MASTERARBEIT | MASTER'S THESIS

Titel | Title

McKean Stochastic Differential Equations and Selected Applications in Mathematical Finance

verfasst von | submitted by

Johannes Rotheneder BSc BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 821
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Mathematik
Betreut von | Supervisor:	Univ.-Prof. Dipl.-Math. Dr. rer.nat. Uwe Schmock

Zusammenfassung

Stochastische Differentialgleichungen (SDGLn) McKean'scher Art besitzen Koeffizienten, die neben der Zeit und dem Pfad der Lösung auch von der Wahrscheinlichkeitsverteilung der Lösung abhängen. Sie treten bei der wahrscheinlichkeitstheoretischen Betrachtung von Systemen auf, die aus vielen wechselwirkenden Bestandteilen, auch Partikel genannt, bestehen. In diesem Kontext kann die Wechselwirkung eines einzelnen Partikels mit dem Rest des Systems durch eine Wechselwirkung mit der Verteilung eines einzelnen repräsentativen Partikels modelliert werden – der Verteilung der Lösung der McKean-SDGL.

Diese Arbeit stellt die Theorie der McKean-SDGLn und ausgewählte Anwendungen in der Finanzmathematik vor. Wir betrachten McKean-SDGLn auf einem unendlichen Zeithorizont, deren Koeffizienten und zugrundeliegenden Brown'sche Bewegungen Werte in reellen oder komplexen euklidischen Räumen annehmen. Die Koeffizienten können von der Zeit, dem zurückliegenden Pfad der Lösung, dem Wahrscheinlichkeitsmaß des zurückliegenden Pfades der Lösung sowie von zusätzlichem, einer Adaptiertheitsbedingung unterliegendem Zufall abhängen. Unter der Annahme von Lipschitz-Stetigkeit und beschränktem Wachstum der Koeffizienten beweisen wir die starke Existenz und pfadweise Eindeutigkeit von Lösungen. Unter denselben Voraussetzungen, jedoch unter Ausschluss der Abhängigkeit der Koeffizienten von zusätzlichem Zufall, beweisen wir außerdem die schwache Eindeutigkeit und beschreiben die Partikelapproximation, welche eine Methode zur numerischen Simulation von Lösungen von McKean-SDGLn bereitstellt. Anschließend beweisen wir deren Konvergenz und leiten eine explizite Konvergenzrate unter der Annahme her, dass die Koeffizienten vom Maßargument nur über die eindimensionalen Randverteilungen der Lösung abhängen.

Der zweite Teil der Arbeit widmet sich ausgewählten Anwendungen von McKean-SDGLn in der Finanzmathematik. Eine wichtige Anwendung bei der Bewertung exotischer Optionen ist die Kalibrierung lokaler stochastischer Volatilitätsmodelle an Markt-Smiles impliziter Volatilitäten. Dieses Problem lässt sich mithilfe einer Partikelmethode lösen, welche bei mehreren stochastischen Faktoren im Modell vorteilhaft ist gegenüber Verfahren, die auf partiellen Differentialgleichungen basieren. Im Kontext der Zinsmodellierung diskutieren wir eine Mean-Field-Version des LIBOR-Marktmodells (LMM), welche dem Problem der Explosion von Terminzinssätzen im klassischen LMM durch eine Abhängigkeit der SDGL-Koeffizienten von der Varianz der Terminzinssatzprozesse begegnet.

Schlagworte: Stochastische Differentialgleichungen McKean'scher Art, Partikelmethode, Ausbreitung des Chaos, Lokale stochastische Volatilitätsmodelle, Mean-Field LIBOR Marktmodelle

Abstract

Stochastic differential equations (SDEs) of McKean type have coefficients which depend, in addition to time and the trajectory of the solution, also on the probability distribution of the solution. They arise in the probabilistic study of systems consisting of many interacting agents, also referred to as particles. In this setting, the interaction of a single particle with the rest of the system can sometimes be modelled by an interaction with the distribution of a single representative particle—the distribution of the solution of the McKean SDE.

This thesis presents the theory of McKean SDEs and selected applications in mathematical finance. We consider McKean SDEs on an infinite time horizon, with coefficients and driving Brownian motions taking values in real or complex Euclidean spaces. The coefficients may depend on time, the past path of the solution, the probability distribution of the past path of the solution, and additional randomness subject to an adaptedness condition. Under Lipschitz continuity and bounded growth conditions, we prove strong existence and pathwise uniqueness of solutions. In the same setting, but excluding dependence of coefficients on additional randomness, we furthermore prove weak uniqueness and describe the particle approximation, which provides a method for the numerical simulation of solutions of McKean SDEs. Subsequently, we prove its convergence and establish an explicit rate of convergence in the case where the coefficients depend on the measure argument only through the marginal distributions.

The second part of the thesis is devoted to applications of McKean SDEs to mathematical finance. An important application to the pricing of exotic options is the calibration of the local stochastic volatility model to market smiles of implied volatilities by means of a particle method, which has advantages over PDE-based methods in the presence of many stochastic factors driving the model. In the context of interest rate modelling, we discuss a mean-field version of the LIBOR market model (LMM), which addresses the problem of explosion of forward interest rates in the classical LMM through a dependence of SDE coefficients on the variance of the forward interest rate process.

Several appendices provide the required foundation. This includes pseudometric spaces, measure and probability theory, a comprehensive discussion of classical SDEs, and selected results from the theory of Wasserstein spaces.

Keywords: McKean stochastic differential equations, propagation of chaos, particle method, local stochastic volatility, mean-field LIBOR market model

Mathematics Subject Classification (MSC2020): 60H10 (primary), 60H05, 60H30, 60K35, 91G20 (secondary)

Acknowledgements

First and foremost, I want to express my deepest gratitude to my advisor, Uwe Schmock, for his guidance and enthusiasm throughout writing this thesis. His example, standards and mentorship have been truly formative for my development as a mathematician.

I could not have completed this journey without the support of my family, whose encouragement throughout the years has been invaluable. I am also most grateful to my friends who have shared the highs and lows of academic life and provided much-needed perspective and levity along the way.

Finally, and most importantly, I thank my partner, Gloria. Your wisdom, humour and unwavering love were indispensable, and I could not have done this without you.

Johannes Rotheneder

Contents

1. Introduction

Stochastic differential equations (SDEs), which generalize first-order ordinary differential equations, govern the time evolution of stochastic processes. They are central tools in mathematical finance, as their solutions can be used to model financial quantities such as asset prices, interest rates, and volatilities. Classical SDEs describe systems whose future dynamics depend on their past trajectories and a source of randomness represented by Brownian motion. The coefficients of such SDEs are functions of time and the governed stochastic process itself.

In the probabilistic study of systems consisting of many interacting agents (also referred to as particles), the interaction of a single particle with the rest of the system can sometimes be modelled by an interaction with the empirical distribution of all particles. In the limit of infinitely many particles, the interaction with the empirical distribution of the system is, under some assumptions, equivalent to an interaction with the distribution of a single representative particle. In this way arise SDEs with coefficients depending additionally on the probability distribution of the solution, raising the level of complexity. SDEs of this type, introduced by H.P. McKean in [40], are known as McKean SDEs, and have found applications in a number of fields including physics, biology, game theory, and mathematical finance.

A McKean SDE Arising in Interest Rate Modelling

An example of a McKean SDE arising in interest rate modelling, see [17], is

$$\mathrm{d}X_t = \exp\bigl(-4(\mathbb{V}[X_t] - \tfrac{3}{2})^+\bigr) X_t \,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+, \tag{1.1}$$

where $B = (B_t)_{t\in\mathbb{R}_+}$ is a real-valued standard Brownian motion and $\mathbb{V}[X_t]$ denotes the variance of the solution at time t. Without the exponential factor, (1.1) is the SDE describing geometric Brownian motion with zero drift and volatility equal to one. The added dependence on the variance has the effect that, once the variance of the solution exceeds the threshold value $\frac{3}{2}$, the volatility driving further fluctuations is reduced, and the reduction becomes exponentially stronger as the variance increases. The example (1.1) raises several interesting questions, e.g. regarding existence and uniqueness of solutions and approaches for the simulation of such SDEs in light of the fact that the law of the solution is unknown a priori. As outlined below, these questions will be addressed with some generality. In particular, we will present and prove in Theorem 2.56 the convergence of the particle approximation, which allows for the approximation of independent solutions of McKean SDEs by solutions of a related, suitably coupled system of classical SDEs. Figure 1.1 shows approximate solutions of (1.1) computed using this particle method. The analytical solution of (1.1), see Example 3.26, exceeds the threshold level of variance, $\frac{3}{2}$, shortly before time 1. From then

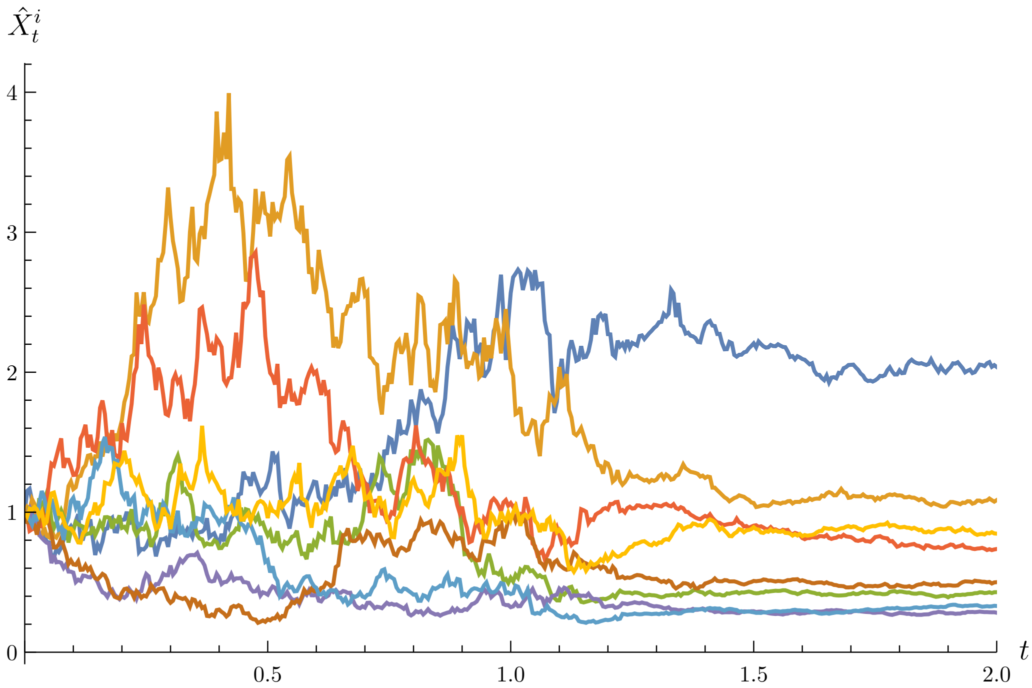

Figure 1.1. Eight sample paths of approximate solutions of the McKean SDE (1.1), computed using the particle approximation described in Subsection 2.4.2 with $N = 250$ particles.

on, the volatility reduction induced by the variance-dependent exponential factor in (1.1) can be readily observed.

Structure of this Thesis

This thesis[1] is devoted to the study of McKean SDEs and selected applications in mathematical finance. It is structured into three parts. In Chapter 2, we develop the theory of McKean SDEs. In Chapter 3, we present selected applications. Finally, an extensive appendix provides the required mathematical foundation.

Theory of McKean SDEs and the Particle Method

Chapter 2 focuses on the theoretical aspects of McKean SDEs, most importantly strong existence, weak and pathwise uniqueness of solutions as well as the particle method and its convergence. We consider simultaneously:

(a) The infinite time horizon.

[1] The text in this thesis is extensively hyperlinked, which we hope improves the clarity of the exposition and the proofs. With a suitable PDF reader (e.g. TEXShop and `Skim` on MacOS, Evince on Linux or the excellent Sioyek on MacOS, Linux and Windows), a snapshot of a link's destination can be previewed in the main text without actually following the link, which reduces context switches and is especially useful in proofs, say to retrace a chain of substitutions.

(b) Coefficients and driving Brownian motions with values in real or complex Euclidean spaces.

(c) Dependence of coefficients on time, the past trajectory of the solution, the probability distribution of the past trajectory of the solution, and additional randomness subject to an adaptedness condition.

(d) The $\mathcal{L}^p$-case for $p \in [2, \infty)$ rather than just the $\mathcal{L}^2$-case. That is, we admit p-integrable initial conditions as well as SDE coefficients depending on distributions with finite p-th moments satisfying a Lipschitz continuity condition involving the p-th order Wasserstein metric (Definition 2.19).

Section 2.1 introduces the setting and basic definitions surrounding McKean SDEs. In Section 2.2, we establish the main strong existence and pathwise uniqueness result, Theorem 2.32, under (global in the state and local in the time variable) Lipschitz continuity and bounded growth conditions on the coefficients (Definitions 2.19 and 2.20). In Section 2.3, we present two important classes of examples of McKean SDEs, McKean–Vlasov SDEs (Definition 2.37) and McKean SDEs with interaction of scalar type (Definition 2.40). Finally, in Section 2.4, we discuss chaotic sequences of probability measures, introduce the particle method, which allows for the approximation of a system of independent solutions of a McKean SDE by a system of coupled classical SDEs, and prove its convergence. For coefficients which depend on the measure argument only through the one-dimensional marginal distributions and which have a Lipschitz property adapted to this setting, we furthermore obtain an explicit rate of convergence of the particle approximation (Theorem 2.58).

Applications of McKean SDEs to Mathematical Finance

Chapter 3 is devoted to applications of McKean SDEs to option pricing and interest rate modelling. In Section 3.1, we consider the calibration of local stochastic volatility models (LSVMs) to market smiles of implied volatilities by means of a particle method. LSVMs, see Definition 3.1, extend the classical Black–Scholes model, which governs the time evolution of the price of a risky asset, by allowing the volatility to depend on time, the risky asset's price and an additional stochastic process, the so-called stochastic volatility. With this extension, important empirically observed features of financial markets such as the smile or skew of implied volatilities can be reproduced, allowing for more accurate pricing of exotic derivatives. However, the added complexity complicates the calibration of such models to market-observed option prices as well as the pricing of derivatives, as PDE methods suffer from the curse of dimensionality when the number of stochastic factors increases. The particle method presented in Subsection 3.1.4 is an efficient alternative method for the calibration and option pricing using the LSVM, which easily incorporates additional stochastic factors such as stochastic interest rates.

In Section 3.2, we turn to interest rate modelling and discuss a mean-field version of the LIBOR market model (LMM) for forward interest rates. The classical LMM, presented in Subsection 3.2.1, is used e.g. in the insurance sector for the valuation of long-term guarantees under the Solvency II regulatory framework. Unfortunately, this model suffers from the problem of explosion, which means that in a significant fraction of scenarios (say, more than 1%), the modelled forward interest rates exceed realistic thresholds. Subsection 3.2.2 considers McKean SDEs of a specific form (see (3.55)), where the dependence on the measure argument is of a local rather than global Lipschitz type. Such a dependence appears with variance-dependent coefficients as in (1.1), and also in the mean-field LMM. In Theorem 3.19, we obtain strong existence and pathwise uniqueness also in this setting and subsequently discuss the concrete examples of arithmetic and geometric Brownian motion with variance-dependent volatility dampening (Examples 3.24 and 3.26). In Subsection 3.2.3, we then return to LMMs and describe how the problem of explosion in the classical LMM is addressed in the mean-field LMM by allowing for a dependence of the volatilities of the forward rates on their laws (in particular their variances), thereby reducing the probability of explosion.

Appendices

Several appendices provide the required foundation. Appendix A discusses pseudometric spaces and develops a central fixed point result, Theorem A.16. Appendix B collects results from measure and probability theory used throughout this thesis, including an extension result for probability measures on $C(\mathbb{R}_+, \mathbb{K}^n)$, Proposition B.20. Appendix C contains a comprehensive discussion of classical SDEs with coefficients depending on time, the past trajectory of the solution and additional randomness, subject to an adaptedness condition. We cover basic terminology, strong existence and pathwise uniqueness (Theorem C.34), the existence of flows (Theorem C.43) and weak uniqueness (Corollary C.57). Appendix D reviews background from the theory of Wasserstein spaces, which provide important topological and measurable structures underpinning McKean SDEs. Appendix E contains a summary of the major and minor results presented in this thesis, grouped into literature-based results and results developed independently of or extending the literature. Finally, Appendix F lists open questions and topics for further study.

2. McKean Stochastic Differential Equations

In this chapter, we introduce the setting and basic definitions surrounding McKean SDEs (Section 2.1) and then establish strong existence and pathwise uniqueness of solutions under (global in the state and local in the time variable) Lipschitz continuity and bounded growth conditions (Section 2.2, in particular Theorem 2.32). In Section 2.3, we present some prominent examples of McKean SDEs, those being McKean–Vlasov SDEs (Definition 2.37) and McKean SDEs with interaction of scalar type (Definition 2.40). Finally, in Section 2.4, we discuss chaotic sequences of probability measures, introduce the particle method and then prove its convergence.

2.1. Introduction and Definitions

Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. For $n \in \mathbb{N}$, let W_n denote the vector space $C(\mathbb{R}_+, \mathbb{K}^n)$ of $\mathbb{K}^n$-valued continuous functions on $\mathbb{R}_+$. For every $t \in \mathbb{R}_+$, define

$$p_t(w) = \max_{s\in[0,t]} |w(s)|, \qquad w \in W_n, \tag{2.1}$$

the seminorm of uniform convergence on $[0, t]$. Let $(c_k)_{k\in\mathbb{N}}$ be a null sequence in $\mathbb{R}^*_+$. Then

$$\varrho_{\mathrm{ucc}}(w, \tilde{w}) = \sup_{k\in\mathbb{N}} \left(c_k \wedge p_k(w - \tilde{w})\right), \qquad w, \tilde{w} \in W_n, \tag{2.2}$$

defines a bounded metric inducing the topology $\mathcal{O}_{\mathrm{ucc}}$ of compact convergence on W_n. Let

$$\pi_s \colon W_n \to \mathbb{K}^n, \quad w \mapsto w(s), \qquad s \in \mathbb{R}_+, \tag{2.3}$$

be the canonical projections and equip W_n with the σ-algebra

$$\mathcal{G} = \sigma(\pi_s : s \in \mathbb{R}_+), \tag{2.4}$$

and the filtration $\mathbb{G} = (\mathcal{G}_t)_{t\in\mathbb{R}_+}$ defined by[1]

$$\mathcal{G}_t = \sigma\big(\pi_s : s \in [0, t]\big), \qquad t \in \mathbb{R}_+. \tag{2.5}$$

Let

$$W_n^t := C([0, t], \mathbb{K}^n), \qquad t \in \mathbb{R}_+, \tag{2.6}$$

[1] No completeness assumption is placed on the σ-algebra $\mathcal{G}$ and the filtration $\mathbb{G}$.

equipped with the topology $\mathcal{O}_{W_n^t}$ of uniform convergence on $[0,t]$ and the associated Borel σ-algebra $\mathcal{B}_{W_n^t}$. The topology $\mathcal{O}_{W_n^t}$ is induced by the norm

$$p_t'(w) := \max_{s\in[0,t]} |w(s)|, \qquad w \in W_n^t, \tag{2.7}$$

which by analogy with (2.1) we will usually also denote by p_t. We also define for each $t \in \mathbb{R}_+$ the projection

$$\pi_{[0,t]} \colon W_n \to W_n^t, \quad w \mapsto w\restriction_{[0,t]}, \tag{2.8}$$

and, for all $s \le t$ in $\mathbb{R}_+$,

$$\pi_{[0,s]}^{[0,t]} \colon W_n^t \to W_n^s, \quad w \mapsto w\restriction_{[0,s]}. \tag{2.9}$$

The following lemma collects several well-known results[2] about the uniform topologies on the spaces (W_n^t, p_t'), $t \in \mathbb{R}_+$, and the topology of compact convergence on $(W_n, \varrho_{\mathrm{ucc}})$, as well as the connections of these topologies to related canonical σ-algebras.

