseteq C_c(\mathbb{R}^d)$, $G_0^{s,\zeta}$ is continuous on C_sH. Furthermore, since $\mathcal{A}_{s,\zeta}$ is nonempty and compact in C_sH, $M_0^{s,\zeta}$ is nonempty and compact in C_sH as well. Define iteratively for $k \in \mathbb{N}_0$

$$\begin{aligned} G_{k+1}^{s,\zeta} : C_sH \to \mathbb{R},\ (\mu_t)_{t\geqslant s} &\mapsto \int_{\mathbb{R}^d} h_{n_{m_{k+1}^s}} d\mu_{q_{m_{k+1}^s}}, \\ u_{k+1}^{s,\zeta} &:= \sup_{\mu \in M_k^{s,\zeta}} G_{k+1}^{s,\zeta}(\mu), \\ M_{k+1}^{s,\zeta} &:= \big((G_{k+1}^{s,\zeta})^{-1}(u_{k+1}^{s,\zeta})\big) \cap M_k^{s,\zeta}. \end{aligned}$$

The same assertions as for $G_0^{s,\zeta}$ and $M_0^{s,\zeta}$ are true for $G_{k+1}^{s,\zeta}$ and $M_{k+1}^{s,\zeta}$. Since $M_{k+1}^{s,\zeta} \subseteq M_k^{s,\zeta}$ and C_sH is Hausdorff, we obtain

$$M^{s,\zeta} := \bigcap_{k\geqslant 0} M_k^{s,\zeta} \neq \emptyset$$

(Exercise 10.2). When $\mu^{(i)} = (\mu_t^{(i)})_{t\geqslant s} \in M^{s,\zeta}$ for $i \in \{1,2\}$, by construction we have

$$\int_{\mathbb{R}^d} h_{n_{m_k^s}} d\mu^{(1)}_{q_{m_k^s}} = \int_{\mathbb{R}^d} h_{n_{m_k^s}} d\mu^{(2)}_{q_{m_k^s}}, \quad k \in \mathbb{N}_0.$$

Since $\{(n_{m_k^s}, q_{m_k^s}),\ k \in \mathbb{N}_0\} = \mathbb{N} \times \mathbb{Q}_s$, this yields $\int h_n d\mu_q^{(1)} = \int h_n d\mu_q^{(2)}$ for all $(n,q) \in \mathbb{N} \times \mathbb{Q}_s$ and hence $\mu_q^{(1)} = \mu_q^{(2)}$ for all $q \in \mathbb{Q}_s$, because $\mathcal{H}$ is measure separating. Since $\mu^{(1)}$ and $\mu^{(2)}$ are continuous in the Hausdorff space H, $\mu^{(1)} = \mu^{(2)}$ follows. Consequently, $M^{s,\zeta} \subseteq \mathcal{A}_{s,\zeta}$ is a singleton, i.e. $M^{s,\zeta} = \{\mu^{s,\zeta}\}$ for some $\mu^{s,\zeta} \in \mathcal{A}_{s,\zeta}$.

It remains to show that the family $\{\mu^{s,\zeta}\}_{s\geqslant 0, \zeta \in A_s}$ has the flow property. To this end, let (s,ζ) be admissible and $0 \leqslant s < r < t$. Consider the admissible (!) initial condition $(r, \mu_r^{s,\zeta})$ and let $\gamma = (\gamma_t)_{t\geqslant r}$ be the unique element in $M^{r,\mu_r^{s,\zeta}}$ according to the above selection, i.e. $\gamma = \mu^{r,\mu_r^{s,\zeta}}$ in our notation. We need to show

$$\gamma_t = \mu_t^{s,\zeta}, \quad \forall\, t \geqslant r. \tag{4.2.2}$$

Set $\eta := \mu^{s,\zeta} \circ_r \gamma \in \mathcal{A}_{s,\zeta}$. Due to the iterative maximizing selection of the first part of the proof, we have

$$\int_{\mathbb{R}^d} h_{n_{m_0^s}} d\mu^{s,\zeta}_{q_{m_0^s}} \geqslant \int_{\mathbb{R}^d} h_{n_{m_0^s}} d\eta_{q_{m_0^s}}. \tag{4.2.3}$$

If $q_{m_0^s} \in [s,r)$, then $\eta_{q_{m_0^s}} = \mu^{s,\zeta}_{q_{m_0^s}}$ by definition and we have equality in (4.2.3). If $q_{m_0^s} \geqslant r$, then $q_{m_0^s} = q_{m_0^r}$ and by the characterizing property of γ in $\mathcal{A}_{r,\mu_r^{s,\zeta}}$, and since $(\mu_t^{s,\zeta})_{t\in[r,\infty)} \in \mathcal{A}_{r,\mu_r^{s,\zeta}}$, we obtain

$$\int_{\mathbb{R}^d} h_{n_{m_0^s}} d\mu^{s,\zeta}_{q_{m_0^s}} \leqslant \int_{\mathbb{R}^d} h_{n_{m_0^s}} d\gamma_{q_{m_0^s}} = \int_{\mathbb{R}^d} h_{n_{m_0^s}} d\eta_{q_{m_0^s}},$$

and hence we have equality in (4.2.3) in any case. Next, consider m_1^s: since (4.2.3) is an equality, both $(\mu_t^{s,\zeta})_{t\geqslant s}$ and $(\eta_t)_{t\geqslant s}$ belong to $M_0^{s,\zeta}$. Using the characterization of $\mu^{s,\zeta}$ again, we obtain

$$\int_{\mathbb{R}^d} h_{n_{m_1^s}} d\mu^{s,\zeta}_{q_{m_1^s}} \geqslant \int_{\mathbb{R}^d} h_{n_{m_1^s}} d\eta_{q_{m_1^s}}, \tag{4.2.4}$$

clearly with equality if $q_{m_1^s} \in [s,r)$. If $q_{m_1^s} \geqslant r$ and $q_{m_0^s} \in [s,r)$, then $m_1^s = m_0^r$, and hence

$$\int_{\mathbb{R}^d} h_{n_{m_1^s}} d\mu^{s,\zeta}_{q_{m_1^s}} \leqslant \int_{\mathbb{R}^d} h_{n_{m_1^s}} d\gamma_{q_{m_1^s}} = \int_{\mathbb{R}^d} h_{n_{m_1^s}} d\eta_{q_{m_1^s}} \tag{4.2.5}$$

by the characterizing property of γ, which gives equality in (4.2.4). If $q_{m_0^s}, q_{m_1^s} \geqslant r$, then $m_0^s = m_0^r$, $m_1^s = m_1^r$ and both $\mu^{s,\zeta}$ and γ are in $M_0^{r,\mu_r^{s,\zeta}}$, which also gives (4.2.5). Hence, equality in (4.2.4) holds in any case. By iteration we obtain

$$\int_{\mathbb{R}^d} h_{n_{m_k^s}} d\mu^{s,\zeta}_{q_{m_k^s}} = \int_{\mathbb{R}^d} h_{n_{m_k^s}} d\eta_{q_{m_k^s}}, \quad \forall\, k \in \mathbb{N}_0,$$

and hence, since $\mathcal{H}$ is measure separating,

$$\mu_q^{s,\zeta} = \eta_q, \quad \forall q \in \mathbb{Q}_s,$$

thus in particular $\mu_q^{s,\zeta} = \eta_q = \gamma_q$ for all $q \in \mathbb{Q}_r$. Since both curves are continuous with values in H, we obtain (4.2.2), which closes the proof. □

Remark 4.2.6. *The previous proof works for any countable measure separating family from $C_c(\mathbb{R}^d)$, any enumeration and any dense countable subset of $[s,\infty)$ instead of $\mathbb{Q}_s$. The selected flow depends on these choices.*

The iterative selection method from the previous proof allows to also prove the following characterization.

Proposition 4.2.7. *In the situation of the previous theorem, the following are equivalent:*

(i) There exists at most one solution flow to the NLFPE with respect to $\{\mathcal{A}_{s,\zeta}\}_{(s,\zeta)\in[0,\infty)\times\mathcal{SP}}$.

(ii) $|\mathcal{A}_{s,\zeta}| \leqslant 1$ *for all* $(s,\zeta)\in\mathbb{R}_+\times\mathcal{SP}$.

Proof. For the nontrivial implication of the assertion, assume there is an admissible initial condition $(s',\zeta')\in[0,T)\times\mathcal{SP}$ with $|\mathcal{A}_{s',\zeta'}|\geqslant 2$. As mentioned in the previous remark, we may choose an enumeration ξ and a family of measure separating functions $\mathcal{H}=\{h_n, n\in\mathbb{N}\}\subseteq C_c(\mathbb{R}^d)$ with $\mathcal{H}=-\mathcal{H}$. Consider the flow $\{\mu^{s,\zeta}\}$ with (s,ζ) running through all admissible initial conditions, constructed as in the proof of the previous theorem subject to this $\mathcal{H}$ and ξ. By assumption, there exists $\gamma\in\mathcal{A}_{s',\zeta'}$ with $\mu^{s',\zeta'}\neq\gamma$, and since both curves are continuous, there is $q\in\mathbb{Q}_{s'}$ such that $\mu^{s',\zeta'}_q\neq\gamma_q$. Thus, considering $-h$ instead of h if necessary, there is $h\in\mathcal{H}$ such that

$$\int_{\mathbb{R}^d} h\,d\gamma_q > \int_{\mathbb{R}^d} h\,d\mu^{s',\zeta'}_q. \tag{4.2.6}$$

Now consider a new enumeration ξ' such that according to ξ' we have $(h_{n_{m_0^{s'}}}, q_{n_{m_0^{s'}}}) = (h,q)$, and denote the flow subject to $\mathcal{H}$ and ξ' by $\{\eta^{s,\zeta}\}$ (the sets of admissible initial conditions remain unchanged). Selecting as in the proof of the previous theorem gives

$$\int_{\mathbb{R}^d} h\,d\eta^{s',\zeta'}_q = \sup_{\mu\in\mathcal{A}_{s',\zeta'}}\Big(\int_{\mathbb{R}^d} h\,d\mu_q\Big).$$

Therefore, taking into account (4.2.6), we conclude

$$\int_{\mathbb{R}^d} h\,d\eta^{s',\zeta'}_q \geqslant \int_{\mathbb{R}^d} h\,d\gamma_q > \int_{\mathbb{R}^d} h\,d\mu^{s',\zeta'}_q.$$

Hence $\eta^{s',\zeta'}\neq\mu^{s',\zeta'}$, which contradicts (i) and finishes the proof. □