Lemma 2.1 (Topological and measurability properties of the spaces W_n and W_n^t)**.**

(a) For each $t \in \mathbb{R}_+$, $(W_n^t, \mathcal{O}_{W_n^t})$ is a Polish space.

(b) The topology $\mathcal{O}_{\mathrm{ucc}}$ of compact convergence in W_n is Polish.

(c) Let $\mathcal{O}_{p_t}$ denote the topology[3] on W_n induced by the seminorm p_t (see Example A.2(b) and Definition A.3). Then

$$\mathcal{G}_t = \sigma(\mathcal{O}_{p_t}) = \pi_{[0,t]}^{-1}(\mathcal{B}_{W_n^t}), \qquad t \in \mathbb{R}_+. \tag{2.10}$$

(d) The σ-algebra $\mathcal{G}$ from (2.4) coincides with the Borel σ-algebra $\sigma(\mathcal{O}_{\mathrm{ucc}})$ of the topology of compact convergence on W_n.

(e) For all $s \le t$ in $\mathbb{R}_+$, the map $\pi_{[0,s]}^{[0,t]}$ from (2.9) is $\mathcal{B}_{W_n^t}$-$\mathcal{B}_{W_n^s}$-measurable.

Proof. (a) For completeness of the norm p_t' inducing $\mathcal{O}_{W_n^t}$, see [42, p. 268]. Separability follows from componentwise application[4] of the Stone–Weierstrass theorem ([36, Theorem 15.2]), according to which for every $t \in \mathbb{R}_+$, the countable set of polynomials with rational coefficients is dense in $C([0,t],\mathbb{R})$ w.r.t. the supremum norm.

(b) See [36, Theorem 21.30] and [53, Exercise 2.121].

[2] See the proof below for detailed references. We will generally cite references directly in the statement of a result when that result is substantially identical to one appearing in the literature, and in this case, unless otherwise noted, the citation also covers the proof. In case a statement has been considerably modified, but the proof is drawn from or essentially identical to a reference, we cite the reference in the proof only (as for Lemma 2.1). When a result or proof is based on ideas from and at the same time materially extends one or several different sources, we briefly summarize the genealogy of ideas and cite relevant references before the statement of the result (as for Lemma 2.31 and Theorem 2.32).

[3] This topology is second countable by separability of $C([0,t],\mathbb{K}^n)$ with respect to the topology of uniform convergence induced by p_t. It is not Hausdorff, since $p_t(f-g)=0$ does not imply $f=g$ in W_n.

[4] In the case $\mathbb{K}=\mathbb{C}$, the Stone–Weierstrass theorem for $C([0,t],\mathbb{R})$ is applied separately to each real and imaginary component.

(c) Fix $t \in \mathbb{R}_+$. We begin with the second equality in (2.10). Let p'_t be the norm inducing $\mathcal{O}_{W_n^t}$ from (2.7). Then

$$p_t(w - \tilde{w}) = p'_t(\pi_{[0,t]}(w) - \pi_{[0,t]}(\tilde{w})), \qquad w, \tilde{w} \in W_n.$$

By Lemma A.6 (a), it follows that

$$\mathcal{O}_{p_t} = \pi_{[0,t]}^{-1}(\mathcal{O}_{W_n^t}), \tag{2.11}$$

and thus

$$\sigma(\mathcal{O}_{p_t}) = \sigma(\pi_{[0,t]}^{-1}(\mathcal{O}_{W_n^t})) = \pi_{[0,t]}^{-1}(\sigma(\mathcal{O}_{W_n^t})) = \pi_{[0,t]}^{-1}(\mathcal{B}_{W_n^t}).$$

For the first equality in (2.10), observe that for all $s \in [0, t]$, the projection π_s from (2.3) is continuous w.r.t. p_t and thus $\sigma(\mathcal{O}_{p_t})$-$\mathcal{B}_{\mathbb{K}^n}$-measurable. From (2.5), it follows that $\mathcal{G}_t \subseteq \sigma(\mathcal{O}_{p_t})$. For the reverse inclusion, note that separability of $\mathcal{O}_{W_n^t}$ (which holds by part (a)) implies by (2.11) separability of $\mathcal{O}_{p_t}$. Consequently, every open set in $\mathcal{O}_{p_t}$ is a countable union of open balls w.r.t. p_t. Each such ball is a preimage of an open subset of $\mathbb{R}_+$ under the map

$$W_n \ni w \mapsto p_t(w - w_0) \in \mathbb{R}_+, \tag{2.12}$$

for some $w_0 \in W_n$. Hence it suffices to show that every map of the form (2.12) is $\mathcal{G}_t$-measurable. This follows from (2.5), the representation

$$p_t(w - w_0) = \sup_{s \in [0,t] \cap \mathbb{Q}} |\pi_s(w) - \pi_s(w_0)|, \qquad w \in W_n,$$

and the fact that countable suprema of measurable functions are measurable.

(d) [36, Theorem 21.31] For every $t \in \mathbb{R}_+$, the projection π_t from (2.3) is continuous w.r.t. the topology $\mathcal{O}_{\mathrm{ucc}}$. From (2.4), it follows that $\mathcal{G} \subseteq \sigma(\mathcal{O}_{\mathrm{ucc}})$. For the reverse inclusion, note that every open ball in $\mathcal{O}_{\mathrm{ucc}}$ is a preimage of an open set in $\mathbb{R}_+$ under the map

$$W_n \ni w \mapsto \varrho_{\mathrm{ucc}}(w, w_0) \in \mathbb{R}_+, \tag{2.13}$$

for some $w_0 \in W_n$. We have seen in part 5 of the proof that $W_n \ni w \mapsto p_k(w - w_0) \in \mathbb{R}_+$ is $\mathcal{G}_k$-measurable for each $k \in \mathbb{N}$. Hence from the definition (2.2) of ϱ_{ucc} and the fact that $\mathcal{G} = \sigma(\bigcup_{t \in \mathbb{R}_+} \mathcal{G}_t)$, every map of the form (2.13) is $\mathcal{G}$-measurable. Consequently, every open ball w.r.t. ϱ_{ucc} is in $\mathcal{G}$. Using that $\mathcal{O}_{\mathrm{ucc}}$ is separable by part (b), every set in $\mathcal{O}_{\mathrm{ucc}}$ is a countable union of open balls w.r.t. ϱ_{ucc}. Therefore, $\mathcal{O}_{\mathrm{ucc}} \subseteq \mathcal{G}$ and thus $\sigma(\mathcal{O}_{\mathrm{ucc}}) \subseteq \mathcal{G}$.

(e) This follows from the fact that for all $s \le t$ in $\mathbb{R}_+$, $\pi_{[0,s]}^{[0,t]}$ is continuous w.r.t. $\mathcal{O}_{W_n^t}$-$\mathcal{O}_{W_n^s}$, and thus $\mathcal{B}_{W_n^t}$-$\mathcal{B}_{W_n^s}$-measurable. ☐

For every measurable space $(S, \mathcal{S})$, let $\mathcal{P}(S, \mathcal{S})$ denote the space of probability measures on $(S, \mathcal{S})$. Almost always, the σ-algebra $\mathcal{S}$ on S will be clear from the context, and in this case (in particular if $\mathcal{S}$ is the Borel σ-algebra of a given topology on S), then we write $\mathcal{P}(S)$ instead.

The study of McKean SDEs is facilitated by the introduction of the Wasserstein metrics, which allow to metrize, for every Polish space (S, d), the associated Wasserstein space $\mathcal{P}_p(S)$ from Definition D.1. The most important underlying Polish spaces in our context are $S = \mathbb{K}^n$ and $S = W_n^t$, $t \in \mathbb{R}_+$. Appendix D gives a brief overview of the main definitions and results from the theory of Wasserstein spaces which we will use throughout this thesis.

Definition 2.2 (The space $\widehat{\mathcal{P}}_p(W_n)$)**.** For each $p \in [1, \infty)$, let

$$\widehat{\mathcal{P}}_p(W_n) := \left\{ \mathbb{Q} \in \mathcal{P}(W_n) : \mathbb{Q} \circ \pi_{[0,t]}^{-1} \in \mathcal{P}_p(W_n^t) \text{ for all } t \in \mathbb{R}_+ \right\}. \tag{2.14}$$

The space $\widehat{\mathcal{P}}_p(W_n)$ is the set of all probability measures $\mathbb{Q}$ on W_n with the property that for every $t \in \mathbb{R}_+$, the pushforward measure $\mathbb{Q} \circ \pi_{[0,t]}^{-1}$ on $(W_n^t, \mathcal{B}_{W_n^t})$ has a finite p-th moment w.r.t. the seminorm p_t from (2.1). In the sequel, this space will serve as the domain of the measure argument of coefficients of McKean SDEs, see (2.34).

For every $t \in \mathbb{R}_+$, the projection $\pi_{[0,t]}$ from (2.8) induces a projection $\Pi_{[0,t]} \colon \widehat{\mathcal{P}}_p(W_n) \to \mathcal{P}_p(W_n^t)$ defined by

$$\Pi_{[0,t]}(\mathbb{Q}) = \mathbb{Q} \circ \pi_{[0,t]}^{-1}, \qquad \mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n). \tag{2.15}$$

In addition, we denote by

$$\Pi_t \colon \widehat{\mathcal{P}}_p(W_n) \to \mathcal{P}_p(\mathbb{K}^n), \quad \mathbb{Q} \mapsto \mathbb{Q} \circ \pi_t^{-1}, \qquad t \in \mathbb{R}_+, \tag{2.16}$$

the projection on $\widehat{\mathcal{P}}_p(W_n)$ which maps $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$ to its marginal distribution at time t. In order to simplify the notation, we will sometimes write

$$\mathbb{Q}_{[0,t]} = \Pi_{[0,t]}(\mathbb{Q}) \qquad \text{and} \qquad \mathbb{Q}_t = \Pi_t(\mathbb{Q}) \tag{2.17}$$

for each $t \in \mathbb{R}_+$.

Remark 2.3. For each $t \in \mathbb{R}_+$, the map $\Pi_{[0,t]}$ defined in (2.15) is surjective. To see this, fix $t \in \mathbb{R}_+$, let $Q \in \mathcal{P}_p(W_n^t)$ and consider the map

$$\iota \colon W_n^t \to W_n, \quad (w_s)_{s \in [0,t]} \mapsto (w_{s \wedge t})_{s \in \mathbb{R}_+}. \tag{2.18}$$

Since each projection $\pi_r \circ \iota \colon W_n^t \ni w \mapsto w_{r \wedge t} \in \mathbb{K}^n$ is continuous w.r.t. the topology $\mathcal{O}_{W_n^t}$ of uniform convergence on W_n^t, it follows that ι is $\mathcal{B}_{W_n^t}$-$\mathcal{G}$-measurable. Hence the pushforward measure $\mathbb{Q} := Q \circ \iota^{-1} \in \mathcal{P}(W_n)$ is well-defined. For each $u \in \mathbb{R}_+$, it follows by the law of the unconscious statistician (see [36, Theorem 4.10]) that

$$\int_{W_n} p_u^p(\tilde{w}) \, \mathrm{d}\mathbb{Q}(\tilde{w}) = \int_{W_n^t} p_u^p(\iota(w)) \, \mathrm{d}Q(w) = \int_{W_n^t} \underbrace{p_{u \wedge t}^p(w)}_{\le p_t^p(w)} \, \mathrm{d}Q(w) < \infty.$$

Hence $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$. Since $\pi_{[0,t]} \circ \iota$ is the identity map on W_n^t, we have

$$\Pi_{[0,t]}(\mathbb{Q}) = (Q \circ \iota^{-1}) \circ \pi_{[0,t]}^{-1} = Q,$$

proving surjectivity of $\Pi_{[0,t]}$.

We next define notation for the most important Wasserstein metrics and pseudometrics appearing in the sequel.

Definition 2.4 (Wasserstein metric on $\mathcal{P}_p(\mathbb{K}^n)$)**.** Let $\mathbb{K}^n$ carry the Euclidean topology. For every $p \in [1,\infty)$, we denote by d_p the Wasserstein metric of order p (Definition D.2) on $\mathcal{P}_p(\mathbb{K}^n)$, and equip $\mathcal{P}_p(\mathbb{K}^n)$ with the corresponding Borel σ-algebra $\mathcal{B}_{\mathcal{P}_p(\mathbb{K}^n)}$.

We will also use frequently the Wasserstein metrics for the underlying spaces (W_n^t, p_t) for $t \in \mathbb{R}_+$, and a corresponding family of pseudometrics on the spaces $\widehat{\mathcal{P}}_p(W_n)$ for $p \in [1,\infty)$. This will later be useful to apply the theory of pseudometric spaces, in particular Banach's fixed point theorem (Theorem A.10), on $\widehat{\mathcal{P}}_p(W_n)$. Recall from Definition D.2 that for every Polish space (S,d) and all probability measures P and Q on $\mathcal{B}_S$, $\mathcal{P}(P,Q)$ denotes the set of all probability measures $R \in \mathcal{P}(S \times S)$ which have P and Q as marginals.

Definition 2.5 (Wasserstein metric on $\mathcal{P}_p(W_n^t)$, pseudometrics on $\widehat{\mathcal{P}}_p(W_n)$)**.** For each $t \in \mathbb{R}_+$ and every $p \in [1,\infty)$, equip the space $\mathcal{P}_p(W_n^t)$ with the p-th order Wasserstein distance

$$\mathcal{T}_{t,p}(P,Q) = \inf_{R \in \mathcal{P}(P,Q)} \int_{W_n^t \times W_n^t} p_t^p(w - \tilde{w})\,\mathrm{d}R(w,\tilde{w})\,, \qquad P,Q \in \mathcal{P}_p(W_n^t), \tag{2.19}$$

and the p-th order Wasserstein metric (see Lemma D.5),

$$\mathcal{W}_{t,p} = \mathcal{T}_{t,p}^{1/p}. \tag{2.20}$$

Let $\mathcal{B}_{\mathcal{P}_p(W_n^t)}$ denote the Borel σ-algebra induced by $\mathcal{W}_{t,p}$ on $\mathcal{P}_p(W_n^t)$. Then we define in addition

$$\widehat{\mathcal{T}}_{t,p}(\mathbb{Q},\tilde{\mathbb{Q}}) = \mathcal{T}_{t,p}(\mathbb{Q}_{[0,t]},\tilde{\mathbb{Q}}_{[0,t]}) \qquad \text{and} \qquad \widehat{\mathcal{W}}_{t,p}(\mathbb{Q},\tilde{\mathbb{Q}}) = \mathcal{W}_{t,p}(\mathbb{Q}_{[0,t]},\tilde{\mathbb{Q}}_{[0,t]}). \tag{2.21}$$

for all $\mathbb{Q},\tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}_p(W_n)$, see also Definition 2.2.

Remark 2.6 (Properties of $\mathcal{W}_{t,p}$ and $\widehat{\mathcal{W}}_{t,p}$)**.** Fix $t \in \mathbb{R}_+$ and $p \in [1,\infty)$.

(a) By Theorem D.6, convergence of probability measures in $\mathcal{P}_p(W_n^t)$ w.r.t. $\mathcal{W}_{t,p}$ implies weak convergence. Since the topology of weak convergence on $\mathcal{P}_p(W_n^t)$ is metrizable (e.g. by the Prokhorov metric, see [53, Theorem 8.50]), it follows that the topology $\mathcal{O}_{\mathcal{W}_{t,p}}$ induced on $\mathcal{P}_p(W_n^t)$ by the metric $\mathcal{W}_{t,p}$ is finer than the topology of weak convergence.

(b) Since (W_n^t, p_t) is a Polish space by Lemma 2.1 (a), it follows from Theorem D.8 that $(\mathcal{P}_p(W_n^t), \mathcal{W}_{t,p})$ is a Polish space.

(c) By Remark 2.3, the map $\Pi_{[0,t]}$ from (2.15) is surjective. Thus by Remark 2.6 (b) and Lemma A.6 (b), $(\widehat{\mathcal{P}}_p(W_n), \widehat{\mathcal{W}}_{t,p})$ is a complete pseudometric space. In addition, by Lemma A.6 (a), the topology $\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}$ induced on $\widehat{\mathcal{P}}_p(W_n)$ by $\widehat{\mathcal{W}}_{t,p}$ coincides with the initial topology induced by $\Pi_{[0,t]}$, i.e.,

$$\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}} = \Pi_{[0,t]}^{-1}(\mathcal{O}_{\mathcal{W}_{t,p}}). \tag{2.22}$$

In particular, since $\mathcal{O}_{\mathcal{W}_{t,p}}$ is second countable by Remark 2.6 (b), so is $\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}$.

(d) By definiteness of the metric $\mathcal{W}_{t,p}$ and the definition of $\widehat{\mathcal{W}}_{t,p}$, we have for all $\mathbb{Q},\tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}_p(W_n)$, using the notation (2.17), the equivalence

$$\widehat{\mathcal{W}}_{t,p}(\mathbb{Q},\tilde{\mathbb{Q}}) = 0 \qquad \Longleftrightarrow \qquad \mathbb{Q}_{[0,t]} = \tilde{\mathbb{Q}}_{[0,t]}.$$

In particular, if $\widehat{\mathcal{W}}_{k,p}(\mathbb{Q},\tilde{\mathbb{Q}}) = 0$ for all $k \in \mathbb{N}$, then by (2.15) and Lemma 2.1(c), $\mathbb{Q}{\restriction_{\mathcal{G}_k}} = \tilde{\mathbb{Q}}{\restriction_{\mathcal{G}_k}}$ for all $k \in \mathbb{N}$. Since $\bigcup_{k\in\mathbb{N}} \mathcal{G}_k$ is an algebra generating $\mathcal{G}$, $\mathbb{Q} = \tilde{\mathbb{Q}}$ follows from the usual uniqueness criterion for probability measures, see [53, Corollary 15.68]. Therefore, $(\widehat{\mathcal{W}}_{k,p})_{k\in\mathbb{N}}$ is a separating family of pseudometrics (Definition A.13) on $\widehat{\mathcal{P}}_p(W_n)$.