4.2.1 Applications

Recall that a subset $A\subseteq X$ of a topological space X is *relatively compact*, if its closure is a compact subset of X. In particular, a closed relatively compact set is compact. For two topological spaces X,Y, the *compact-open topology* on $C(X,Y)$ (the space of continuous maps from X to Y) is the topology with subbase

$$\{f\in C(X,Y) : f(K)\subseteq O\},\quad K\subseteq X \text{ compact}, O\subseteq Y \text{ open}.$$

For our applications, we will use the following general version of the Arzelá-Ascoli theorem

Proposition 4.2.8 (Arzelà-Ascoli theorem, Thm.47.1 [15])**.** *Let* I *be an interval and* (Y,d) *a metric space. A subset* $\mathcal{F}\subseteq C(I,Y)$ *is relatively compact in the compact-open topology if and only if* $\mathcal{F}$ *is pointwise relatively compact and equicontinuous, i.e. if* *(i)* $\{f(t), f \in \mathcal{F}\}$ *is relatively compact in* Y *for all* $t \in I$

(ii) *For all* $t \in I$ *and* $\varepsilon > 0$ *there is* $\delta > 0$ *such that*

$$r \in I, |t-r| < \delta \implies \sup_{f \in \mathcal{F}} d(f(t), f(r)) < \varepsilon.$$

Remark 4.2.9. *Let* Y *(with a fixed topology) be metrizable.*

(i) *The topology* τ_{lu} *on* C_sY *is independent of the choice of compatible metric on* Y*. This follows from the fact that for any such metric,* τ_{lu} *coincides with the compact-open topology on* C_sY *and the straightforward observation that the compact-open topology on* C_sY *only depends on the topology of* Y*, not on its metric.*

(ii) *Whether a subset* $\mathcal{F} \subseteq C_sY$ *is equicontinuous generally depends on the choice of compatible metric on* Y*. However, the Arzelá-Ascoli theorem asserts an equivalence between a) relative compactness of* $\mathcal{F}$ *and b.1) pointwise relative compactness plus b.2) equicontinuity. Since properties a) and b.1) for* $\mathcal{F}$ *are clearly independent of the choice of compatible metric on* Y*, it follows that equicontinuity of a pointwise relatively compact set* $\mathcal{F}$ *is independent of the choice of compatible metric on* Y*.*

The bottomline of the previous remark for our application is: If $\mathcal{A}_{s,\zeta} \subseteq C_sH$ is pointwise relatively compact and we want to prove relative compactness of $\mathcal{A}_{s,\zeta}$ w.r.t. τ_{lu} via Arzelá-Ascoli's theorem, we may choose *any* compatible metric on (H, τ) to prove equicontinuity.

Linear equations. Consider the usual linear FPE

$$\partial_t \mu_t = \partial^2_{ij}\big(a_{ij}(t,x)\mu_t\big) - \partial_i\big(b_i(t,x)\mu_t\big) \tag{4.2.7}$$

and suppose the coefficients $a_{ij}, b_i : (0,\infty) \times \mathbb{R}^d \to \mathbb{R}$, $1 \leqslant i,j \leqslant d$, are Borel and satisfy

Assumption A1.

(A1.i) $\int_0^T \sup_{x\in\mathbb{R}^d} \big(|a_{ij}(t,x)| + |b_i(t,x)|\big) dt < \infty, \quad \forall T > 0, i,j \leqslant d.$

(A1.ii) $x \mapsto a_{ij}(t,x)$ and $x \mapsto b_i(t,x)$ are continuous for dt-a.e. $t > 0$.

We already known that in this case $SP_{s,\zeta} = SP^1_{s,\zeta}$ for $\zeta \in \mathcal{P}$ and each curve in $SP_{s,\zeta}$ is weakly continuous. Consider

$$\mathcal{A}_{s,\zeta} := \begin{cases} SP_{s,\zeta} & , \text{ if } \zeta \in \mathcal{P} \\ \emptyset & , \text{ if } \zeta \in \mathcal{SP}\backslash\mathcal{P}, \end{cases} \tag{4.2.8}$$

which is flow-admissible by Example 4.2.2.

Proposition 4.2.10. *Suppose Assumption A1 holds and that* $SP_{s,\zeta}$ *is nonempty for each* $(s,\zeta) \in [0,\infty) \times \mathcal{P}$*. Then there is a solution flow for* (4.2.7) *in* $\{\mathcal{A}_{s,\zeta}\}_{s\geqslant 0, \zeta \in \mathcal{SP}}$*.*

Proof. Let $(H,\tau) = (\mathcal{SP}, \tau_v)$. By Theorem 4.2.3 and Remark 4.2.4 (i), it suffices to prove that each $\mathcal{A}_{s,\zeta}$ is a compact subset of C_sH w.r.t. τ_{lu}, so we prove $\mathcal{A}_{s,\zeta}$ is closed, pointwise relatively compact and equicontinuous in order to apply Proposition 4.2.8. Since $(\mathcal{SP}, \tau_v)$ is a compact metrizable space (see in particular Remark 1.1.5 (iii)), pointwise relative compactness follows.

Concerning closedness, since $(\mathcal{SP}, \tau_v)$ is metrizable, also $(C_s\mathcal{SP}, \tau_{lu})$ is metrizable, hence sequential. Thus it suffices to prove that the limit of any τ_{lu}-converging sequence in $\mathcal{A}_{s,\zeta}$ belongs to $\mathcal{A}_{s,\zeta}$. So, let $\mu^{(n)} = (\mu_t^{(n)})_{t\geqslant s}, n \geqslant 1$, be a τ_{lu}-converging sequence in $\mathcal{A}_{s,\zeta}$ with limit $\mu \in C_s\mathcal{SP}$ and let $\varphi \in C_c^2(\mathbb{R}^d)$. Due to (A1.ii), we have $L_{a,b}\varphi(t) \in C_c(\mathbb{R}^d)$ dt-a.s., hence

$$\int_{\mathbb{R}^d} L_{a,b}\varphi(t)\, d\mu_t^{(n)} \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}^d} L_{a,b}\varphi(t)\, d\mu_t \quad dt\text{-a.s.},$$

and by (A1.i), Lebesgue's dominated convergence theorem gives

$$\int_s^t \int_{\mathbb{R}^d} L_{a,b}\varphi\, d\mu_\tau^{(n)} d\tau \underset{n\to\infty}{\longrightarrow} \int_s^t \int_{\mathbb{R}^d} L_{a,b}\varphi\, d\mu_\tau d\tau, \quad \forall t > s.$$

Therefore, $\mu \in \mathcal{A}_{s,\zeta}$.

Regarding equicontinuity, thanks to Remark 4.2.9, we may consider the following convenient τ_v-compatible metric (Exercise 11.1) on $\mathcal{SP}$:

$$d_v(\zeta_1, \zeta_2) := \sum_{l\geqslant 1} 2^{-l} C_l^{-1} \bigg[\bigg| \int_{\mathbb{R}^d} f_l d\zeta_1 - \int_{\mathbb{R}^d} f_l d\zeta_2 \bigg| \wedge 1 \bigg], \quad \zeta_1, \zeta_2 \in \mathcal{SP},$$

where $\{f_l, l \in \mathbb{N}\} =: \mathcal{F} \subseteq C_c^2(\mathbb{R}^d)$ is arbitrary but fixed and consists of nontrivial elements such that the closure of $\mathcal{F}$ with respect to uniform convergence contains $C_c(\mathbb{R}^d)$. We choose

$$C_l := 1 + D_l, \quad D_l := (d^2 + d) \max_{1\leqslant i,j\leqslant d} \{||\partial_i f_l||_\infty, ||\partial_{ij} f_l||_\infty\}.$$

We obtain for each $\mu \in \mathcal{A}_{s,\zeta}$ and arbitrary $s \leqslant t_1 \leqslant t_2$:

$$\begin{aligned} d_v(\mu_{t_1}, \mu_{t_2}) &\leqslant \sum_{l\geqslant 1} 2^{-l} C_l^{-1} \bigg[\int_{t_1}^{t_2} \int_{\mathbb{R}^d} |L_{a,b} f_l(t)| d\mu_t dt \wedge 1 \bigg] \\ &\leqslant \sum_{l\geqslant 1} 2^{-l} \bigg[\int_{t_1}^{t_2} \max_{1\leqslant i,j\leqslant d} \sup_{x\in\mathbb{R}^d} \big(|a_{ij}(t,x)| + |b_i(t,x)| \big) dt \bigg]. \end{aligned} \tag{4.2.9}$$

By (A1.i), for any $\varepsilon > 0$, there is $\delta > 0$ independent of μ such that

$$t_1, t_2 \geqslant s,\ |t_1 - t_2| \leqslant \delta \implies \mathrm{d}_v(\mu_{t_1}, \mu_{t_2}) \leqslant \varepsilon.$$

Consequently $\mathcal{A}_{s,\zeta}$ is equicontinuous (even uniformly), which completes the proof. □

With the same proof, one can prove the existence of a solution flow with respect to $\mathcal{A}_{s,\zeta} = SP_{s,\zeta}$ for all $\zeta \in \mathcal{SP}$ (under the assumption that each of these sets is non-empty). The advantage of the choice (4.2.8) is that the corresponding flow consists of probability solutions.

Remark 4.2.11. *Estimate (4.2.9) is independent of the initial measure ζ, so we obtain even relative compactness of $\cup_{\zeta\in\mathcal{P}}\mathcal{A}_{s,\zeta}$.*

Nonlinear equations. Consider $\mathcal{B}((0,\infty))\otimes\tau_v\otimes\mathcal{B}(\mathbb{R}^d)$-measurable coefficients $a_{ij}, b_i : (0,\infty)\times\mathcal{SP}\times\mathbb{R}^d \to \mathbb{R}$, satisfying

Assumption A2.

(A2.i) $(t,\zeta,x)\mapsto a_{ij}(t,\zeta,x)$ and $(t,\zeta,x)\mapsto b_i(t,\zeta,x)$ are bounded on $(0,T)\times\mathcal{SP}\times\mathbb{R}^d$ for all $T>0$.

(A2.ii) $x\mapsto a_{ij}(t,\zeta,x)$ and $x\mapsto b_i(t,\zeta,x)$ are continuous for each $\zeta\in\mathcal{SP}$ and dt-a.e. $t>0$.

(A2.iii) If $\zeta_n \longrightarrow \zeta$ vaguely in $\mathcal{SP}$, then $a_{ij}(t,\zeta_n,x)\longrightarrow a_{ij}(t,\zeta,x)$ and $b_i(t,\zeta_n,x)\longrightarrow b_i(t,\zeta,x)$ locally uniformly in $x\in\mathbb{R}^d$ for each $t>0$.

Note that (A2.iii) excludes the case of Nemytskii-coefficients.