(e) For all $s \le t$ in $\mathbb{R}_+$,

$$\widehat{\mathcal{T}}_{s,p} \le \widehat{\mathcal{T}}_{t,p} \qquad \text{and} \qquad \widehat{\mathcal{W}}_{s,p} \le \widehat{\mathcal{W}}_{t,p}. \tag{2.23}$$

To see this, it suffices by (2.21) to show that for all $P, Q \in \mathcal{P}_p(W_n^t)$,

$$\mathcal{T}_{s,p}(P_{[0,s]}, Q_{[0,s]}) \le \mathcal{T}_{t,p}(P,Q), \tag{2.24}$$

where $P_{[0,s]} \coloneqq P \circ {\pi^{[0,t]}_{[0,s]}}^{-1}$ and $Q_{[0,s]} \coloneqq Q \circ {\pi^{[0,t]}_{[0,s]}}^{-1}$, see also (2.9). Let $R \in \mathcal{P}(P,Q)$. By measurability of $\pi^{[0,t]}_{[0,s]}$, see Lemma 2.1(e), the map

$$\varphi\colon W_n^t \times W_n^t \ni (w,\tilde{w}) \mapsto (\pi^{[0,t]}_{[0,s]}(w), \pi^{[0,t]}_{[0,s]}(\tilde{w})) \in W_n^s \times W_n^s$$

is measurable. Hence $R_{[0,s]} \coloneqq R \circ \varphi^{-1} \in \mathcal{P}(W_n^s \times W_n^s)$ is well defined. In addition, $R_{[0,s]} \in \mathcal{P}(P_{[0,s]}, Q_{[0,s]})$, since for all $A, B \in \mathcal{B}_{W_n^s}$,

$$R_{[0,s]}(A\times B) = R\left({\pi^{[0,t]}_{[0,s]}}^{-1}(A) \times {\pi^{[0,t]}_{[0,s]}}^{-1}(B)\right) = \begin{cases} P_{[0,s]}(A) & \text{if } B = W_n^s, \\ Q_{[0,s]}(B) & \text{if } A = W_n^s. \end{cases}$$

By (2.19) and the law of the unconscious statistician, we obtain

$$\begin{aligned} \mathcal{T}_{s,p}(P_{[0,s]}, Q_{[0,s]}) &\le \int_{W_n^s\times W_n^s} p_s^p(w-\tilde{w})\,\mathrm{d}R_{[0,s]}(w,\tilde{w}) \\ &= \int_{W_n^t\times W_n^t} p_s^p(w-\tilde{w})\,\mathrm{d}R(w,\tilde{w}) \le \int_{W_n^t\times W_n^t} p_t^p(w-\tilde{w})\,\mathrm{d}R(w,\tilde{w})\,. \end{aligned}$$

Since $R \in \mathcal{P}(P,Q)$ was arbitrary, comparison with (2.19) yields (2.24).

(f) Similar arguments as in Remark 2.6(e) show the inequality

$$d_p(\mathbb{Q}_t, \tilde{\mathbb{Q}}_t) \le \widehat{\mathcal{W}}_{t,p}(\mathbb{Q},\tilde{\mathbb{Q}}), \qquad \mathbb{Q},\tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}_p(W_n), \tag{2.25}$$

for all $t \in \mathbb{R}_+$ and every $p \in [1,\infty)$, and where d_p denotes the Wasserstein metric on $\mathcal{P}_p(\mathbb{K}^n)$ from Definition 2.4.

From now on, we fix $p \in [1, \infty)$. The next result, Proposition 2.7, shows that $\widehat{\mathcal{P}}_p(W_n)$ is a Polish space w.r.t. the topology induced by the family $(\widehat{\mathcal{W}}_{k,p})_{k\in\mathbb{N}}$ of pseudometrics (see Definition A.12). This topology is metrizable, see (2.26), and can be seen as the analogue for $\widehat{\mathcal{P}}_p(W_n)$ to the topology of compact convergence on W_n.

Proposition 2.7 ($\widehat{\mathcal{P}}_p(W_n)$ is a Polish space)**.** *Consider the space $\widehat{\mathcal{P}}_p(W_n)$ defined in (2.14), let $(\widehat{\mathcal{W}}_{k,p})_{k\in\mathbb{N}}$ be the separating family of complete pseudometrics from Remarks 2.6(c) and (d), and denote by $\mathcal{O}_{\widehat{\mathcal{P}}_p(W_n)}$ the topology induced by $(\widehat{\mathcal{W}}_{k,p})_{k\in\mathbb{N}}$ (see Definition A.12). Then $(\widehat{\mathcal{P}}_p(W_n), \mathcal{O}_{\widehat{\mathcal{P}}_p(W_n)})$ is a Polish space, and for each null sequence $(c_k)_{k\in\mathbb{N}}$ in $\mathbb{R}_+^*$, a compatible metric is given by*

$$\widehat{\mathcal{W}}_p(\mathbb{Q}, \tilde{\mathbb{Q}}) := \max_{k\in\mathbb{N}}\big(c_k \wedge \widehat{\mathcal{W}}_{k,p}(\mathbb{Q}, \tilde{\mathbb{Q}})\big), \qquad \mathbb{Q}, \tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}_p(W_n). \tag{2.26}$$

Proof. We conduct the proof in three steps. In Step 1, verify that $\widehat{\mathcal{W}}_p$ defined by (2.26) metrizes $\mathcal{O}_{\widehat{\mathcal{P}}_p(W_n)}$. Subsequently in Steps 2 and 3, we show that this metric is complete and separable.

Step 1. Since $(\widehat{\mathcal{W}}_{k,p})_{k\in\mathbb{N}}$ is a separating family of complete pseudometrics, it follows from Remark A.15(b) that $\widehat{\mathcal{W}}_p$ as defined in (2.26) metrizes $\mathcal{O}_{\widehat{\mathcal{P}}_p(W_n)}$.

Step 2 (The metric $\widehat{\mathcal{W}}_p$ is complete). Let $(\mathbb{Q}_m)_{m\in\mathbb{N}}$ be a Cauchy sequence[5] in $\widehat{\mathcal{P}}_p(W_n)$ w.r.t. $\widehat{\mathcal{W}}_p$. By (2.26), $(\mathbb{Q}_m)_{m\in\mathbb{N}}$ is Cauchy w.r.t. $\widehat{\mathcal{W}}_{k,p}$, for each $k \in \mathbb{N}$. From Remark 2.6(c), $\widehat{\mathcal{W}}_{k,p}$ is a complete pseudometric (Definition A.5(b)) for each $k \in \mathbb{N}$. Hence there exist probability measures $(\mathbb{Q}^{(k)})_{k\in\mathbb{N}}$ in $\widehat{\mathcal{P}}_p(W_n)$ with

$$\widehat{\mathcal{W}}_{k,p}(\mathbb{Q}_m, \mathbb{Q}^{(k)}) \xrightarrow[m\to\infty]{} 0, \qquad k \in \mathbb{N}. \tag{2.27}$$

Fix $k \le l$ in $\mathbb{N}$ for a moment. By the monotonicity property (2.23),

$$\widehat{\mathcal{W}}_{k,p}(\mathbb{Q}_m, \mathbb{Q}^{(l)}) \xrightarrow[m\to\infty]{} 0. \tag{2.28}$$

From (2.27) and (2.28), $(\mathbb{Q}_m)_{m\in\mathbb{N}}$ converges w.r.t. $\widehat{\mathcal{W}}_{k,p}$ to both $\mathbb{Q}^{(k)}$ and $\mathbb{Q}^{(l)}$. Hence by the triangle inequality for $\widehat{\mathcal{W}}_{k,p}$,

$$\widehat{\mathcal{W}}_{k,p}(\mathbb{Q}^{(k)}, \mathbb{Q}^{(l)}) = 0.$$

From Remark 2.6(d),

$$\mathbb{Q}^{(k)}_{[0,k]} = \mathbb{Q}^{(l)}_{[0,k]}, \qquad k \le l \text{ in } \mathbb{N}. \tag{2.29}$$

Corollary B.21 yields the existence of a unique extension $\mathbb{Q} \in \mathcal{P}(W_n)$ with

$$\mathbb{Q}_{[0,k]} = \mathbb{Q}^{(k)}_{[0,k]}, \qquad k \in \mathbb{N}. \tag{2.30}$$

[5] In the context of Proposition 2.7, we do not use the notation defined in (2.17), but rather reserve the subscript to denote sequence elements.

For each $k \in \mathbb{N}$, (2.30) together with the definition (2.21) of the pseudometric $\widehat{\mathcal{W}}_{k,p}$ implies that $\widehat{\mathcal{W}}_{k,p}(\mathbb{Q}^{(k)}, \mathbb{Q}) = 0$. By the triangle inequality for $\widehat{\mathcal{W}}_{k,p}$ and (2.27),

$$\widehat{\mathcal{W}}_{k,p}(\mathbb{Q}_m, \mathbb{Q}) \leq \widehat{\mathcal{W}}_{k,p}(\mathbb{Q}_m, \mathbb{Q}^{(k)}) + \widehat{\mathcal{W}}_{k,p}(\mathbb{Q}^{(k)}, \mathbb{Q}) = \widehat{\mathcal{W}}_{k,p}(\mathbb{Q}_m, \mathbb{Q}^{(k)}) \xrightarrow[m\to\infty]{} 0,$$

i.e. $(\mathbb{Q}_m)_{m\in\mathbb{N}}$ converges to $\mathbb{Q}$ w.r.t. $\widehat{\mathcal{W}}_{k,p}$. Since this holds for all $k \in \mathbb{N}$, we obtain from (2.26) that $\widehat{\mathcal{W}}_p(\mathbb{Q}_m, \mathbb{Q}) \xrightarrow[m\to\infty]{} 0$. Therefore, $\widehat{\mathcal{W}}_p$ is complete.

Step 3 (The space $(\widehat{\mathcal{P}}_p(W_n), \widehat{\mathcal{W}}_p)$ is separable). Fix $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$ and $\varepsilon > 0$. For each $N \in \mathbb{N}$, $(\mathcal{P}_p(W_n^N), \mathcal{W}_{N,p})$ is Polish by Remark 2.6 (b) and thus has a countable dense subset $\tilde{\mathcal{D}}_N$. By Remark 2.3, the projections $\Pi_{[0,N]}$, $N \in \mathbb{N}$ are surjective, hence for each $Q \in \tilde{\mathcal{D}}_N$, there exists a $\mathbb{Q}_Q \in \widehat{\mathcal{P}}_p(W_n)$ with $\Pi_{[0,N]}(\mathbb{Q}_Q) = Q$. We will show that the countable set

$$\mathcal{D} := \bigcup_{N\in\mathbb{N}} \{\mathbb{Q}_Q : Q \in \tilde{\mathcal{D}}_N\}$$

is dense in $\widehat{\mathcal{P}}_p(W_n)$. Let $N \in \mathbb{N}$ be such that $\max_{k\geq N} c_k \leq \varepsilon$. For each $\tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}_p(W_n)$, using (2.26),

$$\begin{aligned} \widehat{\mathcal{W}}_p(\mathbb{Q}, \tilde{\mathbb{Q}}) &\leq \max_{k\leq N} \overbrace{\big(c_k \wedge \widehat{\mathcal{W}}_{k,p}(\mathbb{Q}, \tilde{\mathbb{Q}})\big)}^{\leq \widehat{\mathcal{W}}_{N,p}(\mathbb{Q},\tilde{\mathbb{Q}})} \vee \varepsilon \\ &\leq \widehat{\mathcal{W}}_{N,p}(\mathbb{Q}, \tilde{\mathbb{Q}}) \vee \varepsilon, \end{aligned} \tag{2.31}$$

where we used the monotonicity property (2.23) from Remark 2.6 (e) for the last inequality. Since the set $\tilde{\mathcal{D}}_N$ obtained at the beginning of Step 3 is dense in $(\mathcal{P}_p(W_n^N), \mathcal{W}_{N,p})$, there exists a $\tilde{Q} \in \tilde{\mathcal{D}}_N$ with $\mathcal{W}_{N,p}(\mathbb{Q}_{[0,N]}, \tilde{Q}) \leq \varepsilon$. Let $\tilde{\mathbb{Q}} := \mathbb{Q}_{\tilde{Q}}$ be the corresponding preimage in $\mathcal{D}$. Then from (2.21),

$$\widehat{\mathcal{W}}_{N,p}(\mathbb{Q}, \tilde{\mathbb{Q}}) = \mathcal{W}_{N,p}(\mathbb{Q}_{[0,N]}, \tilde{\mathbb{Q}}_{[0,N]}) \leq \varepsilon.$$

Combining the previous display with (2.31), it follows that

$$\widehat{\mathcal{W}}_p(\mathbb{Q}, \tilde{\mathbb{Q}}) \leq \varepsilon. \qquad \square$$

Definition 2.8 (Measurable structure on $\widehat{\mathcal{P}}_p(W_n)$)**.** For every $t \in \mathbb{R}_+$, let $\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}$ denote the topology induced by the pseudometric $\widehat{\mathcal{W}}_{t,p}$ on $\widehat{\mathcal{P}}_p(W_n)$, see Definition A.3. By Remark 2.6 (e), it holds that $\mathcal{O}_{\widehat{\mathcal{W}}_{s,p}} \subseteq \mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}$ for all $s \leq t$ in $\mathbb{R}_+$. We equip the space $\widehat{\mathcal{P}}_p(W_n)$ with the σ-algebra

$$\mathcal{H} := \sigma\Big(\bigcup_{t\in\mathbb{R}_+} \mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}\Big), \tag{2.32}$$

and the filtration $\mathbb{H} = (\mathcal{H}_t)_{t\in\mathbb{R}_+}$, which is defined as the family of Borel σ-algebras

$$\mathcal{H}_t = \sigma\big(\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}\big), \qquad t \in \mathbb{R}_+. \tag{2.33}$$

Remark 2.9. (a) For every $t \in \mathbb{R}_+$, Remark 2.6 (c) implies that $\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}} = \Pi_{[0,t]}^{-1}(\mathcal{O}_{\mathcal{W}_{t,p}})$, and thus

$$\mathcal{H}_t = \sigma(\Pi_{[0,t]}^{-1}(\mathcal{O}_{\mathcal{W}_{t,p}})) = \Pi_{[0,t]}^{-1}(\mathcal{B}_{\mathcal{P}_p(W_n^t)}).$$

Therefore $\mathcal{H}_t$ is the initial σ-algebra on $\widehat{\mathcal{P}}_p(W_n)$ induced by $\Pi_{[0,t]}$ when $\mathcal{P}_p(W_n^t)$ is equipped with the topology induced by $\mathcal{W}_{t,p}$ and the associated Borel σ-algebra $\mathcal{B}_{\mathcal{P}_p(W_n^t)}$.

(b) By Lemma D.15, $\mathcal{H}_t$ coincides with the σ-algebra generated by the map $\Pi_{[0,t]}$ from (2.15) when $\mathcal{P}_p(W_n^t)$ is equipped with the weak topology and the corresponding Borel σ-algebra.

Lemma 2.10. *The process* $(\Pi_t)_{t\in\mathbb{R}_+}$ *on* $(\widehat{\mathcal{P}}_p(W_n), \mathcal{H})$ *with* $\Pi_t \colon \widehat{\mathcal{P}}_p(W_n) \ni \mathbb{Q} \mapsto \mathbb{Q}_t \in \mathcal{P}_p(\mathbb{K}^n)$ *defined in* (2.16) *and* (2.17) *is continuous and* $\mathbb{H}$*-progressive when* $\mathcal{P}_p(\mathbb{K}^n)$ *is equipped with the Borel σ-algebra induced by the Wasserstein metric* d_p *on* $\mathcal{P}_p(\mathbb{K}^n)$ *(see Definition 2.4).*

Proof. We verify that for every $t \in \mathbb{R}_+$, the restriction $\Pi\restriction_{[0,t]\times\widehat{\mathcal{P}}_p(W_n)}$ is $\mathcal{B}_{[0,t]} \otimes \mathcal{H}_t$-measurable. Since $\mathcal{H}_t$ is the Borel σ-algebra associated with the pseudometric $\widehat{\mathcal{W}}_{t,p}$, see (2.33), it suffices by Lemma B.10 to show that $\Pi\restriction_{[0,t]\times\widehat{\mathcal{P}}_p(W_n)}$ is continuous separately in each argument, where $[0,t]$ carries the usual Euclidean topology, $\widehat{\mathcal{P}}_p(W_n)$ carries the topology induced by the pseudometric $\widehat{\mathcal{W}}_{t,p}$ and the codomain $\mathcal{P}_p(\mathbb{K}^n)$ is metrized by the Wasserstein metric d_p.

For each fixed $s \in [0,t]$, continuity of $\widehat{\mathcal{P}}_p(W_n) \ni \mathbb{Q} \mapsto \Pi_s(\mathbb{Q}) \in \mathcal{P}_p(\mathbb{K}^n)$ follows, since by (2.25) and (2.23),

$$d_p(\Pi_s(\mathbb{Q}), \Pi_s(\tilde{\mathbb{Q}})) = d_p(\mathbb{Q}_s, \tilde{\mathbb{Q}}_s) \le \widehat{\mathcal{W}}_{s,p}(\mathbb{Q}, \tilde{\mathbb{Q}}) \le \widehat{\mathcal{W}}_{t,p}(\mathbb{Q}, \tilde{\mathbb{Q}})$$

holds for all $\mathbb{Q}, \tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}_p(W_n)$.

It remains to show continuity of $[0,t] \ni s \mapsto \Pi_s(\mathbb{Q}) \in \mathcal{P}_p(\mathbb{K}^n)$ for every fixed $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$. Consider the probability space $(W_n, \mathcal{G}, \mathbb{Q})$. On this space, the canonical process $(\pi_s)_{s\in\mathbb{R}_+}$ is continuous with law $\mathbb{Q}$, and $\Pi_s(\mathbb{Q}) = \mathbb{Q}_s = \mathcal{L}_{\mathbb{Q}}[\pi_s]$ for every $s \in [0,t]$. For each sequence $(s_k)_{k\in\mathbb{N}}$ in $[0,t]$ converging to some $s \in [0,t]$, we have $\pi_{s_k} \to \pi_s$ as $k \to \infty$, pointwise on W_n. This implies that $\mathbb{Q}_{s_k} = \mathcal{L}_{\mathbb{Q}}[\pi_{s_k}] \to \mathcal{L}_{\mathbb{Q}}[\pi_s] = \mathbb{Q}_s$ weakly as $k \to \infty$. In addition, since $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$, it follows by (2.14) that $\mathbb{E}_{\mathbb{Q}}[\sup_{r\in[0,t]}|\pi_r|^p] < \infty$. By dominated convergence,

$$\int_{\mathbb{K}^n} |x|^p \,\mathrm{d}\mathbb{Q}_{s_k}(x) = \mathbb{E}_{\mathbb{Q}}[|\pi_{s_k}|^p] \xrightarrow[k\to\infty]{} \mathbb{E}_{\mathbb{Q}}[|\pi_s|^p] = \mathbb{E}_{\mathbb{Q}}[|\pi_s|^p] = \int_{\mathbb{K}^n} |x|^p \,\mathrm{d}\mathbb{Q}_s(x)\,.$$

Using Theorem D.6, (c) ⇒ (a), $d_p(\mathbb{Q}_{s_k}, \mathbb{Q}_s) \xrightarrow[k\to\infty]{} 0$ follows. □

Compared with the case of classical SDEs, where coefficients are assumed to be progressive w.r.t. the filtrations $\mathbb{G}$ respectively $\mathbb{G}\otimes\mathbb{F}$ in the case of random dependence, see Definitions C.6 and C.8, we will require in the case of McKean SDEs progressive measurability of coefficients w.r.t. a filtration which reflects the dependence on the measure argument.

Definition 2.11 (The filtrations $\mathbb{G} \otimes \mathbb{H}$ and $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$)**.** Consider the measurable spaces $(W_n, \mathcal{G})$ and $(\widehat{\mathcal{P}}_p(W_n), \mathcal{H})$, see (2.4) and (2.32), and let $\mathbb{G} = (\mathcal{G}_t)_{t\in\mathbb{R}_+}$ and $\mathbb{H} = (\mathcal{H}_t)_{t\in\mathbb{R}_+}$ be the filtrations defined in (2.5) and (2.33), respectively.