Let $\mathcal{A}_{s,\zeta}$ be as in (4.2.8). As in the linear case, under Assumption A2 we have $SP_{s,\zeta} = SP^1_{s,\zeta}$ for all $\zeta\in\mathcal{P}$.

Proposition 4.2.12. *Suppose Assumption A2 is fulfilled and $SP_{s,\zeta}$ is nonempty for each $(s,\zeta)\in[0,\infty)\times\mathcal{P}$. Then there exists a solution flow for the NLFPE in $\{\mathcal{A}_{s,\zeta}\}_{s\geqslant 0,\zeta\in\mathcal{SP}}$.*

Proof. Set $(H,\tau) = (\mathcal{SP},\tau_v)$. As in the linear case, we use the Arzelá-Ascoli theorem 4.2.8 and Theorem 4.2.3, and we prove compactness of $\mathcal{A}_{s,\zeta}\subseteq C_sH$ even with respect to τ_{lu}. Again, pointwise relative compactness follows from the compactness of $(\mathcal{SP},\tau_v)$. Equicontinuity of $\mathcal{A}_{s,\zeta}$ can be prove exactly as in the linear case, using (A2.i) instead of (A1.i). For closedness, assume a sequence $\mu^{(n)} = (\mu^{(n)}_t)_{t\geqslant s}$ from $\mathcal{A}_{s,\zeta}$ τ_{lu}-converges to $\mu=(\mu_t)_{t\geqslant s}$ in $C_s\mathcal{SP}$. We need to prove

$$\int_s^t\int_{\mathbb{R}^d} L_{a,b,\mu^{(n)}_r}\varphi\, d\mu^{(n)}_r dr \underset{n\to\infty}{\longrightarrow} \int_s^t\int_{\mathbb{R}^d} L_{a,b,\mu_r}\varphi\, d\mu_r dr \tag{4.2.10}$$

for each $\varphi\in C^2_c(\mathbb{R}^d)$ and $t\geqslant s$. This follows since

$$\int_{\mathbb{R}^d} L_{a,b,\mu^{(n)}_t}\varphi(t)\, d\mu^{(n)}_t = {}_{C_0^*}\langle \mu^{(n)}_t, L_{a,b,\mu^{(n)}_t}\varphi(t)\rangle_{C_0},$$

where ${}_{C_0^*}\langle\mu,f\rangle_{C_0}$ denotes the dual pairing between $f\in(C_0(\mathbb{R}^d),||\cdot||_\infty)$ and a finite Borel measure μ, understood as an element in the dual space of $C_0(\mathbb{R}^d)$. Since τ_v coincides with the weak-* topology on the topological dual of $C_0(\mathbb{R}^d)$, and since assumptions (A2.ii) and (A2.iii) yield $L_{a,b,\mu_t^{(n)}}\varphi(t) \longrightarrow L_{a,b,\mu_t}\varphi(t)$ in $(C_0(\mathbb{R}^d), ||\cdot||_\infty)$ for each $t \geqslant s$, we get

$$ {}_{C_0^*}\big\langle \mu_t^{(n)}, L_{a,b,\mu_t^{(n)}}\varphi(t)\big\rangle_{C_0} \longrightarrow {}_{C_0^*}\big\langle \mu_t, L_{a,b,\mu_t}\varphi(t)\big\rangle_{C_0}. $$

Now (4.2.10) follows by (A2.i) and Lebesgue's dominated convergence theorem. □

Nemytskii-type coefficients. Under suitable assumptions on the coefficients, Theorem 4.2.3 applies also in the Nemytskii-case. For an example, please see Section 4.2.2. in [17] (Reading exercise 11.2).

5 Nonlinear Markov processes

As said at the end of Chapter 3, the family of solution laws to a well-posed DDSDE does not satisfy the Markov property (recall that for us the Markov property is a property for a *family* of path laws and not for a single stochastic process [that is, for a single law]). Now we present a generalized definition of Markov processes, tailored to apply to such (and, in fact, much more general, i.e. ill-posed) cases. Thereby, we complete the nonlinear analog of the relations between linear FPEs, SDES and Markov processes. The main reference for the content of this chapter is [18].

We write $\Omega_s := C([s,\infty),\mathbb{R}^d)$ (with the topology of locally uniform convergence), $\pi^s_t, t \geqslant s$, for the usual projections on Ω_s, and $\mathcal{F}_{s,r} := \sigma(\pi^s_\tau, s \leqslant \tau \leqslant r)$. We also denote by $\Pi^s_r : \Omega_s \to \Omega_r$ the path projections $\Pi^s_r : w \mapsto w_{|[r,\infty)}$ for $s \leqslant r$.

5.1 Definition, basic properties, relation to classical Markov processes

Below, one should think of $\mathcal{P}_0$ as the class of "allowed initial data".

Definition 5.1.1. Let $\mathcal{P}_0 \subseteq \mathcal{P}$. A *nonlinear Markov process* is a family $(\mathbb{P}_{s,\zeta})_{(s,\zeta)\in\mathbb{R}_+\times\mathcal{P}_0}$ such that $\mathbb{P}_{s,\zeta}$ is a probability measure on $\mathcal{B}(\Omega_s)$ with the properties

(i) $\mu^{s,\zeta}_t := \mathbb{P}_{s,\zeta} \circ (\pi^s_t)^{-1} \in \mathcal{P}_0$ for all $0 \leqslant s \leqslant t$ and $\zeta \in \mathcal{P}_0$.

(ii) The *nonlinear Markov property* holds, i.e. for all $0 \leqslant s \leqslant r \leqslant t$, $\zeta \in \mathcal{P}_0$ and $A \in \mathcal{B}(\mathbb{R}^d)$

$$\mathbb{P}_{s,\zeta}(\pi^s_t \in A|\mathcal{F}_{s,r})(\cdot) = p_{(s,\zeta),(r,\pi^s_r(\cdot))}(\pi^r_t \in A) \quad \mathbb{P}_{s,\zeta} - \text{a.s.}, \tag{5.1.1}$$

where $p_{(s,\zeta),(r,y)}, y \in \mathbb{R}^d$, is the disintegration-family of $\mathbb{P}_{r,\mu^{s,\zeta}_r}$ w.r.t. π^r_r (i.e. in particular $p_{(s,\zeta),(r,y)} \in \mathcal{P}(\Omega_r)$ and $p_{(s,\zeta),(r,y)}(\pi^r_r = y) = 1$).

Note that $\Omega_s \times \mathcal{B}(\Omega_r) \ni (\omega, C) \mapsto p_{(s,\zeta),(r,\pi^s_r(\omega))}(C)$ is equal to the regular conditional probability of $\mathbb{P}_{s,\zeta}$ w.r.t. π^s_r, restricted to $\sigma(\pi^s_u, u \geqslant r)$ (by identifying the latter σ-algebra with $\mathcal{B}(\Omega_r)$) (Exercise 12.1).

The name "*nonlinear* Markov property" stems from the fact that in usual applications the family $\{\mu^{s,\zeta}_t\}_{0\leqslant s\leqslant t,\zeta\in\mathcal{P}_0}$ is a family of solutions to a *nonlinear* FPE.

Proposition 5.1.2. *The one-dimensional time marginals $\mu^{s,\zeta}_t = \mathbb{P}_{s,\zeta} \circ (\pi^s_t)^{-1}$ of a nonlinear Markov process satisfy the* flow property*.*

Proof. We have for all $A \in \mathcal{B}(\mathbb{R}^d)$ and $0 \leqslant s \leqslant r \leqslant t$:

$$\begin{aligned}\mu^{s,\zeta}_t(A) &= \mathbb{E}_{s,\zeta}\left[\mathbb{P}_{s,\zeta}(\pi^s_t \in A|\mathcal{F}_{s,r})\right]\\ &= \mathbb{E}_{s,\zeta}\left[p_{(s,\zeta),(r,\pi^s_r)}(\pi^r_t \in A)\right] = \mathbb{P}_{r,\mu^{s,\zeta}_r}\left(\pi^r_t \in A\right) = \mu^{r,\mu^{s,\zeta}_r}_t(A).\end{aligned}$$

□

Remark 5.1.3. *In contrast to the case of classical Markov processes, it is in general not true that the marginals $\mu_t^{s,\zeta}$ satisfy the (time-inhomogeneous version of the) Chapman–Kolmogorov equations* (2.1.3).

The following proposition shows that the finite-dimensional marginals of the path laws of a nonlinear Markov process (and hence the path laws themselves) are uniquely determined by the family of one-dimensional time marginals $p_{r,t}^{s,\zeta}(x,dz)$, $s \leqslant r, x \in \mathbb{R}^d$, defined in (5.1.2) below.

Proposition 5.1.4. *Let $(\mathbb{P}_{s,\zeta})_{(s,\zeta)\in\mathbb{R}_+\times\mathcal{P}_0}$ be a nonlinear Markov process. For $\zeta \in \mathcal{P}_0, 0 \leqslant s \leqslant r \leqslant t$ and $x \in \mathbb{R}^d$, define $p_{r,t}^{s,\zeta}(x,dz) \in \mathcal{P}$ by*

$$p_{r,t}^{s,\zeta}(x,dz) := p_{(s,\zeta),(r,x)} \circ (\pi_t^r)^{-1}(dz), \tag{5.1.2}$$

which is uniquely determined for $\mu_r^{s,\zeta}$-a.e. $x \in \mathbb{R}^d$. Then for any $n \in \mathbb{N}_0$, $f \in \mathcal{B}_b((\mathbb{R}^d)^{n+1})$ and $s \leqslant t_0 < \cdots < t_n$:

$$\begin{aligned}&\mathbb{E}_{s,\zeta}[f(\pi_{t_0}^s,\dots,\pi_{t_n}^s)]\\ &= \int_{\mathbb{R}^d}\Big(\cdots\int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d} f(x_0,\dots,x_n)\, p_{t_{n-1},t_n}^{s,\zeta}(x_{n-1},dx_n)\Big)p_{t_{n-2},t_{n-1}}^{s,\zeta}(x_{n-2},dx_{n-1})\dots\Big)\mu_{t_0}^{s,\zeta}(dx_0).\end{aligned}$$

Proof. Exercise 12.2. □

Remark 5.1.5. *Even in the case $\mathcal{P}_0 = \mathcal{P}$ it is usually* not *true that $p_{r,t}^{s,\zeta}(x,dz) = \mathbb{P}_{r,\delta_x} \circ (\pi_t^r)^{-1}(dz)$, i.e. the family of one-dimensional time marginals needed to determine the path measures of a nonlinear Markov process is* not *the family of its one-dimensional time marginals, but a "bigger" one, see Remark 5.1.7 below.*

The following result shows that the class of nonlinear Markov processes contains the class of classical normal Markov processes. Let $(\mathbb{P}_{s,x})_{s\geqslant 0,x\in\mathbb{R}^d}$ be a classical normal time-inhomogeneous Markov process and set $\mathbb{P}_{s,\zeta} := \int_{\mathbb{R}^d}\mathbb{P}_{s,x}d\zeta(x)$, $\zeta \in \mathcal{P}$.

Proposition 5.1.6. *$(\mathbb{P}_{s,\zeta})_{(s,\zeta)\in\mathbb{R}_+\times\mathcal{P}}$ is a nonlinear Markov process with $\mathcal{P}_0 = \mathcal{P}$.*

Proof. We have $\mathbb{P}_{r,\mu_r^{s,\zeta}} = \int_{\mathbb{R}^d}\mathbb{P}_{r,y}\,\mu_r^{s,\zeta}(dy)$, $y \mapsto \mathbb{P}_{r,y}(A)$ is measurable for every $A \in \Omega_r$ and, by normality, $\mathbb{P}_{r,y}$ is concentrated on $\{\pi_r^r = y\}$. Hence $\mathbb{P}_{r,y}, y \in \mathbb{R}^d$, is the disintegration family of $\mathbb{P}_{r,\mu_r^{s,\zeta}}$ w.r.t. π_r^r, and thus (5.1.1) holds with $p_{(s,\zeta),(r,\pi_r^s(\cdot))} = \mathbb{P}_{r,\pi_r^s(\cdot)}$, which is the classical Markov property. □

Remark 5.1.7. *If $\{\mathbb{P}_{s,\zeta}\}_{(s,\zeta)\in\mathbb{R}_+\times\mathcal{P}_0}$ is a nonlinear Markov process, consisting of solution laws to a DDSDE, then its one-dimensional time marginal curves $(\mu_t^{s,\zeta})_{t\geqslant s}$, $\mu_t^{s,\zeta} = \mathbb{P}_{s,\zeta}\circ(\pi_t^s)^{-1}$, solve the associated NLFPE, and the curves $(p_{r,t}^{s,\zeta}(x,dz))_{t\geqslant r}$ from* (5.1.2) *are weakly continuous probability solutions to ($\mu^{s,\zeta}$-ℓFPE) with initial datum (r,δ_x) for $\mu_r^{s,\zeta}$-a.e. x. The latter follows from Lemma 1.3.2 (i) and Corollary 1.3.5.*

Hence, if for all $(s,\zeta) \in \mathbb{R}_+\times\mathcal{P}_0$ the equation ($\mu^{s,\zeta}$-ℓFPE) has a unique weakly continuous probability solution for every initial datum $(r,x) \in [s,\infty)\times\mathbb{R}^d$, then ($(s,\zeta)$ fixed) $p_{r,t}^{s,\zeta}$, $s \leqslant r \leqslant t$, are the transition kernels of a linear time-inhomogeneous

Markov process $\{P^{s,\zeta}_{r,x}\}_{(r,x)\in[s,\infty)\times\mathbb{R}^d}$*, see Theorem 2.2.2. The family of these processes is related to the nonlinear Markov process by*

$$\mathbb{P}_{r,\mu^{s,\zeta}_r} = \int_{\mathbb{R}^d} P^{s,\zeta}_{r,x}\, d\mu^{s,\zeta}_r(x), \quad \forall 0 \leqslant s \leqslant r, \zeta \in \mathcal{P}_0$$

(i.e. the RHS is the convex mixture of the path laws of the linear Markov processes). In this case, Proposition 5.1.4 shows that the finite-dimensional marginals of $\mathbb{P}_{s,\zeta}$ *(and hence* $\mathbb{P}_{s,\zeta}$ *itself) are uniquely determined by the transition kernels of a linear Markov process, which depends, however, on* (s,ζ)*.*

5.2 Construction of nonlinear Markov processes

As before, we refer to (3.1.1) and the related stochastic equation (3.3.1) as "the NLFPE" and "the DDSDE". We stress that here we do not impose any regularity on the coefficients, i.e. in particular Nemytskii-type coefficients are included in the theory presented below.

We introduce the following notation [not to be confused with the notation $M^{s,\zeta}$ in the proof of Theorem 4.2.3]. For $(s,\zeta) \in \mathbb{R}_+ \times \mathcal{P}$, we denote the space of weakly continuous probability solutions μ to the NLFPE from (s,ζ) satisfying

$$[(t,x) \mapsto a_{ij}(t,\mu_t,x)], [(t,x) \mapsto b_i(t,\mu_t,x)] \in L^1([0,T]\times\mathbb{R}^d; \mu_t dt), \quad \forall T > 0$$

by $M^{s,\zeta}$. For a weakly continuous curve $\eta : [s,\infty) \in t \mapsto \eta_t \in \mathcal{P}$, we write $M^{s,\zeta}_\eta$ for the set of all weakly continuous probability solutions μ to (η-ℓFPE) from (s,ζ) satisfying for all $i,j \leqslant d$

$$[(t,x) \mapsto a_{ij}(t,\eta_t,x)], [(t,x) \mapsto b_i(t,\eta_t,x)] \in L^1([0,T]\times\mathbb{R}^d; \mu_t dt), \quad \forall T > 0.$$

Recall that μ is an *extreme point* of the convex set $M^{s,\zeta}_\eta$, if $\mu \in M^{s,\zeta}_\eta$ and $\mu = \alpha\mu^1 + (1-\alpha)\mu^2$ for $\alpha \in (0,1)$ and $\mu^1,\mu^2 \in M^{s,\zeta}_\eta$ implies $\mu^1 = \mu^2$. The set of extreme points of $M^{s,\zeta}_\eta$ is denoted by $M^{s,\zeta}_{\eta,\mathrm{ex}}$.

Theorem 5.2.1 (R.-Röckner-nonlinear-Markov-construction). *Let* $\mathcal{P}_0 \subseteq \mathcal{P}$ *and* $\{\mu^{s,\zeta}\}_{(s,\zeta)\in\mathbb{R}_+\times\mathcal{P}_0}$ *be a solution flow to the NLFPE such that* $\mu^{s,\zeta} \in M^{s,\zeta}_{\mu^{s,\zeta},ex}$ *for each* $(s,\zeta) \in \mathbb{R}_+ \times \mathcal{P}_0$*. Then:*

(i) For every $(s,\zeta) \in \mathbb{R}_+ \times \mathcal{P}_0$*, there is a unique weak solution* $X^{s,\zeta}$ *to the DDSDE with initial condition* (s,ζ) *and one-dimensional time marginals equal to* $(\mu^{s,\zeta}_t)_{t\geqslant s}$*.*

(ii) $\{\mathbb{P}_{s,\zeta}\}_{s\geqslant 0,\zeta\in\mathcal{P}_0}$*,* $\mathbb{P}_{s,\zeta} := \mathcal{L}_{X^{s,\zeta}}$*, is a nonlinear Markov process. In particular, its one-dimensional time marginals are* $\mu^{s,\zeta}_t$*,* $0 \leqslant s \leqslant t$*,* $\zeta \in \mathcal{P}_0$*.*

It should be noted that for a solution flow $\{\mu^{s,\zeta}\}_{(s,\zeta)\in\mathbb{R}_+\times\mathcal{P}_0}$ it holds $\mu^{s,\zeta}_t \in \mathcal{P}_0$ for all $0 \leqslant s \leqslant t, \zeta \in \mathcal{P}_0$.

Remark 5.2.2. *(i) Assertion (i) does* not *mean that there is a unique weak solution which additionally satisfies the stated marginal-property, but instead that the subclass of solutions with this marginal property contains exactly one element.*

(ii) Note that the theorem does not *require any uniqueness for the NLFPE. Of course, if the NLFPE is well-posed in $\mathcal{P}_0$, its unique solution family has the flow property, but in the absence of uniqueness, a flow may still be obtained by the methods presented in the previous chapter.*

For the proof, we need the following auxiliary result, which, in view of applications, provides a checkable characterization of the extremality condition of the previous theorem. For a $\mathcal{P}$-valued curve $\mu = (\mu_t)_{t\geqslant s}$ and $C > 0$ set

$$\mathcal{A}_{s,\leqslant}(\mu, C) := \big\{(\eta_t)_{t\geqslant s} \in C([s,\infty),\mathcal{P}) : \eta_t \leqslant C\mu_t,\ \forall t \geqslant s\big\},\quad \mathcal{A}_{s,\leqslant}(\mu) := \bigcup_{C>0} \mathcal{A}_{s,\leqslant}(\mu, C),$$

where continuity is meant w.r.t. the topology of weak convergence of measures.

Lemma 5.2.3. *Let $(s,\zeta) \in \mathbb{R}_+ \times \mathcal{P}$, $\eta \in C([s,\infty),\mathcal{P})$ and $\mu = (\mu_t)_{t\geqslant s} \in M^{s,\zeta}_\eta$. Then*

$$|(M^{s,\zeta}_\eta \cap \mathcal{A}_{s,\leqslant}(\mu))| = 1 \iff \mu \in M^{s,\zeta}_{\eta,ex}.$$

By considering coefficients which do not depend on their measure variable, it is clear that the previous lemma holds in the case of a linear FPE as well (in this case, η disappears from the formulation).

Proof. Clearly, $\mu \in M^{s,\zeta}_\eta \cap \mathcal{A}_{s,\leqslant}(\mu)$. First, suppose $\mu \notin M^{s,\zeta}_{\eta,\mathrm{ex}}$, i.e. there are $\mu^i, i \in \{1,2\}$, in $M^{s,\zeta}_\eta$ and $\alpha \in (0,1)$ such that

$$\mu_t = \alpha\mu^1_t + (1-\alpha)\mu^2_t, \quad t \geqslant s, \tag{5.2.1}$$

and $\mu^1 \neq \mu^2$. Then (5.2.1) implies $\mu^i \in M^{s,\zeta}_\eta \cap \mathcal{A}_{s,\leqslant}(\mu)$, $i \in \{1,2\}$, and hence $|(M^{s,\zeta} \cap \mathcal{A}_{s,\leqslant}(\mu))| \geqslant 2$.