(a) We denote by $\mathbb{G} \otimes \mathbb{H}$ the filtration $(\mathcal{G}_t \otimes \mathcal{H}_t)_{t \geq 0}$ of $(W_n \times \widehat{\mathcal{P}}_p(W_n), \mathcal{G} \otimes \mathcal{H})$.

(b) Given a filtered measurable space $(\Omega, \mathcal{F}, \mathbb{F})$ with $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$, we denote by $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$ the filtration $(\mathcal{G}_t \otimes \mathcal{H}_t \otimes \mathcal{F}_t)_{t \in \mathbb{R}_+}$ of $(W_n \times \widehat{\mathcal{P}}_p(W_n) \times \Omega,\ \mathcal{G} \otimes \mathcal{H} \otimes \mathcal{F})$.

2.2. Strong Existence and Pathwise Uniqueness for McKean SDEs under Lipschitz Continuity and Bounded Growth Conditions

McKean stochastic differential equations were introduced by H.P. McKean in [40] in a form similar to (2.43) below. We consider in this section McKean SDEs with coefficients b and σ depending on time, the trajectory of the solution, the law of the solution and, possibly, the underlying randomness. The dependence of these functions is assumed to be non-anticipative in the sense that the future evolution of the solution process at every time $t \in \mathbb{R}_+$ is determined only by information from times up to t. That is, we admit as coefficients $\mathbb{G} \otimes \mathbb{H}$-progressive functions

$$b\colon \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \to \mathbb{K}^n \quad \text{and} \quad \sigma\colon \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \to \mathbb{K}^{n \times d}, \tag{2.34}$$

or, given a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ (Definition C.5 (a)), $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$-progressive functions

$$b\colon \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \times \Omega \to \mathbb{K}^n \quad \text{and} \quad \sigma\colon \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \times \Omega \to \mathbb{K}^{n \times d}, \tag{2.35}$$

with the filtrations $\mathbb{G} \otimes \mathbb{H}$ and $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$ from Definition 2.11.

Definition 2.12 (McKean stochastic differential equation)**.** A *McKean stochastic differential equation* (MSDE) with $\mathbb{G} \otimes \mathbb{H}$-progressive coefficients b and σ as in (2.34) has the form

$$\mathrm{d}X_t = b(t, X, \mathcal{L}[X])\,\mathrm{d}t + \sigma(t, X, \mathcal{L}[X])\,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+, \tag{2.36}$$

or, given a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$-progressive coefficients b and σ as in (2.35),

$$\mathrm{d}X_t = b(t, X, \mathcal{L}[X], \,\cdot\,)\,\mathrm{d}t + \sigma(t, X, \mathcal{L}[X], \,\cdot\,)\,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+, \tag{2.37}$$

where in both cases $\mathcal{L}[X]$ denotes the law of the solution X and $B = (B_t)_{t \in \mathbb{R}_+}$ is an $\mathbb{R}^d$- or $\mathbb{C}^d$-valued Brownian motion.[6] In the first case, we speak of *nonrandom dependence* of the SDE coefficients on the solution. Otherwise, we speak of *random dependence*.

The notations (2.36) and (2.37) are made precise by defining what constitutes a solution of these SDEs. This is done starting with Definition 2.25 below. As with classical SDEs (see Definition C.9), we distinguish several special cases of (2.36).

Definition 2.13 (Instantaneous and homogeneous dependence)**.** Consider the setting of Definition 2.12.

[6] We use $\cdot$ to highlight the presence of random dependence of coefficients. This leads to a minor abuse of notation when these coefficients appear in stochastic integrals with variable limits, see Remark C.16, which we tolerate in order to improve readability.

(a) If $\hat{b}\colon \mathbb{R}_+ \times \mathbb{K}^n \times \mathcal{P}_p(\mathbb{K}^n) \times \Omega \to \mathbb{K}^n$ and $\hat{\sigma}\colon \mathbb{R}_+ \times \mathbb{K}^n \times \mathcal{P}_p(\mathbb{K}^n) \times \Omega \to \mathbb{K}^{n\times d}$ are such that for all $(t, w, \mathbb{Q}, \omega) \in \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \times \Omega$,

$$b(t, w, \mathbb{Q}, \omega) = \hat{b}(t, w_t, \mathbb{Q}_t, \omega) \quad \text{and} \quad \sigma(t, w, \mathbb{Q}, \omega) = \hat{\sigma}(t, w_t, \mathbb{Q}_t, \omega), \tag{2.38}$$

see also (2.17), then we speak of *instantaneous dependence* of the SDE coefficients on the solution. In this case, we make no notational distinction between b and $\hat{b}$ and σ and $\hat{\sigma}$, respectively, and write the SDE (2.36) as

$$\mathrm{d}X_t = b(t, X_t, \mathcal{L}[X_t], \,\cdot\,)\,\mathrm{d}t + \sigma(t, X_t, \mathcal{L}[X_t], \,\cdot\,)\,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+. \tag{2.39}$$

(b) If in the setting of part (a), $\hat{b}$ and $\hat{\sigma}$ do not depend on ω, i.e. if

$$b(t, w, \mathbb{Q}) = \hat{b}(t, w_t, \mathbb{Q}_t) \quad \text{and} \quad \sigma(t, w, \mathbb{Q}) = \hat{\sigma}(t, w_t, \mathbb{Q}_t), \tag{2.40}$$

for all $(t, w, \mathbb{Q}) \in \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n)$, for suitable functions

$$\hat{b}\colon \mathbb{R}_+ \times \mathbb{K}^n \times \mathcal{P}_p(\mathbb{K}^n) \to \mathbb{K}^n \quad \text{and} \quad \hat{\sigma}\colon \mathbb{R}_+ \times \mathbb{K}^n \times \mathcal{P}_p(\mathbb{K}^n) \to \mathbb{K}^{n\times d}, \tag{2.41}$$

then we write

$$\mathrm{d}X_t = b(t, X_t, \mathcal{L}[X_t])\,\mathrm{d}t + \sigma(t, X_t, \mathcal{L}[X_t])\,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+. \tag{2.42}$$

(c) If in the setting of part (b), $\hat{b}$ and $\hat{\sigma}$ do not depend on t, then we write

$$\mathrm{d}X_t = b(X_t, \mathcal{L}[X_t])\,\mathrm{d}t + \sigma(X_t, \mathcal{L}[X_t])\,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+. \tag{2.43}$$

and speak of *time-homogeneous dependence* of the coefficients on the solution.

Remark 2.14 (History of McKean SDEs and related literature)**.** McKean SDEs were introduced in [40], where the case (2.43) of time-homogeneous dependence of the coefficients on the solution was discussed. Since then, various extensions have been considered. Our work is based on [33] (who consider a finite time horizon, the case $\mathbb{K} = \mathbb{R}$ and $p = 2$ with nonrandom, instantaneous dependence, as in (2.42), with the driving Brownian motion replaced by a Lévy process) and [10, §1.3.1] (finite time horizon, the case $\mathbb{K} = \mathbb{R}$ and $p = 2$ with instantaneous and random dependence (see (2.39))).
The form (2.34) of the coefficients considered in this thesis appears in similar generality in [11, Example 1, Point 4] (with a finite time horizon and $\mathbb{K} = \mathbb{R}$), but neither [11] nor the sources cited in [11, Example 1, Point 4] actually consider such SDEs in full generality.

Pathwise dependence of coefficients on both the solution and its law is considered in the setting of stochastic optimal control theory in several sources, e.g. [37] (the case $\mathbb{K} = \mathbb{R}$, with dependence of the drift on both the (past) path of solution and its law and dependence of the volatility on the (past) path of the solution but not its law), [13] and [19, §2.2] ($\mathbb{K} = \mathbb{R}$, finite time horizon, $p = 2$ and nonrandom dependence).[7]

[7] In [13], SDEs for Hilbert space-valued solutions are considered. The last two sources, [13] and [19] are quite technical and beyond the scope of this thesis.

Of the literature surveyed, the works closest to this thesis in spirit and results is [2], where a finite time horizon, $\mathbb{K} = \mathbb{R}$, $p \geq 2$ and dependence of coefficients on time, the (past) path of the solution as well as the curve of (past) one-dimensional marginal distributions of the solution are considered.[8] [2, Theorem 1.1] establishes strong existence and pathwise uniqueness result for McKean SDEs with coefficients as described above, assuming (see [2, Assumption (I)]) Lipschitz continuity and continuity in the time argument of coefficients. [2, Theorem 1.2] establishes convergence of the particle approximation and the propagation of chaos. The analogues of these two results from [2] in this text are Theorems 2.32, 2.56, 2.58 and 2.60. The main differences between our setting and that of [2] is that we consider the infinite time horizon, the cases $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$ and dependence of coefficients on randomness (only in the context of Theorem 2.32) and the law of the past path of solutions (instead of the curve of one-dimensional marginal distributions).

In order to obtain strong existence and uniqueness, pathwise and in law, of solutions of (2.36) under continuity and boundedness assumptions on the coefficients (Theorem 2.32), but also when discussing the propagation of chaos (Section 2.4), we also consider, for $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$ fixed, McKean SDEs of the form

$$\mathrm{d}X_t = b(t, X, \mathbb{Q})\,\mathrm{d}t + \sigma(t, X, \mathbb{Q})\,\mathrm{d}B_t\,, \tag{2.44}$$

where b and σ are as in (2.34). Furthermore, given a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and a suitable random probability measure[9] $\mathbb{Q}\colon \Omega \to \widehat{\mathcal{P}}_p(W_n)$, we also discuss McKean SDEs of the form

$$\mathrm{d}X_t = b(t, X, \mathbb{Q}, \,\cdot\,)\,\mathrm{d}t + \sigma(t, X, \mathbb{Q}, \,\cdot\,)\,\mathrm{d}B_t\,, \tag{2.45}$$

for b and σ as in (2.35).[10] We begin with an elementary measurability result which ensures well-posedness of these SDEs.

Lemma 2.15 ($\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$-progressive measurability and composition with adapted random measures)**.** *Let $(\Omega, \mathcal{F}, \mathbb{F})$ with $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$ be a filtered measurable space, $f\colon \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \times \Omega \to \mathbb{K}^m$ a $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$-progressive map and $\mathbb{Q}\colon \Omega \to \widehat{\mathcal{P}}_p(W_n)$ a random probability measure which is measurable w.r.t. $\mathcal{F}_t$-$\mathcal{H}_t$ for each $t \in \mathbb{R}_+$.*

(a) The map

$$\tilde{f}\colon \mathbb{R}_+ \times W_n \times \Omega \to \mathbb{K}^m, \quad \tilde{f}(t, w, \omega) := f(t, w, \mathbb{Q}(\omega), \omega)$$

is $\mathbb{G} \otimes \mathbb{F}$-progressive (see Definition C.6).

(b) Let $X\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^n$ be an $\mathbb{F}$-progressive process. Then the composition

$$\mathbb{R}_+ \times \Omega \ni (t, \omega) \mapsto f(t, X(\omega), \mathbb{Q}(\omega), \omega) \in \mathbb{K}^m \tag{2.46}$$

is $\mathbb{F}$-progressive.

---

[8] In a related paper, [3], a convergence result for a numerical particle method (i.e. convergence of a particle-based interpolated Euler scheme) is shown. This result, while highly relevant for applications, is beyond the scope of this thesis.

[9] In the proof of the existence and uniqueness theorem 2.32 for strong solutions of MSDEs, the measure $\mathbb{Q}$ in (2.45) will always be constant. In the context of the propagation of chaos, $\mathbb{Q}$ will be an empirical measure (Definition 2.44) associated with certain approximating processes.

[10] For readability, we write $\mathbb{Q}$ instead of $\mathbb{Q}(\,\cdot\,)$ in (2.45).

Proof. (a) Fix $t \in \mathbb{R}_+$ and observe that the restriction $\tilde{f}\restriction_{[0,t]\times W_n\times\Omega}$ is the composition of the restriction $f\restriction_{[0,t]\times W_n\times\widehat{\mathcal{P}}_p(W_n)\times\Omega}$ and the map

$$\varphi\colon [0,t]\times W_n\times\Omega \ni (s,w,\omega)\mapsto (s,w,\mathbb{Q}(\omega),\omega)\in[0,t]\times W_n\times\widehat{\mathcal{P}}_p(W_n)\times\Omega.$$

By the assumed $\mathbb{G}\otimes\mathbb{H}\otimes\mathbb{F}$-progressive measurability of f, $f\restriction_{[0,t]\times W_n\times\widehat{\mathcal{P}}_p(W_n)\times\Omega}$ is $\mathcal{B}_{[0,t]}\otimes\mathcal{G}_t\otimes\mathcal{H}_t\otimes\mathcal{F}_t$-measurable. Therefore, it suffices to check that φ is $(\mathcal{B}_{[0,t]}\otimes\mathcal{G}_t\otimes\mathcal{F}_t)$-$(\mathcal{B}_{[0,t]}\otimes\mathcal{G}_t\otimes\mathcal{H}_t\otimes\mathcal{F}_t)$-measurable. The required componentwise measurability is trivial for the first, second and fourth component maps of φ, and holds by assumption for the third component map, $(s,w,\omega)\mapsto\mathbb{Q}(\omega)$. Hence φ is $(\mathcal{B}_{[0,t]}\otimes\mathcal{G}_t\otimes\mathcal{F}_t)$-$(\mathcal{B}_{[0,t]}\otimes\mathcal{G}_t\otimes\mathcal{H}_t\otimes\mathcal{F}_t)$-measurable, as claimed.

(b) This follows from Lemma C.7(a) applied to $\tilde{f}$. □

Remark 2.16 (The SDEs (2.44) and (2.45) can be viewed as classical SDEs)**.**

(a) Consider $\mathbb{G}\otimes\mathbb{H}$-progressive coefficients b and σ as in (2.34), and fix $\mathbb{Q}\in\widehat{\mathcal{P}}_p(W_n)$. Since sections of measurable maps are measurable,

$$\begin{aligned}
&\tilde{b}\colon \mathbb{R}_+\times W_n\to\mathbb{K}^n, && \tilde{b}(t,w):=b(t,w,\mathbb{Q}), \quad \text{and}\\
&\tilde{\sigma}\colon \mathbb{R}_+\times W_n\to\mathbb{K}^{n\times d}, && \tilde{\sigma}(t,w):=\sigma(t,w,\mathbb{Q})
\end{aligned}\tag{2.47}$$

are $\mathbb{G}$-progressive. Hence (2.44) can be viewed as a classical SDE of the form (C.6) with coefficients $\tilde{b}$ and $\tilde{\sigma}$.

(b) Consider $\mathbb{G}\otimes\mathbb{H}\otimes\mathbb{F}$-progressive coefficients b and σ as in (2.35) and a random probability measure $\mathbb{Q}\colon\Omega\to\widehat{\mathcal{P}}_p(W_n)$ which is measurable[11] w.r.t. $\mathcal{F}_t$-$\mathcal{H}_t$ for each $t\in\mathbb{R}_+$. Then

$$\begin{aligned}
&\tilde{b}\colon \mathbb{R}_+\times W_n\times\Omega\to\mathbb{K}^n, && \tilde{b}(t,w,\omega):=b(t,w,\mathbb{Q}(\omega),\omega), \quad \text{and}\\
&\tilde{\sigma}\colon \mathbb{R}_+\times W_n\times\Omega\to\mathbb{K}^{n\times d}, && \tilde{\sigma}(t,w,\omega):=\sigma(t,w,\mathbb{Q}(\omega),\omega),
\end{aligned}\tag{2.48}$$

are $\mathbb{G}\otimes\mathbb{F}$-progressive by Lemma 2.15(a). Hence (2.45) can be viewed as a classical SDE of the form (C.7) with coefficients $\tilde{b}$ and $\tilde{\sigma}$.

Remark 2.17 (Interpretation of $\mathbb{G}\otimes\mathbb{H}\otimes\mathbb{F}$-progressive measurability)**.** Let $(\Omega,\mathcal{F},\mathbb{F})$ with $\mathbb{F}=(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be a filtered measurable space. Progressive measurability of coefficients w.r.t. the filtration $\mathbb{G}\otimes\mathbb{H}\otimes\mathbb{F}$ means that the coefficients $b(s,w,\mathbb{Q},\omega)$ and $\sigma(s,w,\mathbb{Q},\omega)$ depend, for times s before a given time $t\in\mathbb{R}_+$, only on the "past" path $w\restriction_{[0,t]}$ and the "past" marginal distribution $\mathbb{Q}_{[0,t]}$. This notion can be made precise as follows: for each $t\in\mathbb{R}_+$, define

$$\begin{aligned}
\varphi_t\colon [0,t]\times W_n\times\widehat{\mathcal{P}}_p(W_n)\times\Omega &\to [0,t]\times W_n^t\times\mathcal{P}_p(W_n^t)\times\Omega,\\
(s,w,\mathbb{Q},\omega)&\mapsto (s,w\restriction_{[0,t]},\mathbb{Q}_{[0,t]},\omega).
\end{aligned}$$

[11] See Lemma D.17(b) for a class of random measures $\mathbb{Q}\colon\Omega\to\widehat{\mathcal{P}}_p(W_n)$ with the measurability property required in Remark 2.16(b).