Now assume $\mu \in M^{s,\zeta}_{\eta,\mathrm{ex}}$ and let $\nu \in M^{s,\zeta}_\eta \cap \mathcal{A}_{s,\leqslant}(\mu)$. Then for every $t \geqslant s$ there is $\varrho_t : \mathbb{R}^d \to \mathbb{R}_+$, $\mathcal{B}(\mathbb{R}^d)$-measurable, such that $\nu_t = \varrho_t\,\mu_t$, and $\varrho_t \leqslant C$ for all $t \geqslant s$ for some $C \in (1,\infty)$. Furthermore, for $t \geqslant s$,

$$\mu_t = \frac{1}{C}\varrho_t\,\mu_t + (1 - \frac{1}{C}\varrho_t)\,\mu_t = \frac{1}{C}\nu_t + (1-\frac{1}{C})\lambda_t,$$

where $\lambda_t := \frac{1-\frac{1}{C}\varrho_t}{1-\frac{1}{C}}\,\mu_t$. Clearly, for each $t \geqslant s$, the measure λ_t is nonnegative and satisfies $\lambda_t(\mathbb{R}^d) = 1$. Moreover $t \mapsto \lambda_t$ is weakly continuous and belongs to $M^{s,\zeta}_\eta$. Since $\mu^{s,\zeta} \in M^{s,\zeta}_{\eta,\mathrm{ex}}$, it follows $\mu_t = \lambda_t$, and hence $\nu_t = \mu_t$ for all $t \geqslant s$. □

As a further preparation, we need part (ii) of the following lemma. Part (i) is not used here, but may be of independent interest. If two nonnegative Borel measures ζ_1, ζ_2 satisfy $\zeta_1 \ll \zeta_2$ and $\zeta_2 \ll \zeta_1$, we write $\zeta_1 \sim \zeta_2$.

Note: In the following lemma, by "solution" we mean weakly continuous probability solutions satisfying (1.3.2).

Lemma 5.2.4. *Consider a linear FPKE with initial datum* $(s_0, \zeta_0) \in \mathbb{R}_+ \times \mathcal{P}$*. Then:*

(i) If solutions are unique from (s_0, ζ_0)*, then solutions are also unique from any* (s_0, η) *such that* $\eta \in \mathcal{P}$*,* $\eta \sim \zeta_0$*.*

(ii) If $(\nu_t^{s_0,\zeta_0})_{t \geqslant s}$ *is the unique solution in* $\mathcal{A}_{s_0,\leqslant}(\nu^{s_0,\zeta_0})$ *from* (s_0, ζ_0)*, then in this class solutions are also unique from any* $(s_0, g\,\zeta_0)$ *with* $g \in \mathcal{B}_b^+(\mathbb{R}^d)$*,* $\int_{\mathbb{R}^d} g(x)\,\zeta_0(dx) = 1$*, and* $\delta \leqslant g$ *for some* $\delta > 0$*.*

The proof can be found as the proof of Lemma 3.7. in [18].

Proof of Theorem 5.2.1.

We shall need the following auxiliary result, which is taken from [21], see Proposition 2.6. therein.

Lemma 5.2.5. *Let* $0 \leqslant s \leqslant r$*,* $P \in \mathcal{P}(\Omega_s)$ *a solution to a linear martingale problem with initial time* s*, and* $\varrho : \Omega_s \to \mathbb{R}_+$ *a bounded* $\mathcal{F}_{s,r}$*-measurable probability density (w.r.t.* P*). Then* $(\varrho P) \circ (\Pi_r^s)^{-1}$ *solves the same martingale problem with initial time* r*.*

We can now prove Theorem 5.2.1.

Proof of Theorem 5.2.1. (i) The existence of a weak solution $X^{s,\zeta}$ to the DDSDE for each initial datum (s, ζ) follows from Theorem 3.4.1. Concerning uniqueness, note that by assumption and Lemma 5.2.3, for each $0 \leqslant s \leqslant r$, $\zeta \in \mathcal{P}_0$, ($\mu^{s,\zeta}$-ℓFPE) has a unique solution from $(r, \mu_r^{s,\zeta})$ in $\mathcal{A}_{r,\leqslant}(\mu^{s,\zeta})$ (this solution is $(\mu_t^{s,\zeta})_{t \geqslant r}$).

Claim: For any $(s, \zeta) \in \mathbb{R}_+ \times \mathcal{P}_0$ and $r \geqslant s$, solutions to the corresponding linear martingale problem with one-dimensional time marginals in $\mathcal{A}_{r,\leqslant}(\mu^{s,\zeta})$ are unique from $(r, \mu_r^{s,\zeta})$.

Proof of Claim: Fix $(s, \zeta) \in \mathbb{R}_+ \times \mathcal{P}_0$, $r \geqslant s$, and let P^1, P^2 be such solutions. Their one-dimensional time marginal curves $(P_t^i)_{t \geqslant r}$,

$$P_t^i := P^i \circ (\pi_t^r)^{-1},$$

solve ($\mu^{s,\zeta}$-ℓFPE) from $(r, \mu_r^{s,\zeta})$, and hence

$$P_t^i = \mu_t^{s,\zeta}, \ \forall\, t \geqslant r, \quad i \in \{1, 2\}. \tag{5.2.2}$$

For $n \in \mathbb{N}$, let

$$\mathcal{H}_r^{(n)} := \{\Pi_{i=1}^n h_i(\pi_{t_i}^r) \,|\, h_i \in \mathcal{B}_b^+(\mathbb{R}^d), h_i \geqslant c_i \text{ for some } c_i > 0, r \leqslant t_1 < \cdots < t_n\},$$

$$\mathcal{H}_r := \bigcup_{n \in \mathbb{N}} \mathcal{H}_r^{(n)}$$

and note that $\mathcal{H}_r$ is closed under pointwise multiplication and $\sigma(\mathcal{H}_r) = \mathcal{B}(\Omega_r)$. Hence, by induction in $n \in \mathbb{N}$ and a monotone class argument, it suffices to prove

$$\mathbb{E}_{P^1}[H] = \mathbb{E}_{P^2}[H] \text{ for all } H \in \mathcal{H}_r^{(n)} \tag{5.2.3}$$

for each $n \in \mathbb{N}$. For $n = 1$, (5.2.3) holds by (5.2.2). For the induction step from n to $n+1$, fix $r \leqslant t_1 < \cdots < t_n < t_{n+1}$ and functions $h_i, i \in \{1, \dots, n+1\}$, as specified in the definition of $\mathcal{H}_r^{(n+1)}$, and set

$$\varrho : \Omega_r \to \mathbb{R}_+, \quad \varrho := \frac{\Pi_{i=1}^n h_i(\pi_{t_i}^r)}{\mathbb{E}_{P^1}[\Pi_{i=1}^n h_i(\pi_{t_i}^r)]},$$

where the denominator is greater or equal to $\Pi_{i=1}^n c_i > 0$. Note that ϱ is $\mathcal{F}_{r,t_n}$-measurable and

$$\frac{1}{c} \leqslant \varrho \leqslant c \text{ pointwise for some } c > 1, \tag{5.2.4}$$

(where c depends on h_i, t_i and n) and $\mathbb{E}_{P^i}[\varrho] = 1$ for $i \in \{1,2\}$ by the induction hypothesis. Since for every $f \in \mathcal{B}_b^+(\mathbb{R}^d)$ we have

$$\int_{\Omega_r} f(\pi_{t_n}^r)\,(\varrho P^i) = \Big[\int_{\Omega_r} \Pi_{i=1}^n h_i(\pi_{t_i}^r) f(\pi_{t_n}^r)\, P^i\Big] \big(\mathbb{E}_{P^1}[\Pi_{i=1}^n h_i(\pi_{t_i}^r)]\big)^{-1},$$

and since the induction hypothesis implies that these terms are equal for $i \in \{1,2\}$, it follows that

$$(\varrho\, P^1) \circ (\pi_{t_n}^r)^{-1} = (\varrho\, P^2) \circ (\pi_{t_n}^r)^{-1}. \tag{5.2.5}$$

By Lemma 5.2.5, the path measures $(\varrho P^i) \circ (\Pi_{t_n}^r)^{-1}$, $i \in \{1,2\}$, on $\mathcal{B}(\Omega_{t_n})$ solve the same linear martingale problem from time t_n and, by (5.2.5), with identical initial condition. Consequently, their curves of one-dimensional time marginals $\eta^i = (\eta_t^i)_{t \geqslant t_n} := ((\varrho\, P^i) \circ (\pi_t^r)^{-1})_{t \geqslant t_n}$, $i \in \{1,2\}$, solve ($\mu^{s,\zeta}$-ℓFPE). For any $A \in \mathcal{B}(\mathbb{R}^d)$ and $t \geqslant t_n$, we have by (5.2.2)

$$\eta_t^i(A) = \int_{\Omega_r} \varrho(w) 1_A(\pi_t^r(w))\, P^i(dw) \leqslant c\, P_t^i(A) = c\, \mu_t^{s,\zeta}(A), \quad i \in \{1,2\},$$

for c as in (5.2.4), and consequently $\eta^i \in \mathcal{A}_{t_n,\leqslant}(\mu^{s,\zeta})$. Similarly, $\eta_t^i(A) \geqslant \frac{1}{c}\mu_t^{s,\zeta}(A)$ for all $A \in \mathcal{B}(\mathbb{R}^d)$ and $t \geqslant t_n$. In particular, for $t = t_n$, it follows that $\eta_{t_n}^i = g^i\, \mu_{t_n}^{s,\zeta}$ for some measurable $g^i : \mathbb{R}^d \to \mathbb{R}_+$ such that $\frac{1}{c} \leqslant g^i \leqslant c$, and $\int_{\mathbb{R}^d} g^i\, d\mu_{t_n}^{s,\zeta} = 1$. By assumption, Lemma 5.2.3 and 5.2.4 (ii), we obtain $(\eta_t^1)_{t \geqslant t_n} = (\eta_t^2)_{t \geqslant t_n}$, so in particular $\eta_{t_{n+1}}^1 = \eta_{t_{n+1}}^2$. Now we have

$$\frac{\mathbb{E}_{P^i}\big[\Pi_{i=1}^{n+1} h_i(\pi_{t_i}^r)\big]}{\mathbb{E}_{P^1}\big[\Pi_{i=1}^{n} h_i(\pi_{t_i}^r)\big]} = \int_{\Omega_r} \varrho(w) h_{n+1}(\pi_{t_{n+1}}^r(w))\, P^i(dw) = \int_{\mathbb{R}^d} h_{n+1}(x)\, \eta_{t_{n+1}}^i(dx)$$

for $i \in \{1,2\}$, and conclude

$$\mathbb{E}_{P^1}\big[\Pi_{i=1}^{n+1} h_i(\pi_{t_i}^r)\big] = \mathbb{E}_{P^2}\big[\Pi_{i=1}^{n+1} h_i(\pi_{t_i}^r)\big],$$

which gives (5.2.3) for $n+1$, and hence completes the proof of the claim.

Since $\mu^{s,\zeta} \in \mathcal{A}_{\leqslant,s}(\mu^{s,\zeta})$, the assertion now follows from the equivalence of the linear martingale problem and the associated SDE.

(ii) The family $(\mathbb{P}_{s,\zeta})_{(s,\zeta)\in\mathbb{R}_+\times\mathcal{P}_0}$ satisfies

(i) $\mathbb{P}_{s,\zeta}\in\mathcal{P}(\Omega_s)$ and $\mathbb{P}_{s,\zeta}\circ(\pi^s_t)^{-1}=\mu^{s,\zeta}_t$ for all $t\geqslant s$,

(ii) $\mathbb{P}_{s,\zeta}$ is the path law of the unique weak DDSDE solution with one-dimensional time marginals $(\mu^{s,\zeta}_t)_{t\geqslant s}$.

To prove the nonlinear Markov property, let $0\leqslant s\leqslant r\leqslant t$ and $\zeta\in\mathcal{P}_0$. Disintegrating $\mathbb{P}_{r,\mu^{s,\zeta}_r}$ with respect to π^r_r yields

$$\mathbb{P}_{r,\mu^{s,\zeta}_r}(\cdot)=\int_{\mathbb{R}^d}p_{(s,\zeta),(r,y)}(\cdot)\,\mu^{s,\zeta}_r(dy) \tag{5.2.6}$$

as measures on $\mathcal{B}(\Omega_r)$, where the $\mu^{s,\zeta}_r$-almost surely determined family $p_{(s,\zeta),(r,y)}$, $y\in\mathbb{R}^d$, of Borel probability measures on Ω_r is as in Definition 5.1.1.

By Lemma 1.3.2, for $\mu^{s,\zeta}_r$-a.e. $y\in\mathbb{R}^d$, $p_{(s,\zeta),(r,y)}$ solves the $\mu^{s,\zeta}$-linearized martingale problem from (r,δ_y). Hence, for any $\varrho\in\mathcal{B}^+_b(\mathbb{R}^d)$ with $\int_{\mathbb{R}^d}\varrho\,d\mu^{s,\zeta}_r=1$, the measure $\mathbb{P}_\varrho\in\mathcal{P}(\Omega_r)$,

$$\mathbb{P}_\varrho:=\int_{\mathbb{R}^d}p_{(s,\zeta),(r,y)}\,\varrho(y)\,d\mu^{s,\zeta}_r(dy), \tag{5.2.7}$$

solves the same linearized martingale problem with initial datum $(r,\varrho\,\mu^{s,\zeta}_r)$. Let $n\in\mathbb{N}$, $s\leqslant t_1<\cdots<t_n\leqslant r$, $h\in\mathcal{B}^+_b((\mathbb{R}^d)^n)$ such that $a^{-1}\leqslant h\leqslant a$ for some $a>1$, and let $\tilde g:\mathbb{R}^d\to\mathbb{R}_+$ be the bounded, $\mu^{s,\zeta}_r$-a.s. uniquely determined map such that

$$\mathbb{E}_{s,\zeta}\big[h(\pi^s_{t_1},\dots,\pi^s_{t_n})|\sigma(\pi^s_r)\big]=\tilde g(\pi^s_r)\quad \mathbb{P}_{s,\zeta}-a.s.$$

Let $g:=c_0\tilde g$, where $c_0>0$ is such that $\int_{\mathbb{R}^d}g\,d\mu^{s,\zeta}_r=1$, and let $\mathbb{P}_g$ be as in (5.2.7), with g replacing ϱ, with initial condition $(r,g\,\mu^{s,\zeta}_r)$. Also, consider $\theta:\Omega_s\to\mathbb{R}$, $\theta:=c_0h(\pi^s_{t_1},\dots,\pi^s_{t_n})$, i.e. $\mathbb{E}_{s,\zeta}[\theta]=1$. Set

$$\mathbb{P}^\theta:=(\theta\,\mathbb{P}_{s,\zeta})\circ(\Pi^s_r)^{-1}.$$

Note that $g(\mathbb{R}^d),\theta(\Omega_s)\subseteq[a^{-1}c_0,ac_0]$ $\mu^{s,\zeta}_r$-a.s., so in particular $g\,\mu^{s,\zeta}_r\sim\mu^{s,\zeta}_r$. By Lemma 5.2.5 , also $\mathbb{P}^\theta$ solves the same linearized martingale problem with initial datum $(r,g\mu^{s,\zeta}_r)$, since for all $A\in\mathcal{B}(\mathbb{R}^d)$

$$\begin{aligned}\mathbb{P}^\theta\circ(\pi^r_r)^{-1}(A)&=\int_{\Omega_s}1_A(\pi^s_r(w))\theta(w)\,\mathbb{P}_{s,\zeta}(dw)\\&=\int_{\Omega_s}1_A(\pi^s_r(w))g(\pi^s_r(w))\,\mathbb{P}_{s,\zeta}(dw)=\int_A g(y)\,\mu^{s,\zeta}_r(dy).\end{aligned}$$

In particular, both one-dimensional time marginal curves $(\mathbb{P}_g\circ(\pi^r_t)^{-1})_{t\geqslant r}$ and $(\mathbb{P}^\theta\circ(\pi^r_t)^{-1})_{t\geqslant r}$ solve ($\mu^{s,\zeta}$-ℓFPE) from $(r,g\,\mu^{s,\zeta}_r)$. Moreover,

$$\mathbb{P}^\theta\circ(\pi^r_t)^{-1},\,\mathbb{P}_g\circ(\pi^r_t)^{-1}\leqslant ac_0\,\mu^{s,\zeta}_t,\quad \forall t\geqslant r, \tag{5.2.8}$$

i.e. both these one-dimensional time marginal curves belong to $\mathcal{A}_{r,\leqslant}(\mu^{s,\zeta})$. Indeed, (5.2.8) can be seen as follows. For all $t \geqslant r$ and $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}\mathbb{P}^\theta \circ (\pi_t^r)^{-1}(A) &= \int_{\Omega_s} \theta(w) 1_A(\pi_t^s(w))\, \mathbb{P}_{s,\zeta}(dw) \\ &\leqslant c_0 a \int_{\Omega_s} 1_A(\pi_t^s(w))\, \mathbb{P}_{s,\zeta}(dw) = c_0 a\, \mu_t^{s,\zeta}(A).\end{aligned}$$

Similarly, by (5.2.6),

$$\mathbb{P}_g \circ (\pi_t^r)^{-1}(A) = \int_{\mathbb{R}^d} p_{(s,\zeta),(r,y)}(\pi_t^r \in A) g(y)\, \mu_r^{s,\zeta}(dy) \leqslant a c_0\, \mathbb{P}_{r,\mu_r^{s,\zeta}}(\pi_t^r \in A) = a c_0\, \mu_t^{s,\zeta}(A).$$

Hence by the assumption, Lemma 5.2.3 and Lemma 5.2.4 (ii)

$$\mathbb{P}_g \circ (\pi_t^r)^{-1} = \mathbb{P}^\theta \circ (\pi_t^r)^{-1}, \quad \forall t \geqslant r,$$

and therefore for $t \geqslant r$ and $A \in \mathcal{B}(\mathbb{R}^d)$

$$\begin{aligned}\mathbb{E}_{s,\zeta}\big[h(\pi_{t_1}^s, \dots, \pi_{t_n}^s) 1_{\pi_t^s \in A}\big] &= c_0^{-1} \mathbb{P}^\theta \circ (\pi_t^r)^{-1}(A) = c_0^{-1} \mathbb{P}_g \circ (\pi_t^r)^{-1}(A) \\ &= c_0^{-1} \int_{\Omega_s} p_{(s,\zeta),(r,\pi_r^s(\omega))}(\pi_t^r \in A) g(\pi_r^s(\omega))\, \mathbb{P}_{s,\zeta}(d\omega) \\ &= \int_{\Omega_s} p_{(s,\zeta),(r,\pi_r^s(\omega))}(\pi_t^r \in A) h(\pi_{t_1}^s(\omega), \dots, \pi_{t_n}^s(\omega))\, \mathbb{P}_{s,\zeta}(d\omega).\end{aligned}$$

Here we used the $\sigma(\pi_r^s)$-measurability of $\Omega_s \ni \omega \mapsto p_{(s,\zeta),(r,\pi_r^s(\omega))}(\pi_t^r \in A)$ for the final equality. By a monotone class-argument, (5.1.1) follows.

□

Since the nonlinear Markov property is always fulfilled for $s = r$ (!), the following corollary follows from the previous proof.

Corollary 5.2.6. *Let $\mathfrak{P}_0 \subseteq \mathcal{P}_0 \subseteq \mathcal{P}$ and let $\{\mu^{s,\zeta}\}_{(s,\zeta)\in\mathbb{R}_+\times\mathcal{P}_0}$ be a solution flow to the NLFPE such that $\mu_t^{s,\zeta} \in \mathfrak{P}_0$ for all $0 \leqslant s \leqslant t, \zeta \in \mathfrak{P}_0$ (then $\{\mu^{s,\zeta}\}_{(s,\zeta)\in\mathbb{R}_+\times\mathfrak{P}_0}$ is a solution flow) and $\mu_t^{s,\zeta} \in \mathfrak{P}_0$ for all $0 \leqslant s < t, \zeta \in \mathcal{P}_0$. Also assume $\mu^{s,\zeta} \in M^{s,\zeta}_{\mu^{s,\zeta},ex}$ for all $(s,\zeta) \in \mathbb{R}_+ \times \mathfrak{P}_0$.*

Then there exists a nonlinear Markov process $(\mathbb{P}_{s,\zeta})_{(s,\zeta)\in\mathbb{R}_+\times\mathcal{P}_0}$ such that $\mathbb{P}_{s,\zeta} \circ (\pi_t^s)^{-1} = \mu_t^{s,\zeta}$ for every $(s,\zeta) \in \mathbb{R}_+ \times \mathcal{P}_0$ and $t \geqslant s$, consisting of path laws of weak solutions to the corresponding DDSDE. Moreover, for $\zeta \in \mathfrak{P}_0$, $\mathbb{P}_{s,\zeta}$ is the path law of the unique *weak solution to the DDSDE with one-dimensional time marginals $(\mu_t^{s,\zeta})_{t\geqslant s}$.*

A typical application of Corollary 5.2.6 is as follows: $\mathfrak{P}_0 = \mathcal{P} \cap L^\infty$, $\mathcal{P}_0 = \mathcal{P}$, one has a solution flow $\{\mu^{s,\zeta}\}_{(s,\zeta)\in\mathbb{R}_+\times\mathcal{P}}$ to the NLFPE with $\mu^{s,\zeta} \in \bigcap_{\delta>s} L^\infty((\delta,\infty), L^\infty)$ (also called $L^1 - L^\infty$-regularization), and for every initial datum $(s,\zeta) \in \mathbb{R}_+ \times \mathfrak{P}_0$, solutions to the NLFPE are unique in $\bigcap_{T>s} L^\infty((s,T)\times\mathbb{R}^d)$. From the latter property, one can often prove that the corresponding linearized equations ($\mu^{s,\zeta}$-ℓFPE) have a unique solution in $\bigcap_{T>s} L^\infty((s,T)\times\mathbb{R}^d)$ from $(s,\zeta) \in \mathbb{R}_+ \times \mathfrak{P}_0$. Then the extremality-assumption of the corollary is satisfied.