Then a map $f\colon \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \times \Omega \to \mathbb{K}^m$ is $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$-progressive if and only if for every $t \in \mathbb{R}_+$, there exists a map $f_t\colon [0,t] \times W_n^t \times \mathcal{P}_p(W_n^t) \times \Omega \to \mathbb{K}^m$ which is $(\mathcal{B}_{[0,t]} \otimes \mathcal{B}_{W_n^t} \otimes \mathcal{B}_{\mathcal{P}_p(W_n^t)} \otimes \mathcal{F}_t)$-$\mathcal{B}_{\mathbb{K}^m}$ measurable, such that

$$f\restriction_{[0,t]\times W_n\times\widehat{\mathcal{P}}_p(W_n)\times\Omega} = f_t \circ \varphi_t.$$

This is a consequence of Doob's factorization lemma B.8 (a) and the identity

$$\mathcal{B}_{[0,t]} \otimes \mathcal{G}_t \otimes \mathcal{H}_t \otimes \mathcal{F}_t = \sigma(\varphi_t) = \varphi_t^{-1}(\mathcal{B}_{[0,t]} \otimes \mathcal{B}_{W_n^t} \otimes \mathcal{B}_{\mathcal{P}_p(W_n^t)} \otimes \mathcal{F}_t),$$

which follows from Lemma 2.1 (c) and Remark 2.9. The same arguments show that a map $f\colon \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n)$ is $\mathbb{G} \otimes \mathbb{H}$-progressive if and only if it admits a factorization as above, but where neither φ_t nor f_t depend on ω. For this reason, some sources (e.g. [13, 19]) write McKean SDEs (in the case of nonrandom dependence) in the form

$$\mathrm{d}X_t = b(t, X_{\cdot\wedge t}, \mathcal{L}[X_{\cdot\wedge t}])\,\mathrm{d}t + \sigma(t, X_{\cdot\wedge t}, \mathcal{L}[X_{\cdot\wedge t}])\,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+.$$

Remark 2.18 (Simplified measurability criterion for coefficients of the McKean SDE with nonrandom, instantaneous dependence)**.** In the case of nonrandom instantaneous dependence, i.e. if the coefficients are as in (2.42), the $\mathbb{G} \otimes \mathbb{H}$-progressive measurability required in Definition 2.12 is easily checked. Let $\hat{b}\colon \mathbb{R}_+ \times \mathbb{K}^n \times \mathcal{P}_p(\mathbb{K}^n) \to \mathbb{K}^n$ and $\hat{\sigma}\colon \mathbb{R}_+ \times \mathbb{K}^n \times \mathcal{P}_p(\mathbb{K}^n) \to \mathbb{K}^{n\times d}$ be measurable w.r.t. $\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{B}_{\mathcal{P}_p(\mathbb{K}^n)}$ and define

$$\begin{aligned} &b\colon \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \to \mathbb{K}^n, && b(s,w,\mathbb{Q}) = \hat{b}(s, w_s, \mathbb{Q}_s), \quad \text{and} \\ &\sigma\colon \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \to \mathbb{K}^{n\times d}, && \sigma(s,w,\mathbb{Q}) = \hat{\sigma}(s, w_s, \mathbb{Q}_s), \end{aligned}$$

see also (2.17). We show that b and σ are $\mathbb{G} \otimes \mathbb{H}$-progressive. Define

$$\varphi\colon \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \to \mathbb{R}_+ \times \mathbb{K}^n \times \mathcal{P}_p(\mathbb{K}^n), \quad (s,w,\mathbb{Q}) \mapsto (s, w_s, \mathbb{Q}_s),$$

with which we have

$$b = \hat{b} \circ \varphi \quad \text{and} \quad \sigma = \hat{\sigma} \circ \varphi.$$

Hence it suffices to show that for every $t \in \mathbb{R}_+$, the restriction $\varphi\restriction_{[0,t]\times W_n\times\widehat{\mathcal{P}}_p(W_n)}$ is $(\mathcal{B}_{[0,t]} \otimes \mathcal{G}_t \otimes \mathcal{H}_t)$-$(\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{B}_{\mathcal{P}_p(\mathbb{K}^n)})$-measurable. Indeed, measurability is clear for the first projection $(s,w,\mathbb{Q}) \mapsto s$ and follows for the second projection $(s,w,\mathbb{Q}) \mapsto w_s = \pi_s(w)$, since the canonical process $(\pi_s)_{s\in\mathbb{R}_+}$ on W_n is continuous, $\mathbb{G}$-adapted and thus $\mathbb{G}$-progressive. Finally, measurability of the third projection $[0,t] \times W_n \times \widehat{\mathcal{P}}_p(W_n) \ni (s,w,\mathbb{Q}) \mapsto \mathbb{Q}_s \in \mathcal{P}_p(\mathbb{K}^n)$ follows from the fact that the process $(\Pi_s)_{s\in\mathbb{R}_+}$ on $(\widehat{\mathcal{P}}_p(W_n), \mathcal{H})$ with Π_s defined in (2.16) is $\mathbb{H}$-progressive by Lemma 2.10.

The next two definitions specify the conditions on the coefficients under which there is strong existence and pathwise uniqueness[12] of solutions of (2.36), see Theorem 2.32 below. They are based on analogous definitions from [10, §1.3], made there for coefficients with instantaneous dependence (Definition 2.13 (a)), and incorporate also ideas from [53, §8.4].

[12] Strong existence and pathwise uniqueness of solutions of (2.37) hold also when the Lipschitz continuity property (see (2.49)) in the measure argument is relaxed from a global to a local form, if in addition a specific form of coefficients (see (3.55)) is assumed. This is the subject of Subsection 3.2.2 and has an application in the mean-field LIBOR market model described in Subsection 3.2.3.

Definition 2.19 (Lipschitz property)**.** Two coefficients b and σ as in (2.35) are said to have the *Lipschitz property*, if there exists an increasing function $c\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that for all $t \in \mathbb{R}_+$, $w, \tilde{w} \in W_n$, $\mathbb{Q}, \tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}_p(W_n)$ and all $\omega \in \Omega$,

$$\begin{aligned} \big|b(t,w,\mathbb{Q},\omega) - b(t,\tilde{w},\tilde{\mathbb{Q}},\omega)\big| + \big|\sigma(t,w,\mathbb{Q},\omega) - \sigma(t,\tilde{w},\tilde{\mathbb{Q}},\omega)\big| \\ \le c_t\big(p_t(w-\tilde{w}) + \widehat{\mathcal{W}}_{t,p}(\mathbb{Q},\tilde{\mathbb{Q}})\big), \end{aligned} \tag{2.49}$$

where $|\cdot|$ denotes both the Euclidean vector norm and the Frobenius matrix norm, p_t is given by (2.1) and $\widehat{\mathcal{W}}_{t,p}$ by (2.21).

Definition 2.20 (Linearly bounded growth condition)**.** Two coefficients b and σ as in (2.35) are said to satisfy the linearly bounded growth condition, if there exists an increasing function $D\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that for all $t \in \mathbb{R}_+$, $w \in W_n$, $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$ and all $\omega \in \Omega$,

$$|b(t,w,\mathbb{Q},\omega)| + |\sigma(t,w,\mathbb{Q},\omega)| \le D_t(1 + p_t(w)). \tag{2.50}$$

Lemma 2.21 (Weaker measurability requirements for coefficients with the Lipschitz property)**.**

(a) Let $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$ with $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be a filtered probability space and consider coefficients b and σ as in (2.35) which have the Lipschitz property[13] (Definition 2.19). If

$$\begin{aligned} &\mathbb{R}_+ \times \Omega \ni (t,\omega) \mapsto b(t,w,\mathbb{Q},\omega) \in \mathbb{K}^n \quad \textit{and} \\ &\mathbb{R}_+ \times \Omega \ni (t,\omega) \mapsto \sigma(t,w,\mathbb{Q},\omega) \in \mathbb{K}^{n\times d} \end{aligned}$$

are $\mathbb{F}$-progressive for all $(w,\mathbb{Q}) \in W_n \times \widehat{\mathcal{P}}_p(W_n)$, then b and σ are $\mathbb{G}\otimes\mathbb{H}\otimes\mathbb{F}$-progressive (Definition 2.11).

(b) Let b and σ be coefficients as in (2.34) which satisfy the Lipschitz property (Definition 2.19). If

$$t \mapsto b(t,w,\mathbb{Q}) \qquad \textit{and} \qquad t \mapsto \sigma(t,w,\mathbb{Q})$$

are $\mathcal{B}_{\mathbb{R}_+}$-measurable for all $(w,\mathbb{Q}) \in W_n \times \widehat{\mathcal{P}}_p(W_n)$, then b and σ are $\mathbb{G}\otimes\mathbb{H}$-progressive.

Proof. We can consider only b without loss of generality.

(a) The Lipschitz property implies that for all $(t,\omega) \in \mathbb{R}_+ \times \Omega$ and all $s \in [0,t]$, the restriction

$$W_n \times \widehat{\mathcal{P}}_p(W_n) \ni (w,\mathbb{Q}) \mapsto b\restriction_{[0,t]\times W_n\times\widehat{\mathcal{P}}_p(W_n)\times\Omega}(s,w,\mathbb{Q},\omega) \in \mathbb{K}^n$$

is continuous with respect to the product of the topology $\mathcal{O}_{p_t}$ on W_n and the topology $\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}$ induced by the pseudometric $\widehat{\mathcal{W}}_{t,p}$ on $\widehat{\mathcal{P}}_p(W_n)$ (Definition 2.5). With Lemma B.10, it follows that $b\restriction_{[0,t]\times W_n\times\widehat{\mathcal{P}}_p(W_n)\times\Omega}$ is measurable w.r.t. $\mathcal{B}_{[0,t]}\otimes\mathcal{B}\otimes\mathcal{F}_t$, where $\mathcal{B}$ denotes the Borel σ-algebra $\sigma\big(\mathcal{O}_{p_t}\otimes\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}\big)$ of the product of the two topologies. Since both

[13] More generally, it suffices for b and σ to be continuous w.r.t. the product of the topologies $\mathcal{O}_{p_t}$ and $\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}$ on W_n and $\widehat{\mathcal{P}}_p(W_n)$, respectively.

$\mathcal{O}_{p_t}$ and $\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}$ are second countable (the latter by Remark 2.6 (c)), we have by [53, Theorem 15.53]

$$\sigma\big(\mathcal{O}_{p_t} \otimes \mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}\big) = \sigma(\mathcal{O}_{p_t}) \otimes \sigma\big(\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}\big) = \mathcal{G}_t \otimes \mathcal{H}_t,$$

where we used Lemma 2.1 (c) and (2.33) in the last equality. Since $t \in \mathbb{R}_+$ was arbitrary, it follows that b is $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$-progressive.

(b) This follows directly from part (a). □

For McKean SDEs with instantaneous dependence (Definition 2.13), the following stronger Lipschitz property will be used in Theorem 2.58 below to obtain quantitative bounds on the speed of convergence of the particle approximation (Definition 2.51).

Definition 2.22 (Instantaneous Lipschitz property, [10, §1.3.1])**.** Two coefficients b and σ as in (2.35) are said to have the *instantaneous Lipschitz property*, if there exists an increasing function $c\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that for all $t \in \mathbb{R}_+$, $w, \tilde{w} \in W_n$, $\mathbb{Q}, \tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}_p(W_n)$ and all $\omega \in \Omega$,

$$\begin{aligned} \big|b(t, w, \mathbb{Q}, \omega) - b(t, \tilde{w}, \tilde{\mathbb{Q}}, \omega)\big| + \big|\sigma(t, w, \mathbb{Q}, \omega) - \sigma(t, \tilde{w}, \tilde{\mathbb{Q}}, \omega)\big| \\ \le c_t\big(|w_t - \tilde{w}_t| + d_p(\mathbb{Q}_t, \tilde{\mathbb{Q}}_t)\big), \end{aligned} \tag{2.51}$$

where d_p denotes the Wasserstein metric on $\mathcal{P}_p(\mathbb{K}^n)$ (Definition 2.4) and we have used the notation (2.17) for the one-dimensional marginal distributions of $\mathbb{Q}$ and $\tilde{\mathbb{Q}}$.

Remark 2.23 (Observations surrounding the Lipschitz properties)**.**

(a) (Lower-order Lipschitz properties are stronger) By Remark D.4, $\widehat{\mathcal{W}}_{t,q}$ is increasing in q for each $t \in \mathbb{R}_+$. In addition, $\widehat{\mathcal{P}}_q(W_n) \subseteq \widehat{\mathcal{P}}_p(W_n)$ for all $q \ge p$. Hence if b and σ satisfy the Lipschitz property (2.49) for an order $p \in [1, \infty)$, this is also the case for all orders $q \in [p, \infty)$. The same arguments show that the instantaneous Lipschitz property (2.51) for an order $p \in [1, \infty)$ implies the instantaneous Lipschitz property for all orders $q \in [p, \infty)$.

(b) (The instantaneous Lipschitz property implies the Lipschitz property) Assume that two coefficients b and σ as in (2.35) have the instantaneous Lipschitz property (Definition 2.22). Using (2.25) from Remark 2.6 (f), it follows that b and σ have the Lipschitz property (Definition 2.19).

Remark 2.24 (Strong existence and pathwise uniqueness of the SDE (2.45))**.** Fix $p \in [2, \infty)$ and an initial distribution $\mu \in \mathcal{P}_p(\mathbb{K}^n)$, let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic base (Definition C.5) and $\mathbb{Q}\colon \Omega \to \widehat{\mathcal{P}}_p(W_n)$ a random probability measure which is $\mathcal{F}_t$-$\mathcal{H}_t$-measurable for each $t \in \mathbb{R}_+$ (see Definition 2.8). Consider $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$-progressive (Definition 2.11) coefficients b and σ as in (2.35) and recall that by Remark 2.16 (b), the SDE (2.45) can be viewed as a classical SDE with $\mathbb{G} \otimes \mathbb{F}$-progressive coefficients $\tilde{b}$ and $\tilde{\sigma}$ defined in (2.48). Assume that b and σ have the Lipschitz property and satisfy the bounded growth condition (Definitions 2.19 and 2.20).

(a) The coefficients $\tilde{b}$ and $\tilde{\sigma}$ have the Lipschitz property and satisfy the bounded growth condition (Definitions C.25 and C.27) for classical SDEs. Hence by Theorem C.34, there is strong existence and pathwise uniqueness (Definitions C.19 and C.22) for the SDE (2.45) with initial distribution μ.

(b) By Remark C.35, every solution $X = (X_t)_{t\in\mathbb{R}_+}$ of (2.45) satisfies $\mathcal{L}[X] \in \widehat{\mathcal{P}}_p(W_n)$.

(c) In the case of nonrandom dependence, i.e. if b and σ are as in (2.34), then there is uniqueness in law of solutions by Corollary C.57.[14]

The concepts of strong solutions (Definition 2.25), strong existence (Definition 2.26), weak solutions and weak existence (Definition 2.27) and weak uniqueness (Definition 2.28) are defined in analogy to the classical case (Appendix C). We follow again Convention C.14, i.e. every definition made only for the McKean SDE (2.37) with random dependence also includes the more specific case (2.36) of nonrandom dependence. Different definitions for the cases of nonrandom and random dependence arise because of the stochastic base being fixed in the latter case, see also Remark C.12.

Definition 2.25 (Strong solution of the McKean SDE)**.** Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$ be a set-up (Definition C.5), and consider $\mathbb{G}\otimes\mathbb{H}\otimes\mathbb{F}$-progressive coefficients b and σ as in (2.35). A continuous $\mathbb{K}^n$-valued $\mathbb{F}$-adapted stochastic process X is called a *strong solution* of the McKean SDE (2.37) with initial condition ξ, if it is a strong solution in the classical sense (Definition C.15) for the SDE (2.45) with $\mathbb{Q} = \mathcal{L}[X]$.

Definition 2.26 (Strong existence of solutions of the McKean SDE)**.** Let $\mu \in \mathcal{P}_p(\mathbb{K}^n)$ be an initial distribution.

(a) For $\mathbb{G}\otimes\mathbb{H}$-progressive functions b and σ as in (2.34), we say that there is strong existence of a solution of the MSDE (2.36) with initial distribution μ for $\mathbb{R}^d$- or $\mathbb{K}^d$-valued Brownian motion, if for every set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B)$ (Definition C.5(b)) with $\mathcal{L}_{\mathbb{P}}[\xi] = \mu$ and where the Brownian motion B is $\mathbb{R}^d$- or $\mathbb{K}^d$-valued, respectively, there exists a strong solution in the sense of Definition 2.25.

(b) Fix a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. For $\mathbb{G}\otimes\mathbb{H}\otimes\mathbb{F}$-progressive functions as in (2.35), we say that there is strong existence of a solution of the MSDE (2.36) with initial distribution μ for $\mathbb{R}^d$- or $\mathbb{K}^d$-valued Brownian motion, if for every $\mathbb{R}^d$- respectively $\mathbb{K}^d$-valued $(\mathbb{F}, \mathbb{P})$-Brownian motion $B = (B_t)_{t\in\mathbb{R}_+}$ and every $\mathcal{F}_0$-measurable initial random vector ξ with $\mathcal{L}_{\mathbb{P}}[\xi] = \mu$, there exists a strong solution in the sense of Definition 2.25.

Definition 2.27 (Weak solution, weak existence of the McKean SDE)**.** Let $\mu \in \mathcal{P}_p(\mathbb{K}^n)$ be an initial distribution.

(a) Let b and σ be $\mathbb{G}\otimes\mathbb{H}$-progressive functions as in (2.34). A *weak solution* of the MSDE (2.36) with initial law μ for $\mathbb{R}^d$- or $\mathbb{K}^d$-valued Brownian motion is a pair $(\mathfrak{S}, X)$, where $\mathfrak{S} := (\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B)$ is a set-up (Definition C.5(b)) with $\mathcal{L}_{\mathbb{P}}[\xi] = \mu$ and where the

[14] By Example C.24, uniqueness in law cannot be expected to hold in general in the case of random dependence, even for classical SDEs.

Brownian motion B is $\mathbb{R}^d$-valued or $\mathbb{K}^d$-valued, respectively, and a stochastic process $X = (X_t)_{t\in\mathbb{R}_+}$ which is a strong solution (Definition 2.25) of (2.36) on the set-up $\mathfrak{S}$.[15]

(b) Fix a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and let b and σ be $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$-progressive coefficients as in (2.35). A *weak solution* of the MSDE (2.37) with initial law μ for $\mathbb{R}^d$- or $\mathbb{K}^d$-valued Brownian motion is a triple (B, ξ, X) of an $\mathbb{R}^d$-valued respectively $\mathbb{K}^d$-valued Brownian motion B, an $\mathcal{F}_0$-measurable initial random vector ξ with $\mathcal{L}_{\mathbb{P}}[\xi] = \mu$ and a continuous $\mathbb{K}^n$-valued $\mathbb{F}$-adapted stochastic process $X = (X_t)_{t\in\mathbb{R}_+}$ which is a strong solution (Definition 2.25) of (2.37) on the set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$.

(c) When a weak solution of (2.36) respectively (2.37) exists for the initial distribution μ, we speak of *weak existence* for this initial distribution.

Definition 2.28 (Weak uniqueness of solutions of the McKean SDE)**.** Let $\mu \in \mathcal{P}(\mathbb{K}^n)$ be an initial distribution. Consider $\mathbb{G} \otimes \mathbb{H}$-progressive coefficients b and σ as in (2.34). We say that there is *weak uniqueness* or *uniqueness in law* of solutions of the MSDE (2.36) for the initial distribution μ and $\mathbb{R}^d$- or $\mathbb{K}^d$-valued Brownian motion, respectively, if every weak solution X in the sense of Definition 2.27(a) with initial distribution μ has the same law on W_n.

Due to Remark C.23, Definition 2.28 does not include the case of random dependence.

Definition 2.29 (Pathwise uniqueness of solutions of the McKean SDE)**.** Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic base (Definition C.5(a)) and fix $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$-progressive coefficients b and σ as in (2.37) and an initial distribution $\mu \in \mathcal{P}_p(\mathbb{K}^n)$. We say that there is *pathwise uniqueness* for the SDE (2.37) with initial distribution μ for $\mathbb{R}^d$- or $\mathbb{K}^d$-valued Brownian motion, respectively, if whenever (B, X) and $(\tilde{B}, \tilde{X})$ are two weak solutions of (2.37) with initial distribution μ defined on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with $B \overset{\text{uti}}{=} \tilde{B}$ (u.t.i. identical Brownian motions) and $X_0 \overset{\text{a.s.}}{=} \tilde{X}_0$ (a.s. equal initial conditions), then $X \overset{\text{uti}}{=} \tilde{X}$.

Remark 2.30 (SDEs on finite and infinite time horizons)**.** In textbook presentations of the theory of SDEs (see e.g. [10, §1.3] [32, §1.5.4] and [44, §5.2]), as well as in papers on the subject (see e.g. [17]), SDEs are often considered on a finite time horizon $[0, \overline{T}]$. This means that instead of coefficients b and σ as in (2.35), the coefficients are of the form

$$b\colon [0, \overline{T}] \times W_n^{\overline{T}} \times \mathcal{P}_p(W_n^{\overline{T}}) \times \Omega \to \mathbb{K}^n \tag{2.52}$$

and

$$\sigma\colon [0, \overline{T}] \times W_n^{\overline{T}} \times \mathcal{P}_p(W_n^{\overline{T}}) \times \Omega \to \mathbb{K}^{n\times d}, \tag{2.53}$$

see (2.6) and (D.1). The entire theory on SDEs presented in Appendix C and the present chapter can also be developed for this setting. We briefly sketch the required modifications and then argue that the finite time horizon setting can, by appropriately extending the domain of the coefficients, effectively be viewed as a special case of the infinite time horizon setting, meaning that no re-development is in fact needed. Fix an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

[15] To simplify the notation, we will sometimes write (B, X) or even more simply X to denote a weak solution $((\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B), X)$, leaving the other elements of the set-up implicit. The same convention applies in the setting of random dependence (Definition 2.27(b)).