5.3 Applications to nonlinear FPEs and PDEs

We consider some general and explicit cases of NLFPEs to which Theorem 5.2.1 or Corollary 5.2.6 apply.

(i) **Well-posed equations.** If the NLFPE has a unique weakly continuous probability solution $\mu^{s,\zeta}$ with the previously mentioned global in space integrability condition from every initial datum $(s,\zeta) \in \mathbb{R}_+ \times \mathcal{P}$ (or $\mathcal{P}_0$), and each linearized equation ($\mu^{s,\zeta}$-ℓFPE) has a unique weakly continuous probability solution from (s,ζ), Theorem 5.2.1 applies and yields the existence of a uniquely determined nonlinear Markov process with one-dimensional time marginals $\mu_t^{s,\zeta}$, $0 \leqslant s \leqslant t$, $\zeta \in \mathcal{P}$ ($\mathcal{P}_0$).

We stress again that these strong well-posedness results can typically not be proven for equations with Nemytskii-type coefficients.

(ii) **Generalized PME.** Consider

$$\partial_t u(t) = \Delta\beta(u) - \operatorname{div}\big(DB(u(t))u(t)\big) \tag{5.3.1}$$

under the following assumptions.

(B1) $\beta(0) = 0, \beta \in C^2(\mathbb{R}), \beta' \geqslant 0.$

(B2) $B \in C^1(\mathbb{R}) \cap C_b(\mathbb{R}), B \geqslant 0.$

(B3) $D \in L^\infty(\mathbb{R}^d;\mathbb{R}^d), \operatorname{div} D \in L^2_{\text{loc}}(\mathbb{R}^d), (\operatorname{div} D)^- \in L^\infty(\mathbb{R}^d).$

(B4) $\forall K \subset \mathbb{R}$ compact: $\exists\, \alpha_K > 0$ with $|B(r)r - B(s)s| \leqslant \alpha_K|\beta(r) - \beta(s)|$ $\forall r,s \in K$.

For the class of distributional solutions (in PDE-sense), this equation can be equivalently considered as a NLFPE, see Example (ii) in Section 3.1. The following holds: For each $(s,\zeta) \in \mathbb{R}_+ \times \mathcal{P}_0$, $\mathcal{P}_0 := \mathcal{P} \cap L^\infty$, there is a distributional solution $u^{s,\zeta}$ to (5.3.1) such that $\mu_t^{s,\zeta} = u_t^{s,\zeta}(x)dx$ is a weakly continuous probability solution in $\bigcap_{T>s} L^\infty((s,T) \times \mathbb{R}^d)$, and these solutions have the flow property in $\mathcal{P}_0$. Moreover, $\mu^{s,\zeta}$ is the unique weakly continuous probability solution from (s,ζ) to ($\mu^{s,\zeta}$-ℓFPE) in $\bigcap_{T>s} L^\infty((s,T) \times \mathbb{R}^d) \supseteq \mathcal{A}_{s,\leqslant}(\mu^{s,\zeta})$. For these statements, see [4, Thm.2.2] and [7, Cor.4.2], respectively.

Thus, Theorem 5.2.1 applies and gives a nonlinear Markov process $\{\mathbb{P}_{s,\zeta}\}_{s\geqslant 0,\zeta\in\mathcal{P}_0}$ with one-dimensional time marginals densities $u_t^{s,\zeta}$, where $\mathbb{P}_{s,\zeta}$ is the path law of a restricted-unique weak solution to the associated DDSDE

$$dX_t = B\big(u_t(X_t)\big)D(X_t))dt + \sqrt{\frac{2\beta(u_t(X_t))}{u_t(X_t)}}dB_t, \;\; \mathcal{L}_{X_t}(dx) = u_t\,dx, \;\; t \geqslant s.$$

Bottomline: The nonlinear PDE-solutions $u^{s,\zeta}$ have a probabilistic representation as the one-dimensional time marginal densities of a nonlinear Markov process, which consists of solutions to the associated DDSDE.

(iii) **Classical PME, measure-valued initial data.** For the classical porous media equation

$$\partial_t u = \Delta\big(|u|^{m-1}u\big), \quad m \geqslant 1,$$

it was shown in [16] that for *any* initial datum $(s,\zeta) \in \mathbb{R}_+ \times \mathcal{P}$, there is a unique weakly continuous distributional (in PDE-sense) probability solution $u^{s,\zeta}$ in $\bigcap_{T>\tau>s} L^\infty((\tau,T)\times\mathbb{R}^d)$. In fact, it is shown that $u^{s,\zeta}$ is even L^1-continuous on (s,∞). Clearly, this uniqueness implies the flow property for the solutions $t \mapsto u_t^{s,\zeta}(x)dx$ to the corresponding NLFPE, see Example (i) in Section 3.1. For $\zeta = \delta_{x_0}$, $u^{s,\zeta}$ is the explicit *Barenblatt solution*,

$$u_t^{s,\delta_{x_0}}(x) = (t-s)^{-\alpha}\left[\left(C - k|x-x_0|^2(t-s)^{-2\beta}\right)_+\right]^{\frac{1}{m-1}}, \quad t > s,$$

where $\alpha = \frac{d}{d(m-1)+2}$, $\beta = \frac{\alpha}{d}$, $k = \frac{\alpha(m-1)}{2md}$, $f_+ := \max(f,0)$, and $C = C(m,d) > 0$ is chosen such that $\int_{\mathbb{R}^d} u_t^{s,\zeta}(x)dx = 1$ for all $t > s$. The corresponding McKean-Vlasov equation is

$$dX_t = \sqrt{2u_t(X_t)^{m-1}}dB_t, \quad \mathcal{L}_{X_t} = u_t(x)dx, \quad t \geqslant s, \quad \mathcal{L}_{X_s} = \zeta. \quad (5.3.2)$$

Since assumptions (B1)-(B4) are satisfied, for $\zeta \in \mathcal{P} \cap L^\infty$ we have uniqueness of ($\mu^{s,\zeta}$-ℓFPE) from (s,ζ) in $\bigcap_{T>s} L^\infty((s,T)\times\mathbb{R}^d)$ (compare (ii) above). Thus, Corollary 5.2.6 applies with $\mathfrak{P}_0 = \mathcal{P} \cap L^\infty$ and $\mathcal{P}_0 = \mathcal{P}$, and yields a nonlinear Markov process $(\mathbb{P}_{s,\zeta})_{(s,\zeta)\in\mathbb{R}_+\times\mathcal{P}}$ consisting of path laws $\mathbb{P}_{s,\zeta}$ of weak solutions to (5.3.2) with one-dimensional time marginals densities $u_t^{s,\zeta}$.

Remark 5.3.1. *Corollary 5.2.6 first only implies that $\mathbb{P}_{s,\zeta}$ is uniquely determined if $\zeta \in \mathcal{P} \cap L^\infty$. However, it can be shown from the formula for the finite-dimensional marginals in Proposition 5.1.4 that the entire nonlinear Markov process $\{\mathbb{P}_{s,\zeta}\}_{s\geqslant 0,\zeta\in\mathcal{P}}$ is uniquely determined by $\{u^{s,\zeta}\}_{s\geqslant 0,\zeta\in\mathcal{P}}$.*

For further applications, including the $1D$-Burgers equation and the $2D$ Navier–Stokes equations in vorticity form, please see [18].

5.3.1 p-Laplace equation

The main reference for the subsequent content is [3]. As a particularly interesting example to which the above theory applies, consider the *p-Laplace equation*

$$\frac{\partial}{\partial t} u(t,x) = \operatorname{div}(|\nabla u(t,x)|^{p-2}\nabla u(t,x)), \ (t,x) \in (0,\infty)\times\mathbb{R}^d, \quad (5.3.3)$$

where $p > 2$ (the equation is meaningful for $p > 1$, but here we only consider $p > 2$). It may be considered a nonlinear generalization of the heat equation, which is recovered for $p = 2$. There are explicitly given solutions, called *Barenblatt solutions*, given by

$$w^y(t,x) = t^{-k}\big(C_1 - qt^{-\frac{kp}{d(p-1)}}|x-y|^{\frac{p}{p-1}}\big)_+^{\frac{p-1}{p-2}}, \ (t,x) \in (0,\infty)\times\mathbb{R}^d, \quad (5.3.4)$$

where $k := \left(p-2+\frac{p}{d}\right)^{-1}$, $q := \frac{p-2}{p}\left(\frac{k}{d}\right)^{\frac{1}{p-1}}$, $C_1 \in (0,\infty)$ is the unique constant such that $|w^y(t)|_{L^1(\mathbb{R}^d)} = 1$ for all $t > 0$. Here "solution" is understood in the usual weak sense, i.e. both sides of the equation coincide as Sobolev functions. One can easily show $w^y(t,x)dx \xrightarrow{t\to 0} \delta_y$ weakly.

Fokker–Planck reformulation. At first sight, the p-Laplace equation is not of Fokker–Planck type. However, it was shown in [3] the following result. Its proof is obtained by observing that for $u = w^y$ the RHS of (5.3.5) equals the RHS of (5.3.3) in distributional sense.

Proposition 5.3.2. *The Barenblatt solution $t \mapsto w^y(t,x)dx$ is a weakly continuous probability solution with initial datum δ_y to the nonlinear Fokker–Planck equation*

$$\partial_t u = \Delta\big(|\nabla u|^{p-2}u\big) - div\big(\nabla\big(|\nabla u(t,x)|^{p-2}\big)u\big). \tag{5.3.5}$$

The coefficients $a_{ij}(u,x) = \delta_{ij}|\nabla u(x)|^{p-2}$ and $b_i(u,x) = \partial_i(|\nabla u|^{p-2})$ are defined on $W^{1,1}_{\text{loc}}(\mathbb{R}^d) \times \mathbb{R}^d$. Note that this is a new type of nonlinear FP-equation compared to what we considered before, since now the dependence of the solution is not pointwise via its density, but pointwise via the gradient of its density (which means that, as a function of u, the coefficients depend not pointwise on u, but on its values in a neighborhood of $u(x)$).

The corresponding McKean–Vlasov SDE is

$$\begin{aligned} dX(t) &= \nabla(|\nabla u(t,X(t))|^{p-2})dt + \sqrt{2}\,|\nabla u(t,X(t))|^{\frac{p-2}{2}}dW(t),\ t>0, \\ \mathcal{L}_{X(t)} &= u(t,x)dx,\ t>0, \end{aligned} \tag{5.3.6}$$

and we have the following superposition result. Note that while strictly speaking the nonlinear superposition principle was not formulated for gradient-dependent coefficients, one can simply use the linear result by first freezing w^y (that is, ∇w^y) in the coefficients.

Proposition 5.3.3. *There exists a probabilistically weak solution $(X^y)_{t\geqslant 0}$ to the above McKean–Vlasov SDE such that $\mathcal{L}_{X^y_t}(dx) = w^y(t,x)dx$ for all $t > 0$ and $\mathcal{L}_{X^y_0}(dx) = \delta_y(dx)$.*

This result also holds for general solutions to (5.3.3) which satisfy certain Sobolev regularity assumptions. Here, we focus solely on the Barenblatt solutions.

p-Brownian motion. Set

$$\mathcal{P}_0 := \{w^y(\delta,x)dx,\ y\in\mathbb{R}^d,\ \delta \geqslant 0\} \subseteq \mathcal{P}$$

(note in particular $\{\delta_y, y\in\mathbb{R}^d\} \subseteq \mathcal{P}_0$, where we abuse notation in writing $w^y(0,x)dx = \delta_y$), i.e. "$\mathcal{P}_0$= all possible distributions attained by Barenblatt solutions". Note that for $\zeta \in \mathcal{P}_0$, there is a *unique* pair (y,δ) such that $\zeta = w^y(\delta,x)dx$. Clearly, for $\zeta = w^y(\delta,x)dx$,

$$(\mu^{s,\zeta}_t)_{t\geqslant s} := (w^y(\delta+t-s,x)dx)_{t\geqslant s}$$

is a weakly continuous probability solution to (5.3.5) with initial datum (s,ζ). That $\{\mu^{s,\zeta}\}_{s\geqslant 0,\zeta\in\mathcal{P}_0}$ is a solution flow in $\mathcal{P}_0$ is obvious. We have the following result.

Theorem 5.3.4. *Let $d \geqslant 2$, $p > 2$. There is a nonlinear Markov process $\{P_{s,\zeta}\}_{s\geqslant 0,\zeta\in\mathcal{P}_0}$ such that $P_{s,\zeta}$ has marginals $\mu_t^{s,\zeta}, t \geqslant s$, and is the path law of a weak solution $X^{s,\zeta}$ to (5.3.6).*

Moreover, each $X^{s,\zeta}$ is the unique weak solution to (5.3.6) with marginals $\mu_t^{s,\zeta}, t \geqslant s$. In particular, this nonlinear Markov process is uniquely determined by the equation (5.3.6) and the Barenblatt solutions.

Furthermore, since in the present case the coefficients are not explicitly time-dependent, we find:

Lemma 5.3.5. *$(P_{s,\zeta})_{s\geqslant 0,\zeta\in\mathcal{P}_0}$ is time-homogeneous, i.e. $P_{s,\zeta} = P_{0,\zeta} \circ (\hat{\Pi}_s)^{-1}$ for all $(s,\zeta) \in [0,\infty) \times \mathcal{P}_0$, where*

$$\hat{\Pi}_s : C([0,\infty),\mathbb{R}^d) \to C([s,\infty),\mathbb{R}^d), \quad \hat{\Pi}_s : (\omega(t))_{t\geqslant 0} \mapsto (\omega(t-s))_{t\geqslant s}. \quad (5.3.7)$$

Moreover, for $\zeta = w^y(\delta,x)dx$, we have $P_{0,\zeta} = P_{0,\delta_y} \circ (\tilde{\Pi}^0_\delta)^{-1}$, where we define $\tilde{\Pi}^s_r : C([s,\infty),\mathbb{R}^d) \to C([s,\infty),\mathbb{R}^d)$ via

$$\tilde{\Pi}^s_r : \omega(t)_{t\geqslant s} \mapsto \omega(t+r)_{t\geqslant s}. \quad (5.3.8)$$

As a consequence of this lemma, the nonlinear Markov process $\{P_{s,\zeta}\}_{s\geqslant 0,\zeta\in\mathcal{P}_0}$ is uniquely determined by $\{P_y\}_{y\in\mathbb{R}^d}$, where we set $P_y := P_{0,\delta_y}$.

Definition 5.3.6. Let $d \geqslant 2$, $p > 2$. We call $\{P_y\}_{y\in\mathbb{R}^d}$ *p-Brownian motion*, P_0 *p-Wiener measure*, and any stochastic process X^y with law P_y a *p-Brownian motion with start in y*.

This definition is consistent with the classical case $p = 2$, where the Barenblatt solutions are replaced by the heat kernel and the corresponding (classical) Markov process is Brownian motion ("2-Brownian motion").

Remark 5.3.7. *We point out that for $p > 2$, unlike in the case $p = 2$, the measures P_y are not given as the image measure of P_0 under the translation map $T_y : C([0,\infty),\mathbb{R}^d) \to C([0,\infty),\mathbb{R}^d), T_y(\omega) := \omega + y$. This can be seen, for instance, from the fact that $P_0 \circ (T_y)^{-1}$ is* not *a solution path law of the McKean–Vlasov SDE (5.3.6)–(5.3.7) with initial condition δ_y.*

A crucial restricted uniqueness result for a linear PDE. The proof of Theorem 5.3.4 follows by applying Corollary 5.2.6, for which we aim to show that for every $(s,\zeta) \in [0,\infty) \times \mathfrak{P}_0$, $\zeta = w^y(\delta,x)dx$, where $\mathfrak{P}_0 := \{w^y(\delta,x)dx, \delta > 0\}$, the $\mu^{s,\zeta}$-linearized FPE

$$\partial_t u(t,x) = \Delta\big(|\nabla w^y(\delta+t-s,x)|^{p-2}u(t,x)\big) - \operatorname{div}\big(\nabla\big(|\nabla w^y(\delta+t-s,x)|^{p-2}\big)u(t,x)\big)$$

with initial condition $u(t,x)dx \xrightarrow{t\to s} \zeta$ has a unique solution $u : (s,\infty) \times \mathbb{R}^d \to \mathbb{R}$ under the restriction that there is some $C > 0$ such that $0 \leqslant u(t,x) \leqslant Cw^y(\delta + t - s, x)$ dx-a.s. for every $t \geqslant s$. Here, solution is understood in the sense that $t \mapsto u(t,x)dx$ is a weakly continuous probability solution to the above equation in the sense of Fokker–Planck equations.

Indeed, this is proven in [3]. Note that this is a restricted uniqueness result for a linear PDE with degenerate diffusion- and drift-coefficient (w^y and hence ∇w^y is compactly supported on each $[0,T] \times \mathbb{R}^d$).

Bibliography

[1] L. Ambrosio. *Transport equation and Cauchy problem for non-smooth vector fields*, pages 1–41. Springer Berlin Heidelberg, 2008.

[2] V. Barbu. *Nonlinear differential equations of monotone types in Banach spaces.* Springer Monographs in Mathematics. Springer, New York, 2010.

[3] V. Barbu, M. Rehmeier, and M. Röckner. p–Brownian motion and the p–Laplacian. *arXiv preprint 2409.18744*, 2024.

[4] V. Barbu and M. Röckner. Solutions for nonlinear Fokker–Planck equations with measures as initial data and McKean-Vlasov equations. *Journal of Functional Analysis*, 280(7):108926, 2021.

[5] V. Barbu and M. Röckner. Probabilistic representation for solutions to nonlinear Fokker–Planck equations. *SIAM Journal on Mathematical Analysis*, 50(4):4246–4260, 2018.

[6] V. Barbu and M. Röckner. From nonlinear Fokker–Planck equations to solutions of distribution dependent SDE. *The Annals of Probability*, 48(4):1902–1920, 2020.

[7] V. Barbu and M. Röckner. Uniqueness for nonlinear Fokker–Planck equations and for McKean-Vlasov SDEs: The degenerate case. *arXiv preprint 2203.00122*, 2022.

[8] V. I. Bogachev, T. I. Krasovitskii, and S. V. Shaposhnikov. On nonuniqueness of probability solutions to the cauchy problem for the fokker–planck–kolmogorov equation. *Doklady Mathematics*, 103(3):108–112, 2021.

[9] V. I. Bogachev, N. V. Krylov, M. Röckner, and S. V. Shaposhnikov. *Fokker–Planck–Kolmogorov Equations.* Mathematical Surveys and Monographs 207. American Mathematical Society, 2015.

[10] V. I. Bogachev, M. Röckner, and S. V. Shaposhnikov. On the Ambrosio–Figalli–Trevisan superposition principle for probability solutions to Fokker–Planck–Kolmogorov equations. *Journal of Dynamics and Differential Equations*, 33(2):715–739, 2021.

[11] A. Figalli. Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. *Journal of Functional Analysis*, 254(1):109–153, 2008.

[12] T. D. Frank. *Nonlinear Fokker–Planck Equations: Fundamentals and Applications.* Springer Berlin Heidelberg, 2005.

[13] S. Grube. *Strong solutions to McKean–Vlasov stochastic differential equations with coefficients of Nemytskii-type*. PhD thesis. Bielefeld University, 2022.

[14] O. A. Manita and S. V. Shaposhnikov. Nonlinear parabolic equations for measures. *Algebra i Analiz*, 25(1):64–93, 2013.

[15] J. R. Munkres. *Topology*. Prentice Hall, Incorporated, 2000.

[16] M. Pierre. Uniqueness of the solutions of $u_t - \Delta\phi(u) = 0$ with initial datum a measure. *Nonlinear Analysis: Theory, Methods & Applications*, 6(2):175–187, 1982.

[17] M. Rehmeier. Flow selections for (nonlinear) Fokker–Planck–Kolmogorov equations. *J. Differential Equations*, 328:105–132, 2022.

[18] M. Rehmeier and M. Röckner. On nonlinear Markov processes in the sense of McKean. *arXiv preprint 2212.12424*, 2022.

[19] D. W. Stroock. *Lectures on Stochastic Analysis: Diffusion Theory*. London Mathematical Society Student Texts. Cambridge University Press, 1987.

[20] D. W. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes*. Classics in Mathematics. Springer Berlin Heidelberg, 2007.

[21] D. Trevisan. Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients. *Electronic Journal of Probability*, 21(none):1 – 41, 2016.

[22] G. Da Prato V. I. Bogachev and M. Röckner. On parabolic equations for measures. *Communications in Partial Differential Equations*, 33(3):397–418, 2008.

[23] F.-Y. Wang. Distribution dependent SDEs for Landau type equations. *Stochastic Processes and their Applications*, 128(2):595–621, 2018.