(a) The filtration $\mathbb{F}$ in $(\Omega, \mathcal{F})$ from Definition C.5(a) is replaced by a filtration $\mathbb{F}^{\overline{T}} = (\mathcal{F}_t^{\overline{T}})_{t\in[0,\overline{T}]}$ satisfying the usual completeness assumption (Definition C.3(a)).

(b) The σ-algebra $\mathcal{G}$ on W_n, see (2.4) is replaced by the σ-algebra $\mathcal{G}^{\overline{T}}$ generated by the projections $W_n^{\overline{T}} \ni w \mapsto \pi_s^{[0,\overline{T}]}(w) := w_s \in \mathbb{K}^n$, $s \in [0, \overline{T}]$. The filtration $\mathbb{G}$ in $(W_n, \mathcal{G})$, see (2.5), is replaced by $\mathbb{G}^{\overline{T}}$ with $\mathcal{G}_t^{\overline{T}} := \sigma(\pi_s^{[0,\overline{T}]} : s \in [0, t])$ for every $t \in [0, \overline{T}]$.

(c) The σ-algebra $\mathcal{H}$ on $\widehat{\mathcal{P}}_p(W_n)$, see (2.32), is replaced by $\mathcal{H}^{\overline{T}} := \sigma(\mathcal{O}_{\mathcal{W}_{\overline{T},p}})$, where $\mathcal{W}_{\overline{T},p}$ denotes the Wasserstein metric on $\mathcal{P}_p(W_n^{\overline{T}})$, see Definition 2.5 and Remark 2.6. Let $\mathcal{W}_{t,p}^{\overline{T}}(\mathbb{Q}, \tilde{\mathbb{Q}}) := \mathcal{W}_{t,p}(\mathbb{Q}\circ(\pi_{[0,t]}^{[0,\overline{T}]})^{-1}, \tilde{\mathbb{Q}}\circ(\pi_{[0,t]}^{[0,\overline{T}]})^{-1})$ for all $t \in [0, \overline{T}]$ and all $\mathbb{Q}, \tilde{\mathbb{Q}} \in \mathcal{P}_p(W_n^{\overline{T}})$, cf. (2.9) and (2.21). Then the filtration $\mathbb{H}$ in $(\widehat{\mathcal{P}}_p(W_n), \mathcal{H})$, see (2.33), is replaced by the filtration $\mathbb{H}^{\overline{T}} = (\mathcal{H}_t^{\overline{T}})_{t\in[0,\overline{T}]}$, where $\mathcal{H}_t^{\overline{T}} = \sigma(\mathcal{O}_{\mathcal{W}_{t,p}^{\overline{T}}})$ for each $t \in [0, \overline{T}]$.

With the modifications from parts (a) to (c) above, the entire theory of classical and McKean SDEs presented in Appendix C and the present chapter could be re-derived for $\mathbb{G}^{\overline{T}} \otimes \mathbb{F}^{\overline{T}}$-progressively respectively $\mathbb{G}^{\overline{T}} \otimes \mathbb{H}^{\overline{T}} \otimes \mathbb{F}^{\overline{T}}$-progressively measurable coefficients (cf. Definitions C.6 and 2.11) with the required modifications being either notational or simplifying.[16] However, such a re-derivation is unnecessary, since every SDE on a finite time horizon with $\mathbb{G}^{\overline{T}} \otimes \mathbb{H}^{\overline{T}} \otimes \mathbb{F}^{\overline{T}}$-progressively measurable coefficients as in (2.52) and (2.53) can be viewed as an instance of the general case (2.37) by the following construction. Given the filtration $\mathbb{F}^{\overline{T}} = (\mathcal{F}_t^{\overline{T}})_{t\in[0,\overline{T}]}$, define an extended filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ by $\mathcal{F}_t = \mathcal{F}_{t\wedge\overline{T}}^{\overline{T}}$ for every $t \in \mathbb{R}_+$. Furthermore, let

$$
\begin{aligned}
\varphi\colon \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \times \Omega &\to [0, \overline{T}] \times W_n^{\overline{T}} \times \mathcal{P}_p(W_n^{\overline{T}}) \times \Omega, \\
(s, w, \mathbb{Q}, \omega) &\mapsto (s \wedge \overline{T}, w\restriction_{[0,\overline{T}]}, \mathbb{Q}_{[0,\overline{T}]}, \omega),
\end{aligned}
\tag{2.54}
$$

and

$$
\tilde{b} := \mathbb{1}_{[0,\overline{T}]}(b \circ \varphi), \qquad \tilde{\sigma} := \mathbb{1}_{[0,\overline{T}]}(\sigma \circ \varphi). \tag{2.55}
$$

Routine arguments[17] establish that $\varphi\restriction_{[0,t]\times W_n\times\widehat{\mathcal{P}}_p(W_n)\times\Omega}$ is, for each $t \in \mathbb{R}_+$, $(\mathcal{B}_{[0,t]} \otimes \mathcal{G}_t \otimes \mathcal{H}_t \otimes \mathcal{F}_t)$-$(\mathcal{B}_{[0,t\wedge\overline{T}]} \otimes \mathcal{G}_{t\wedge\overline{T}}^{\overline{T}} \otimes \mathcal{H}_{t\wedge\overline{T}}^{\overline{T}} \otimes \mathcal{F}_{t\wedge\overline{T}}^{\overline{T}})$-measurable. Coupled with the assumed $\mathbb{G}^{\overline{T}} \otimes \mathbb{H}^{\overline{T}} \otimes \mathbb{F}^{\overline{T}}$-progressive measurability of b and σ, it follows that $\tilde{b}$ and $\tilde{\sigma}$ defined in (2.55) are $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$-progressive. Similarly, if b and σ satisfy a Lipschitz property and bounded growth condition analogous to (2.49) and (2.50) but adapted to the finite time horizon setting, then $\tilde{b}$ and $\tilde{\sigma}$ defined in (2.55) have the Lipschitz property and satisfy the bounded growth condition for McKean SDEs (Definitions 2.19 and 2.20), so our main existence and uniqueness result, Theorem 2.32, applies.[18]

[16] For example, in the Lipschitz property and bounded growth condition (Definitions 2.19 and 2.20), constants can replace the functions c and D. Moreover, due to $(W_n^{\overline{T}}, \mathcal{W}_{\overline{T},p})$ being a Polish space by Theorem D.8, the use of Proposition 2.7 in the proof of Theorems 2.32 and 3.19, and with it all of Appendix A, becomes superfluous.

[17] See e.g. the proof of Lemma 2.15.

[18] Similar arguments work under (the finite time horizon analogues of) SDEs with coefficients as in Assumption 3.14.

For this reason, we will refer to results developed in Appendix C and this section also when working with SDEs with coefficients defined on finite time horizons, which are more common in the applied literature. This corresponds to implicitly extending coefficients as in (2.52) and (2.53) using (2.55).

For the proof of Theorem 2.32 below, but also when investigating the propagation of chaos in Section 2.4, Lemma 2.31, which generalizes Lemma C.31 to the setting of McKean SDEs, will be useful. The proof is based on ideas from Step 1 of the proof of [53, Theorem 8.27], adapted incorporating also parts of the proof of [33, Proposition 2] to take into account the added dependencies of the coefficients.

Lemma 2.31 (L^p-continuous dependence of certain stochastic integrals on the input data)**.** *Fix $p \in [2, \infty)$, a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ a filtration and an $\mathbb{R}^d$- or $\mathbb{K}^d$-valued standard $(\mathbb{F}, \mathbb{P})$ Brownian motion $B = (B_t)_{t\in\mathbb{R}_+}$. Consider $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$-progressive coefficients b and σ as in (2.35) which*

(i) have the Lipschitz property (Definition 2.19), or
(ii) have the instantaneous Lipschitz property (Definition 2.22).

(a) Let $\mathbb{Q}^X, \mathbb{Q}^Y \colon \Omega \to \widehat{\mathcal{P}}_p(W_n)$ be maps which are $\mathcal{F}_t$-$\mathcal{H}_t$-measurable for each $t \in \mathbb{R}_+$ (see Definition 2.8) and define

$$\gamma_s(\omega) := \begin{cases} \widehat{\mathcal{W}}_{s,p}(\mathbb{Q}^X(\omega), \mathbb{Q}^Y(\omega)) & \text{in case (i),} \\ d_p(\mathbb{Q}^X_s(\omega), \mathbb{Q}^Y_s(\omega)) & \text{in case (ii),} \end{cases} \tag{2.56}$$

for all $s \in \mathbb{R}_+$ and $\omega \in \Omega$, where we have used the notation (2.17) for the second case. Let X and Y be two continuous semimartingales such that the processes

$$\begin{aligned} \tilde{X} &:= X_0 + \int_0^{\cdot} b(s, X, \mathbb{Q}^X, \cdot\,)\,\mathrm{d}s + \int_0^{\cdot} \sigma(s, X, \mathbb{Q}^X, \cdot\,)\,\mathrm{d}B_s \quad \text{and} \\ \tilde{Y} &:= Y_0 + \int_0^{\cdot} b(s, Y, \mathbb{Q}^Y, \cdot\,)\,\mathrm{d}s + \int_0^{\cdot} \sigma(s, Y, \mathbb{Q}^Y, \cdot\,)\,\mathrm{d}B_s \end{aligned} \tag{2.57}$$

are well-defined.[19] Then for all $t \in \mathbb{R}_+$,

$$\begin{aligned} \mathbb{E}[p_t^p(\tilde{X} - \tilde{Y})] &\le 3^{p-1}\mathbb{E}[|X_0 - Y_0|^p] \\ &\quad + \beta(t) \int_0^t c_s^p \Big(\mathbb{E}[p_s^p(X - Y)] + \mathbb{E}\big[\gamma_s^p\big]\Big)\,\mathrm{d}s\,, \end{aligned} \tag{2.58}$$

where the function c originates from the Lipschitz property and β was defined in (C.22).

(b) Let $X = (X_t)_{t\in\mathbb{R}_+}$ and $Y = (Y_t)_{t\in\mathbb{R}_+}$ be two strong solutions (Definition 2.25) of the McKean SDE (2.37).[20] Then for all $t \in \mathbb{R}_+$,

$$\mathbb{E}[p_t^p(X - Y)] \le \left(3^{p-1}\mathbb{E}[|X_0 - Y_0|^p] + \beta(t) \int_0^t c_s^p\, \gamma_s^p\,\mathrm{d}s\right) \exp\big(t(\beta(t) \vee c_t^p)^2\big)\,, \tag{2.59}$$

with γ as in (2.56), setting[21] $\mathbb{Q}^X := \mathcal{L}[X]$ and $\mathbb{Q}^Y := \mathcal{L}[Y]$.

[19] By Lemma 2.15 (b), the coefficients in (2.57) are $\mathbb{F}$-progressive, so only stochastic integrability (Definition C.1) is in question.

[20] The solutions X and Y may have different starting values X_0 and Y_0.

[21] In the setting of Lemma 2.31 (b), γ is a deterministic function, since $\mathbb{Q}^X$ and $\mathbb{Q}^Y$ are constant.

Proof. (a) For each $N \in \mathbb{N}$ define $\tau_N = \inf\{t \in \mathbb{R}_+ : |X_t - Y_t| \geq N\}$, which is a stopping time by [53, Lemma 3.53(b)]. Since X and Y have continuous paths, it follows for every $\omega \in \Omega$ that the increasing sequence $(\tau_N(\omega))_{N\in\mathbb{N}}$ cannot have an accumulation point in any compact subset of $\mathbb{R}_+$, hence $\tau_N(\omega) \to \infty$ as $N \to \infty$.[22] Fix $N \in \mathbb{N}$ for the remaining part of this step. We will omit this index where appropriate. Define the processes $U \colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^n$ and $V \colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^{n\times d}$ by

$$\begin{aligned} U_s(\omega) &= \mathbb{1}_{[0,\tau_N(\omega)]}(s)\big(b(s, X(\,\cdot\,,\omega), \mathbb{Q}^X(\omega), \omega) - b(s, Y(\,\cdot\,,\omega), \mathbb{Q}^Y(\omega), \omega)\big), \\ V_s(\omega) &= \mathbb{1}_{[0,\tau_N(\omega)]}(s)\big(\sigma(s, X(\,\cdot\,,\omega), \mathbb{Q}^X(\omega), \omega) - \sigma(s, Y(\,\cdot\,,\omega), \mathbb{Q}^Y(\omega), \omega)\big), \end{aligned} \tag{2.60}$$

for all $s \in \mathbb{R}_+$ and $\omega \in \Omega$. To see that the processes U and V are progressive, note that $\mathbb{1}_{[0,\tau_N]}$ is adapted and left-continuous, hence progressive, that the other terms in (2.60) are progressive by Lemma 2.15(b). By the Lipschitz property (2.49) respectively the instantaneous Lipschitz property (2.51) of the coefficients, these processes satisfy, for every $s \in \mathbb{R}_+$ and $\omega \in \Omega$,

$$\begin{aligned} |U_s(\omega)| + |V_s(\omega)| &\leq c_s\, \mathbb{1}_{[0,\tau_N(\omega)]}(s)\, \Big(p_s(X(\,\cdot\,,\omega) - Y(\,\cdot\,,\omega)) + \gamma_s(\omega)\Big) \\ &\leq c_s \Big(p_s\left(X^{\tau_N}(\,\cdot\,,\omega) - Y^{\tau_N}(\,\cdot\,,\omega)\right) + \gamma_s(\omega)\Big). \end{aligned}$$

With subadditivity of $\mathbb{R}_+ \ni x \mapsto x^{1/p}$ and Jensen's inequality for $\mathbb{R}_+ \ni x \mapsto x^p$ (see (C.18)), the previous display implies

$$\begin{aligned} |U_s(\omega)|^p + |V_s(\omega)|^p &\leq \big(|U_s(\omega)| + |V_s(\omega)|\big)^p \\ &\leq 2^{p-1} c_s^p \Big(p_s^p\left(X^{\tau_N}(\,\cdot\,,\omega) - Y^{\tau_N}(\,\cdot\,,\omega)\right) + \gamma_s^p(\omega)\Big). \end{aligned} \tag{2.61}$$

We will omit writing the argument ω in the following. Define

$$\tilde{u}_N(t) = \mathbb{E}[p_t^p(\tilde{X}^{\tau_N} - \tilde{Y}^{\tau_N})], \qquad t \in \mathbb{R}_+. \tag{2.62}$$

Using the defining equation (2.57) for $\tilde{X}$ and $\tilde{Y}$, linearity and commutation with stopping of the stochastic integral, (2.60), the triangle inequality for the seminorm p_t given by (2.1) and finally Jensen's inequality (see (C.18)), it follows that

$$\tilde{u}_N(t) \leq 3^{p-1}\mathbb{E}\bigg[|X_0 - Y_0|^p + p_t^p\bigg(\int_0^{\cdot} U_s \,\mathrm{d}s\bigg) + p_t^p(V \cdot B)\bigg], \quad t \in \mathbb{R}_+. \tag{2.63}$$

For each $t \in \mathbb{R}_+$, rewriting with the definition (2.1) of p_t and using Jensen's inequality (see (C.19)) for $t > 0$,

$$p_t^p\bigg(\int_0^{\cdot} U_s \,\mathrm{d}s\bigg) = \sup_{u\in[0,t]}\bigg|\int_0^u U_s \,\mathrm{d}s\bigg|^p \leq \bigg(\int_0^t |U_s|\mathrm{d}s\bigg)^p \leq t^{p-1}\int_0^t |U_s|^p \,\mathrm{d}s. \tag{2.64}$$

[22] The stopping times $(\tau_N)_{N\in\mathbb{N}}$ are needed for the application of Grönwall's inequality in part 24 of the proof to ensure a local integrability assumption.

Since $V \cdot B$ is a continuous local martingale with quadratic variation process $\operatorname{tr}([V \cdot B]) \overset{\text{uti}}{=} c_{\mathbb{K}} \int_0^{\cdot} |V_s|^2 \, \mathrm{d}s$, it follows from the Burkholder–Davis–Gundy inequality [53, Theorem 6.32] that

$$\begin{aligned}\mathbb{E}\big[p_t^p(V \cdot B)\big] &\leq C_p c_{\mathbb{K}}^{p/2} \mathbb{E}\bigg[\bigg(\int_0^t |V_s|^2 \, \mathrm{d}s\bigg)^{p/2}\bigg] \\ &\leq C_p c_{\mathbb{K}}^{p/2} t^{p/2-1} \mathbb{E}\bigg[\int_0^t |V_s|^p \, \mathrm{d}s\bigg],\end{aligned} \tag{2.65}$$

where for the last inequality we used Jensen's inequality (see (C.19)) with the function $\mathbb{R}_+ \ni x \mapsto x^{p/2}$, which is convex since $p \geq 2$ by assumption. Plugging the estimates (2.64) and (2.65) into (2.63) and using Fubini's theorem for the first inequality, and applying (2.61) for the third inequality shows that, for each $t \in \mathbb{R}_+$,

$$\begin{aligned}\tilde{u}_N(t) &\leq 3^{p-1}\mathbb{E}[|X_0 - Y_0|^p] + 3^{p-1} \int_0^t \mathbb{E}\left[t^{p-1}|U_s|^p + C_p c_{\mathbb{K}}^{p/2} t^{p/2-1}|V_s|^p\right] \mathrm{d}s \\ &\leq 3^{p-1}\mathbb{E}[|X_0 - Y_0|^p] + \underbrace{3^{p-1}\left(t^{p-1} + C_p c_{\mathbb{K}}^{p/2} t^{p/2-1}\right)}_{=\beta(t)/2^{p-1}} \int_0^t \mathbb{E}\left[|U_s|^p + |V_s|^p\right] \mathrm{d}s \\ &\leq 3^{p-1}\mathbb{E}[|X_0 - Y_0|^p] + \beta(t) \int_0^t c_s^p \Big(u_N(s) + \mathbb{E}[\gamma_s^p]\Big) \, \mathrm{d}s\end{aligned} \tag{2.66}$$

with

$$u_N(s) := \mathbb{E}\big[p_s^p\left(X^{\tau_N} - Y^{\tau_N}\right)\big], \quad s \in \mathbb{R}_+. \tag{2.67}$$

The expectation $\mathbb{E}[\gamma_s^p]$ in the last line of (2.66) exists due to Lemma D.19 (a) and is a measurable function[23] of s by Lemma D.19 (b), which ensures well-definedness of the Lebesgue integral $\int_0^t c_s^p \, \mathbb{E}[\gamma_s^p] \, \mathrm{d}s$. By the monotone convergence theorem applied to the expectation, (2.67) implies

$$u_N(s) \xrightarrow[N\to\infty]{} \mathbb{E}[p_s^p(X - Y)],$$

and similarly it follows from (2.62) that

$$\tilde{u}_N(t) \xrightarrow[N\to\infty]{} \mathbb{E}[p_t^p(\tilde{X} - \tilde{Y})],$$

for all $s \in [0, t]$. With another application of monotone convergence, this time for the Lebesgue integral over $[0, t]$ in the last line of (2.66), the claimed bound (2.58) follows from (2.66).

(b) If X and Y are two strong solutions (Definition 2.25) of the McKean SDE (2.37), then the continuous semimartingales $\tilde{X}$ and $\tilde{Y}$ from (2.57) with $\mathbb{Q}^X = \mathcal{L}[X]$ and $\mathbb{Q}^Y = \mathcal{L}[Y]$ are well defined by Definition 2.25 of a strong solution and coincide up to indistinguishability with X and Y, respectively. Hence the functions $\tilde{u}_N$ and u_N defined in the proof of part 13 are equal. By definition of the stopping times $(\tau_N)_{N\in\mathbb{N}}$, we have $u_N(s) \leq \mathbb{E}[|X_0 - Y_0|^p] + N^p$. For proving the bound (2.59), we can assume

[23] It is possible for γ_s^p to have infinite expectation.

w.l.o.g. that $\mathbb{E}[|X_0 - Y_0|^p] < \infty$, hence u_N is bounded for every $N \in \mathbb{N}$ and thus locally integrable w.r.t. the measure μ on $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$, which we define by its density c^p w.r.t. the Lebesgue measure. The bound (2.66) shows that u_N satisfies the integral inequality needed for the application of the extended Grönwall inequality [53, Theorem 8.21] with $I = \mathbb{R}_+$,

$$\alpha(t) := 3^{p-1}\mathbb{E}[|X_0 - Y_0|^p] + \beta(t)\int_0^t c_s^p\,\gamma_s^p\,\mathrm{d}s\,, \qquad t \in \mathbb{R}_+$$

and the function β given by (C.22). Since α is increasing and both β and the density c^p of μ are bounded on $[0,t]$ by $\beta(t) \vee c_t^p$, we have by [53, Remark 8.23 (c)]

$$u_N(t) \le \left(3^{p-1}\mathbb{E}[|X_0 - Y_0|^p] + \beta(t)\int_0^t c_s^p\,\gamma_s^p\,\mathrm{d}s\right)\exp\left(t(\beta(t) \vee c_t^p)^2\right), \tag{2.68}$$

for each $t \in \mathbb{R}_+$. Since $p_t\left(X^{\tau_N} - Y^{\tau_N}\right) \nearrow p_t(X - Y)$ pointwise on Ω as $N \to \infty$, the monotone convergence theorem implies that $u_N(t) \xrightarrow[N\to\infty]{} \mathbb{E}\left[p_t^p(X - Y)\right]$. Combining this with (2.68), we obtain (2.59). $\square$

Theorem 2.32 below, one of the main results in this thesis, establishes strong existence and pathwise uniqueness of solutions of the McKean SDE (2.37) and gives in addition weak uniqueness for the case of nonrandom dependence. The statement of Theorem 2.32 is based on that of the existence and uniqueness result [53, Theorem 8.27] for classical SDEs, while the proof builds on ideas from [26, §10.2] and [33, Proposition 2], where the case of nonrandom, instantaneous dependence is treated on a finite time horizon, as well as [10, §1.3.1], which includes the case of random dependence. For the extension to the infinite time horizon, Proposition 2.7 and Theorem A.16 are used.[24]

Theorem 2.32 (Strong Existence and Pathwise Uniqueness for the McKean SDE)**.** *Fix $p \in [2,\infty)$ and a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ (Definition C.5 (a)). Consider $\mathbb{G}\otimes\mathbb{H}\otimes\mathbb{F}$-progressive coefficients b and σ as in* (2.35) *which*

(a) have the Lipschitz property (Definition 2.19), and

(b) are of linearly bounded growth (Definition 2.20).

Then, for every initial distribution $\mu \in \mathcal{P}_p(\mathbb{K}^n)$ and every $\mathbb{R}^d$- or $\mathbb{K}^d$-valued standard Brownian motion B, there is strong existence and pathwise uniqueness (Definitions 2.26 and 2.29) for the McKean SDE (2.37)*. For each $t \in \mathbb{R}_+$, the seminorm p_t (see* (2.1)*) of the solution process X is p-integrable.*

Furthermore, in the case of nonrandom dependence (Definition 2.12), there is uniqueness in law (Definition 2.28), and given an $\mathcal{F}_0$-measurable initial random vector ξ with law μ, the solution X with $X_0 \overset{\text{a.s.}}{=} \xi$ is progressive w.r.t. the smallest filtration $\widetilde{\mathbb{F}} = (\widetilde{\mathcal{F}}_t)_{t\in\mathbb{R}_+}$ generated by ξ and B and containing all $\mathbb{P}$-null sets of $\sigma(\xi, (B_t)_{t\in\mathbb{R}_+})$.

[24] Theorem 2.32 is extended by Theorem 3.19 in Subsection 3.2.2, where the Lipschitz continuity property (see (2.49)) in the measure argument is relaxed from a global to a local form, assuming in addition a specific form of coefficients (see (3.55)).

Proof. The proof is divided into two main steps. In Step 1, we reduce to an equivalent fixed point problem. In Step 2, we use a version of Banach's fixed point theorem (Theorem A.16) for topologies induced by countable families of pseudometrics[25] to show existence and uniqueness of a fixed point.

Step 1 (Reduction to an equivalent fixed point problem). Fix a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B)$ (Definition C.5(b)) with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ and where the initial random vector ξ has law μ. Every strong solution constructed in this proof will be w.r.t. this set-up. From parts (a) and (b) of Remark 2.24, it follows that for every $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$, the classical SDE (2.45) has a pathwise unique strong solution $X^{\mathbb{Q}}$ with $X_0^{\mathbb{Q}} \overset{\text{a.s.}}{=} \xi$ and $\mathcal{L}[X^{\mathbb{Q}}] \in \widehat{\mathcal{P}}_p(W_n)$. Hence the map

$$\Phi\colon \widehat{\mathcal{P}}_p(W_n) \to \widehat{\mathcal{P}}_p(W_n), \quad \mathbb{Q} \mapsto \mathcal{L}[X^{\mathbb{Q}}] \tag{2.69}$$

is well defined. We make the following three observations:

(i) If X is a strong solution of the McKean SDE (2.37) with $X_0 \overset{\text{a.s.}}{=} \xi$, then $\mathcal{L}[X]$ is a fixed point of Φ.

(ii) If $\mathbb{Q}^* \in \widehat{\mathcal{P}}_p(W_n)$ is a fixed point of Φ, then $X^{\mathbb{Q}^*}$ is a strong solution of the McKean SDE on the given set-up which is pathwise unique among the solutions with law $\mathbb{Q}^*$.

(iii) In the case of nonrandom dependence, for each $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$, $X^{\mathbb{Q}}$ is a solution of the SDE (2.44). By Remark 2.24(c) there is uniqueness in law for this SDE. Hence in this case, Φ is independent of the underlying set-up.

By observation (ii), existence of a fixed point of Φ implies strong existence of solutions. In combination with (i), Φ having a unique fixed point implies pathwise uniqueness of solutions. Moreover, in the case of nonrandom dependence, using observation (iii), uniqueness of fixed points of Φ implies weak uniqueness of solutions of the McKean SDE. Hence it suffices to show that Φ has a unique fixed point $\mathbb{Q}^*$ in $\widehat{\mathcal{P}}_p(W_n)$. Assuming that this is the case, the McKean SDE (2.37) with initial distribution μ is equivalent to the classical SDE

$$\mathrm{d}X_t = b(t, X, \mathbb{Q}^*, \,\cdot\,)\,\mathrm{d}t + \sigma(t, X, \mathbb{Q}^*, \,\cdot\,)\,\mathrm{d}B_t\,, \qquad \mathcal{L}[X_0] = \mu. \tag{2.70}$$

The p-integrability of $p_t(X)$ for all $t \in \mathbb{R}_+$ and the $\widetilde{\mathbb{F}}$-progressive measurability of X in the case of nonrandom dependence then follow directly from the classical strong existence and pathwise uniqueness result (Theorem C.34) applied to (2.70).

Step 2 (Existence and uniqueness of a fixed point of Φ). By Proposition 2.7, the family $(\widehat{\mathcal{W}}_{k,p})_{k\in\mathbb{N}}$ from Definition 2.5 is a separating family of complete pseudometrics on $\widehat{\mathcal{P}}_p(W_n)$, and the associated metric $\widehat{\mathcal{W}}_p$ defined by (2.26) induces a Polish topology on $\widehat{\mathcal{P}}_p(W_n)$. By Theorem A.16, existence and uniqueness of a fixed point of Φ follow if

(i) for each $k \in \mathbb{N}$, Φ is continuous w.r.t. $\widehat{\mathcal{W}}_{k,p}$, and

(ii) for each $k \in \mathbb{N}$, there exists an $m \in \mathbb{N}$ such that the m-fold composition $\Phi^{\circ m}$ of Φ with itself is a contraction w.r.t. $\widehat{\mathcal{W}}_{k,p}$.

Fix $\mathbb{Q}, \tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}_p(W_n)$ for now and define

$$h(t) := \mathbb{E}[p_t^p(X^{\mathbb{Q}} - X^{\tilde{\mathbb{Q}}})], \qquad t \in \mathbb{R}_+, \tag{2.71}$$

[25] See Appendix A for background material on pseudometric spaces.

which is an increasing, $\mathbb{R}_+$-valued function due to the fact that $\mathcal{L}[X^{\mathbb{Q}}]$ and $\mathcal{L}[X^{\tilde{\mathbb{Q}}}]$ are in $\widehat{\mathcal{P}}_p(W_n)$ by construction of the map Φ. Fix $t \in \mathbb{R}_+$ for a while. By construction, $X^{\mathbb{Q}} \overset{\text{uti}}{=} \xi + \int_0^{\cdot} b(s, X^{\mathbb{Q}}, \mathbb{Q}, \,\cdot\,)\,\mathrm{d}s + \int_0^{\cdot} \sigma(s, X^{\mathbb{Q}}, \mathbb{Q}, \,\cdot\,)\,\mathrm{d}B_s$ and $X^{\tilde{\mathbb{Q}}} \overset{\text{uti}}{=} \xi + \int_0^{\cdot} b(s, X^{\tilde{\mathbb{Q}}}, \tilde{\mathbb{Q}}, \,\cdot\,)\,\mathrm{d}s + \int_0^{\cdot} \sigma(s, X^{\tilde{\mathbb{Q}}}, \tilde{\mathbb{Q}}, \,\cdot\,)\,\mathrm{d}B_s$. Hence (2.58) from Lemma 2.31 (a) with $\tilde{X} := X^{\mathbb{Q}}$ and $\tilde{Y} := X^{\tilde{\mathbb{Q}}}$ implies that

$$h(t) \le \beta(t) \int_0^t c_s^p\, h(s)\,\mathrm{d}s + \underbrace{\beta(t) c_t^p \int_0^t \widehat{\mathcal{W}}_{s,p}^p(\mathbb{Q}, \tilde{\mathbb{Q}})\,\mathrm{d}s}_{=:\tilde{\alpha}(t)},$$

where the function c originates from the Lipschitz property (Definition 2.19) and β was defined in (C.22). By Grönwall's inequality ([53, Theorem 8.21 and Remark 8.24 (c)]), using also that $\widehat{\mathcal{W}}_{s,p}^p(\mathbb{Q}, \tilde{\mathbb{Q}}) = \widehat{\mathcal{T}}_{s,p}(\mathbb{Q}, \tilde{\mathbb{Q}})$ by (2.20) and (2.21),

$$h(t) \le \tilde{\alpha}(t) \exp\bigl(t(\beta(t) \vee c_t^p)^2\bigr) = L_t \int_0^t \widehat{\mathcal{T}}_{s,p}(\mathbb{Q}, \tilde{\mathbb{Q}})\,\mathrm{d}s\,, \tag{2.72}$$

where[26]

$$L_t := \beta(t) c_t^p \exp\bigl(t(\beta(t) \vee c_t^p)^2\bigr), \qquad t \in \mathbb{R}_+, \tag{2.73}$$

is an $\mathbb{R}_+$-valued increasing function. Recall the notation $\mathcal{P}(\mu, \nu)$ for the set of couplings of two probability measures μ and ν (Definition D.2). Since $\mathcal{L}_{\mathbb{P}}[(X^{\mathbb{Q}}, X^{\tilde{\mathbb{Q}}})] \in \mathcal{P}(\Phi(\mathbb{Q}), \Phi(\tilde{\mathbb{Q}}))$, we have $\mathcal{L}_{\mathbb{P}}[((X_s^{\mathbb{Q}})_{s\in[0,t]}, (X^{\tilde{\mathbb{Q}}_s})_{s\in[0,t]})] \in \mathcal{P}(\Phi(\mathbb{Q})_{[0,t]}, \Phi(\tilde{\mathbb{Q}})_{[0,t]})$. Consequently, by (2.21) from Definition 2.5,

$$\widehat{\mathcal{T}}_{t,p}(\Phi(\mathbb{Q}), \Phi(\tilde{\mathbb{Q}})) \le \mathbb{E}\Bigl[p_t^p(X^{\mathbb{Q}} - X^{\tilde{\mathbb{Q}}})\Bigr] = h(t). \tag{2.74}$$

Combining (2.72) and (2.74), which both hold for all $t \in \mathbb{R}_+$, we have

$$\widehat{\mathcal{T}}_{t,p}(\Phi(\mathbb{Q}), \Phi(\tilde{\mathbb{Q}})) \le L_t \int_0^t \widehat{\mathcal{T}}_{s,p}(\mathbb{Q}, \tilde{\mathbb{Q}})\,\mathrm{d}s\,, \qquad t \in \mathbb{R}_+. \tag{2.75}$$

We next show by induction that (2.75) implies that

$$\widehat{\mathcal{T}}_{t,p}(\Phi^{\circ N}(\mathbb{Q}), \Phi^{\circ N}(\tilde{\mathbb{Q}})) \le \frac{(L_t \cdot t)^N}{N!} \widehat{\mathcal{T}}_{t,p}(\mathbb{Q}, \tilde{\mathbb{Q}}), \qquad t \in \mathbb{R}_+, N \in \mathbb{N}. \tag{2.76}$$

Indeed, the case $N = 1$ follows from (2.75) and the fact that $\widehat{\mathcal{T}}_{s,p} \le \widehat{\mathcal{T}}_{t,p}$ whenever $0 \le s \le t$, see Remark 2.6 (e). Furthermore, if (2.76) holds for $N \in \mathbb{N}$, then using (2.75) and the inductive hypothesis,

$$\begin{aligned}
\widehat{\mathcal{T}}_{t,p}(\Phi^{\circ(N+1)}(\mathbb{Q}),\Phi^{\circ(N+1)}(\tilde{\mathbb{Q}})) &\le L_t \int_0^t \widehat{\mathcal{T}}_{s,p}(\Phi^{\circ N}(\mathbb{Q}), \Phi^{\circ N}(\tilde{\mathbb{Q}}))\,\mathrm{d}s \\
&\le L_t \int_0^t \frac{(L_s \cdot s)^N}{N!} \widehat{\mathcal{T}}_{s,p}(\mathbb{Q}, \tilde{\mathbb{Q}})\,\mathrm{d}s \le \frac{(L_t \cdot t)^{N+1}}{(N+1)!} \widehat{\mathcal{T}}_{t,p}(\mathbb{Q}, \tilde{\mathbb{Q}}).
\end{aligned}$$

Since $\mathbb{Q}, \tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}_p(W_n)$ were arbitrary, (2.76) for $N = 1$ and (2.19) imply that for each $k \in \mathbb{N}$, Φ is continuous w.r.t. $\widehat{\mathcal{W}}_{k,p}$. In addition, using (2.76) together with the fact that $\widehat{\mathcal{W}}_{t,p} = \widehat{\mathcal{T}}_{t,p}^{1/p}$

[26] From the definition (C.22) of β, we see that L_t depends apart from t only on the order p and the Lipschitz constant c_t (Definition 2.19) of the coefficients b and σ up to time t.

by Definition 2.5 shows that Φ^N is a contraction w.r.t. $\widehat{\mathcal{W}}_{k,p}$ for all large enough $N \in \mathbb{N}$. Therefore, Φ has the properties (i) and (ii) claimed above. $\square$

Remark 2.33. (a) Theorem 2.32 is usually proven (see e.g. [10, Theorem 1.7], [33, Proposition 2] and [56, Theorem 1.1]) for solutions defined on a finite time horizon $[0, T]$. In this case, significantly less fixed point theory is required, since one can work directly with the Polish spaces (W_n^T, p_T) and $(\mathcal{P}_p(W_n^T), \mathcal{W}_{T,p})$, see Definition 2.5 and Remark 2.6(c) for details. This approach can be pursued in the infinite time horizon by constructing a sequence of solutions on longer and longer finite time intervals, verifying their consistency and patching them together to obtain a solution on $\mathbb{R}_+ \times \Omega$. This obviates the need for Proposition B.20, Proposition 2.7 and Theorem A.16, but lengthens the proof of Theorem 2.32 considerably and makes less use of the topological and measurable structure of $\widehat{\mathcal{P}}_p(W_n)$, which is needed in any case to ensure well-definedness of the stochastic integrals implicitly defined by the McKean SDE.

(b) In the proof of Theorem 2.32, we used the fixed point theorem A.16 to obtain existence and uniqueness of a fixed point $\mathbb{Q}^*$ of the map Φ from (2.69). Under the conditions (i) and (ii) established in the proof, Theorem A.16 also implies that for every $\mathbb{Q}^{(0)} \in \widehat{\mathcal{P}}_p(W_n)$, the sequence defined by $\mathbb{Q}^{(m)} := \Phi^{\circ m}(\mathbb{Q}^{(0)})$, $m \in \mathbb{N}$ converges to $\mathbb{Q}^*$ w.r.t. the metric $\widehat{\mathcal{W}}_p$ from Proposition 2.7. The following corollary gives the rate of convergence for the restrictions to finite time horizons of the sequence $\mathbb{Q}^{(m)}$ to $\mathbb{Q}^*$, and also of an associated sequence of solutions of classical SDEs to the solution of the McKean SDE.

(c) In Theorem 2.32, the bounded growth condition (see (2.50) in Definition 2.20) can be replaced by a growth condition of the form

$$|b(t, w, \mathbb{Q}, \omega)| + |\sigma(t, w, \mathbb{Q}, \omega)| \le D_t\big(1 + p_t(w) + \widehat{M}_{t,p}^{1/p}\big), \tag{2.77}$$

for all $t \in \mathbb{R}_+$, $w \in W_n$, $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$ and all $\omega \in \Omega$, where $D\colon \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing function and $\widehat{M}_{t,p}$ is defined in (3.59). Indeed, the proof of Theorem 2.32 uses Definition 2.20 only through Remark 2.24, which remains valid assuming (2.77) instead of (2.50). Hence no modification of the proof is necessary. Note that (2.77) is ensured by the Lipschitz property if the coefficients b and σ satisfy in addition the bound

$$|b(t, 0, \delta_0, \omega)| + |\sigma(t, 0, \delta_0, \omega)| \le \tilde{D}_t, \qquad (t, \omega) \in \mathbb{R}_+ \times \Omega,$$

where 0 and δ_0 denote respectively the zero element in W_n and its associated Dirac measure. In the case of nonrandom dependence of coefficients, the bound in the previous display holds certainly if in addition to the Lipschitz property being satsified, $b(\,\cdot\,, 0, \delta_0)$ and $\sigma(\,\cdot\,, 0, \delta_0)$ are left- or right-continuous. However, replacing (2.50) with (2.77) breaks (2.82) in Remark 2.36 below. Furthermore, the proof of Theorem 3.19 does not carry over to this extended setting.

Corollary 2.34 (Solution of the McKean SDE as a limit of an iterated sequence of solutions of classical SDEs)**.** *Consider the setting of Theorem 2.32. Fix* $\mathbb{Q}^{(0)} \in \widehat{\mathcal{P}}_p(W_n)$ *and define* *iteratively a sequence $(X^{(m)})_{m\in\mathbb{N}}$ of $\mathbb{F}$-adapted continuous processes, where for each $m\in\mathbb{N}$, $X^{(m)}$ is the pathwise unique solution of the classical SDE*

$$\mathrm{d}X_t^{(m)} = b(t, X^{(m)}, \mathbb{Q}^{(m-1)}, \,\cdot\,)\,\mathrm{d}t + \sigma(t, X^{(m)}, \mathbb{Q}^{(m-1)}, \,\cdot\,)\,\mathrm{d}B_t\,, \qquad X_0^{(m)} \overset{\text{a.s.}}{=} \xi, \tag{2.78}$$

see also Remark 2.24 (a), and

$$\mathbb{Q}^{(m)} := \mathcal{L}_{\mathbb{P}}[X^{(m)}]. \tag{2.79}$$

Let $X^* = (X_t^*)_{t\in\mathbb{R}_+}$ be the pathwise unique strong solution of the McKean SDE (2.37) and $\mathbb{Q}^* := \mathcal{L}_{\mathbb{P}}[X^*]$ the associated unique fixed point of the map Φ from (2.69) (see observations (i) and (ii) in the proof of Theorem 2.32). Then for all $m\in\mathbb{N}$ and $t\in\mathbb{R}_+$,

$$\widehat{\mathcal{T}}_{t,p}(\mathbb{Q}^{(m)}, \mathbb{Q}^*) \le \frac{(L_t\cdot t)^m}{m!}\widehat{\mathcal{T}}_{t,p}(\mathbb{Q}^{(0)}, \mathbb{Q}^*) \tag{2.80}$$

and

$$\mathbb{E}[p_t^p(X^{(m)} - X^*)] \le \frac{(L_t\cdot t)^{m+1}}{(m+1)!}\widehat{\mathcal{T}}_{t,p}(\mathbb{Q}^{(0)}, \mathbb{Q}^*), \tag{2.81}$$

where the $\mathbb{R}_+$-valued increasing function L was defined in (2.73).

Proof. Fix $m\in\mathbb{N}$ and $t\in\mathbb{R}_+$. By (2.78), (2.79) and the definition (2.69) of Φ, $\mathbb{Q}^{(m)} = \Phi^{\circ m}(\mathbb{Q}^{(0)})$. Since $\mathbb{Q}^*$ is a fixed point of Φ, (2.76) implies that

$$\widehat{\mathcal{T}}_{t,p}(\mathbb{Q}^{(m)}, \mathbb{Q}^*) = \widehat{\mathcal{T}}_{t,p}(\Phi^{\circ m}(\mathbb{Q}^{(0)}), \Phi^{\circ m}(\mathbb{Q}^*)) \le \frac{(L_t\cdot t)^m}{m!}\widehat{\mathcal{T}}_{t,p}(\mathbb{Q}^{(0)}, \mathbb{Q}^*),$$

establishing (2.80). From (2.71) and (2.72) with $\mathbb{Q} := \mathbb{Q}^{(m)}$ and $\tilde{\mathbb{Q}} := \mathbb{Q}^*$ for the first inequality, using for the second inequality that the previous display holds for all $t\in\mathbb{R}_+$ and for the last inequality the monotonicity property (2.23), we get

$$\begin{aligned}
\mathbb{E}[p_t^p(X^{(m)} - X^*)] &\le L_t\int_0^t \widehat{\mathcal{T}}_{s,p}(\mathbb{Q}^{(m)}, \mathbb{Q}^*)\,\mathrm{d}s \\
&\le L_t\int_0^t \frac{(L_s\cdot s)^m}{m!}\widehat{\mathcal{T}}_{s,p}(\mathbb{Q}^{(0)}, \mathbb{Q}^*)\,\mathrm{d}s \\
&\le \frac{(L_t\cdot t)^{m+1}}{(m+1)!}\widehat{\mathcal{T}}_{t,p}(\mathbb{Q}^{(0)}, \mathbb{Q}^*),
\end{aligned}$$

which shows (2.81). □

Remark 2.35. The approximation of a solution of the McKean SDE by solutions to the associated sequence of classical SDEs defined by (2.78) appears in a similar form in McKean's original paper [40], where it is used to prove strong existence of solutions without making use of a fixed point argument.

Remark 2.36 (Bound for the p-th moments of increments in Theorem 2.32)**.** Consider the setting of Theorem 2.32. By Step 1 of the proof of that theorem, the McKean SDE (2.37) is equivalent to the classical SDE (2.70), where $\mathbb{Q}^*$ denotes the law of the solution on the given set-up.[27] This classical SDE has the $\mathbb{G}\otimes\mathbb{F}$-progressive (see Lemma 2.15 (a)) coefficients $\tilde b$ and $\tilde\sigma$ given for all $(t,w,\omega)\in\mathbb{R}_+\times W_n\times\Omega$ by

$$\tilde b(t,w,\omega) \coloneqq b(t,w,\mathbb{Q}^*,\omega) \quad\text{and}\quad \tilde\sigma(t,w,\omega) \coloneqq \sigma(t,w,\mathbb{Q}^*,\omega).$$

By the assumptions of Theorem 2.32, b and σ have the Lipschitz property and are of linearly bounded growth (Definitions 2.19 and 2.20) in the sense of McKean SDEs. Hence $\tilde b$ and $\tilde\sigma$ have the Lipschitz property and are of linearly bounded growth in the sense of classical SDEs (Definitions C.25 and C.27). From Remark C.35 applied to the classical SDE (2.70), it follows that an $L^p(\mathbb{P})$-bound for the seminorm of the increments is given by

$$\|p_t(X-\xi)\|_{L^p(\mathbb{P})} \le \sum_{k=1}^{\infty}\Big(\frac{(K_t t)^k}{k!}\Big)^{1/p}, \qquad t\in\mathbb{R}_+, \tag{2.82}$$

where the series converges by the ratio test and K is given by[28] (C.42) with $I=1$, the function c from Definition 2.19 and the function D from Definition 2.20.

2.3. Examples of McKean SDEs

We consider in this section, which is based on [10, §1.3.2], two prominent classes of examples of McKean SDEs. In both, the coefficients are as in (2.40), i.e. we have instantaneous, nonrandom dependence of the coefficients on the solution in the sense of Definition 2.13. We apply the convention described in Definition 2.13, making no notational distinction between b and $\hat b$ and σ and $\hat\sigma$ respectively. Due to Remark 2.23 (a), we will, when discussing properties of coefficients, often consider the order $p=1$ without loss of generality.

Definition 2.37 (McKean–Vlasov SDE, [10, §1.3.2])**.** The prototypical McKean SDE is given by the McKean–Vlasov SDE, where the coefficients $b\colon\mathbb{R}_+\times\mathbb{K}^n\times\mathcal{P}_1(\mathbb{K}^n)\to\mathbb{K}^n$ and $\sigma\colon\mathbb{R}_+\times\mathbb{K}^n\times\mathcal{P}_1(\mathbb{K}^n)\to\mathbb{K}^{n\times d}$ are of the form

$$b(t,x,\mu) \coloneqq \int_{\mathbb{K}^n}\tilde b(t,x,y)\,\mathrm{d}\mu(y)\,, \qquad \sigma(t,x,\mu) \coloneqq \int_{\mathbb{K}^n}\tilde\sigma(t,x,y)\,\mathrm{d}\mu(y)\,, \tag{2.83}$$

for suitable (see Lemma 2.39 below) functions

$$\tilde b\colon\mathbb{R}_+\times\mathbb{K}^n\times\mathbb{K}^n\to\mathbb{K}^n \quad\text{and}\quad \tilde\sigma\colon\mathbb{R}_+\times\mathbb{K}^n\times\mathbb{K}^n\to\mathbb{K}^{n\times d}. \tag{2.84}$$

We will often write b instead of $\tilde b$, and similarly σ instead of $\tilde\sigma$.

Remark 2.38 (Higher order interactions [10, §1.3.2])**.** The dependence of b and σ on μ in Definition 2.37 is also called interaction of order 1. Interactions of higher order can be defined, e.g. by

$$b(t,x,\mu) = \int_{\mathbb{K}^n}\int_{\mathbb{K}^n}\tilde b(t,x,y,z)\,\mathrm{d}\mu(y)\,\mathrm{d}\mu(z)$$

and

$$\sigma(t,x,\mu) = \int_{\mathbb{K}^n}\int_{\mathbb{K}^n}\tilde\sigma(t,x,y,z)\,\mathrm{d}\mu(y)\,\mathrm{d}\mu(z)$$

for an interaction of order 2.

[27] By Theorem 2.32, $\mathbb{Q}^*$ is independent of the set-up in the case of nonrandom dependence.
[28] See (C.22) for the definition of the function β appearing in (C.42).

Lemma 2.39 (Admissibility of coefficients of the McKean–Vlasov SDE)**.** *Let $\tilde{b}$ and $\tilde{\sigma}$ be as in (2.84) with the following additional properties:*

(a) For all $x, y \in \mathbb{K}^n$, the sections $t \mapsto \tilde{b}(t,x,y)$ and $t \mapsto \tilde{\sigma}(t,x,y)$ are $\mathcal{B}_{\mathbb{R}_+}$-measurable.

(b) The functions $\tilde{b}$ and $\tilde{\sigma}$ are Lipschitz-continuous uniformly in (x,y), locally in t, meaning that there exists an increasing function $c\colon \mathbb{R}_+ \to \mathbb{R}_+$ with

$$\left|\tilde{b}(t,x,y) - \tilde{b}(t,\tilde{x},\tilde{y})\right| + |\tilde{\sigma}(t,x,y) - \tilde{\sigma}(t,\tilde{x},\tilde{y})| \le c_t(|x-\tilde{x}| + |y-\tilde{y}|), \tag{2.85}$$

for all $t \in \mathbb{R}_+$ and all $x, \tilde{x}, y, \tilde{y} \in \mathbb{K}^n$. Then the coefficients b and σ defined by (2.83) are $\mathbb{G}\otimes\mathbb{H}$-progressive (Definition 2.11) when viewed as maps on $\mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n)$ and satisfy the instantaneous Lipschitz property (Definition 2.22) for each $p \in [1,\infty)$. If in addition

(c) there exists an increasing function $D\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$|\tilde{b}(t,x,y)| + |\tilde{\sigma}(t,x,y)| \le D_t(1+|x|), \qquad (t,x,y) \in \mathbb{R}_+ \times \mathbb{K}^n \times \mathbb{K}^n, \tag{2.86}$$

then b and σ also satisfy the linearly bounded growth condition (Definition 2.20).

Proof. By Lemma B.10, measurability in t (Assumption (a)) and Lipschitz continuity in (x,y) (Assumption (b)) together imply $\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{B}_{\mathbb{K}^n}$-measurability of $\tilde{b}$ and $\tilde{\sigma}$. Let $t \in \mathbb{R}_+$ and let $x_0, y_0 \in \mathbb{K}^n$ be two points. The bound

$$\left|\tilde{b}(t,x,y)\right| \le \left|\tilde{b}(t,x,y) - \tilde{b}(t,x_0,y_0)\right| + \left|\tilde{b}(t,x_0,y_0)\right| \le c_t(|x-x_0| + |y-y_0|) + \left|\tilde{b}(t,x_0,y_0)\right|$$

then shows that $y \mapsto \tilde{b}(t,x,y)$ is in $\mathcal{L}^1(\mu)$ for all $(t,x) \in \mathbb{R}_+ \times \mathbb{K}^n$ and every $\mu \in \mathcal{P}_1(\mathbb{K}^n)$. The same argument applies to $\tilde{\sigma}$, hence the coefficients b and σ given by (2.83) are well defined and, by Fubini's theorem, measurable in (t,x) for each fixed $\mu \in \mathcal{P}_1(\mathbb{K}^n)$. For the instantaneous Lipschitz property, it suffices by (a) to establish the bound

$$|b(t,x,\mu) - b(t,y,\nu)| + |\sigma(t,x,\mu) - \sigma(t,y,\nu)| \le c_t\left(|x-y| + d_1(\mu,\nu)\right), \tag{2.87}$$

where d_1 denotes the Wasserstein metric (Definition 2.4) on $\mathcal{P}_1(\mathbb{K}^n)$. To this end, fix $t \in \mathbb{R}_+$, $x, y \in \mathbb{K}^n$, $\mu, \nu \in \mathcal{P}_p(\mathbb{K}^n)$ and a coupling $\pi \in \mathcal{P}(\mu,\nu)$. Then using (2.85) for the first inequality,

$$\begin{aligned}
&|b(t,x,\mu) - b(t,y,\nu)| + |\sigma(t,x,\mu) - \sigma(t,y,\nu)| \\
&\qquad = \left|\int_{\mathbb{K}^n} \tilde{b}(t,x,z) - \tilde{b}(t,y,\tilde{z})\,\mathrm{d}\pi(z,\tilde{z})\right| + \left|\int_{\mathbb{K}^n} \tilde{\sigma}(t,x,z) - \tilde{\sigma}(t,y,\tilde{z})\,\mathrm{d}\pi(z,\tilde{z})\right| \\
&\qquad \le \int_{\mathbb{K}^n} |\tilde{b}(t,x,z) - \tilde{b}(t,y,\tilde{z})| + |\tilde{\sigma}(t,x,z) - \tilde{\sigma}(t,y,\tilde{z})|\,\mathrm{d}\pi(z,\tilde{z}) \\
&\qquad \le c_t \int_{\mathbb{K}^n} |x-y| + |z-\tilde{z}|\,\mathrm{d}\pi(z,\tilde{z}) = c_t\left(|x-y| + \int_{\mathbb{K}^n} |z-\tilde{z}|\,\mathrm{d}\pi(z,\tilde{z})\right).
\end{aligned}$$

Taking the infimum over $\pi \in \mathcal{P}(\mu,\nu)$ and comparing with (D.2), the bound (2.87) follows. From Remark 2.23 (b), b and σ also have the Lipschitz property (Definition 2.19). Using

Assumption (a) and Lemma 2.21 (b), $\mathbb{G} \otimes \mathbb{H}$-progressive-measurability of b and σ follows. Finally, the addition follows from (2.83) and (2.86) from the estimate

$$|b(t,x,\mu)| + |\sigma(t,x,\mu)| \le \int_{\mathbb{K}^n} |\tilde{b}(t,x,y)| + |\tilde{\sigma}(t,x,y)| \,\mathrm{d}\mu(y) \le D_t(1+|x|),$$

which is valid for all $(t,x,\mu) \in \mathbb{R}_+ \times \mathbb{K}^n \times \mathcal{P}_1(\mathbb{K}^n)$. □

The next type of McKean SDE is given by the mean-field interaction of scalar type, where the dependence of $b(t,x,\mu)$ and $\sigma(t,x,\mu)$ on μ is through a dependence on a finite number m of moments of μ.

Definition 2.40 (Mean-field interaction of scalar type, [10, §1.3.2])**.** The McKean SDE with interaction of scalar type is defined by suitable functions (see Lemma 2.41 below)

$$\begin{aligned} &\tilde{b}\colon \mathbb{R}_+ \times \mathbb{K}^n \times \mathbb{K}^m \to \mathbb{K}^n, \\ &\tilde{\sigma}\colon \mathbb{R}_+ \times \mathbb{K}^n \times \mathbb{K}^m \to \mathbb{K}^{n\times d}, \\ &\varphi\colon \mathbb{K}^n \to \mathbb{K}^m, \end{aligned} \tag{2.88}$$

which give rise to coefficients b and σ as in (2.40) by

$$b(t,x,\mu) \coloneqq \tilde{b}(t,x,\langle\varphi,\mu\rangle), \qquad \sigma(t,x,\mu) \coloneqq \tilde{\sigma}(t,x,\langle\varphi,\mu\rangle), \tag{2.89}$$

for all $t \in \mathbb{R}_+$, $x \in \mathbb{K}^n$ and $\mu \in \mathcal{P}_1(\mathbb{K}^n)$, where $\langle\varphi,\mu\rangle \coloneqq \int_{\mathbb{K}^n} \varphi \,\mathrm{d}\mu$.

Lemma 2.41 (Admissibility of coefficients for the mean-field interaction of scalar type)**.** *Let $\tilde{b}$ and $\tilde{\sigma}$ and $\varphi\colon \mathbb{K}^n \to \mathbb{K}^m$ be functions as in (2.88) with the following properties:*

(a) For all $x \in \mathbb{K}^n$ and $y \in \mathbb{K}^m$, the sections $t \mapsto \tilde{b}(t,x,y)$ and $t \mapsto \tilde{\sigma}(t,x,y)$ are $\mathcal{B}_{\mathbb{R}_+}$-measurable.

(b) The functions $\tilde{b}$ and $\tilde{\sigma}$ are Lipschitz-continuous uniformly in (x,y), locally in t, meaning that there exists an increasing function $c\colon \mathbb{R}_+ \to \mathbb{R}_+$ with

$$|\tilde{b}(t,x,y) - \tilde{b}(t,\tilde{x},\tilde{y})| + |\tilde{\sigma}(t,x,y) - \tilde{\sigma}(t,\tilde{x},\tilde{y})| \le c_t(|x-\tilde{x}| + |y-\tilde{y}|).$$

for all $t \in \mathbb{R}_+$, $x, \tilde{x} \in \mathbb{K}^n$ and $y, \tilde{y} \in \mathbb{K}^m$.

(c) The function φ is Lipschitz continuous.

Then the coefficients b and σ defined by (2.89) are $\mathbb{G} \otimes \mathbb{H}$-progressive (Definition 2.11) when viewed as maps on $\mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n)$ and satisfy the instantaneous Lipschitz property (Definition 2.22) for each $p \in [1,\infty)$.

Proof. By Lemma B.10, measurability in t and Lipschitz continuity in (x,y) together imply $\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{B}_{\mathbb{K}^m}$-measurability of $\tilde{b}$ and $\tilde{\sigma}$. For all $t \in \mathbb{R}_+$, we have

$$\begin{aligned} &|b(t,x,\mu) - b(t,y,\nu)| + |\sigma(t,x,\mu) - \sigma(t,y,\nu)| \\ &\qquad = |\tilde{b}(t,x,\langle\varphi,\mu\rangle) - \tilde{b}(t,y,\langle\varphi,\nu\rangle)| + |\tilde{\sigma}(t,x,\langle\varphi,\mu\rangle) - \tilde{\sigma}(t,y,\langle\varphi,\nu\rangle)| \\ &\qquad \le c_t\left(|x-y| + |\langle\varphi,\mu\rangle - \langle\varphi,\nu\rangle|\right). \end{aligned}$$

We next establish the instantaneous Lipschitz property. Let $K \in \mathbb{R}_+$ be a Lipschitz constant for φ. The estimate

$$|\varphi(x)| \leq |\varphi(0)| + |\varphi(x) - \varphi(0)| \leq |\varphi(0)| + K|x|, \qquad x \in \mathbb{K}^n,$$

shows that φ is integrable w.r.t. every distribution with a finite expectation. In particular, $\int \varphi \, \mathrm{d}\mu$ and $\int \varphi \, \mathrm{d}\nu$ are well-defined. For every coupling $\pi \in \mathcal{P}(\mu, \nu)$,

$$\begin{aligned} |\langle \varphi, \mu \rangle - \langle \varphi, \nu \rangle| &= \left| \int_{\mathbb{K}^n} \varphi(x) \, \mathrm{d}\mu(x) - \int_{\mathbb{K}^n} \varphi(y) \, \mathrm{d}\nu(y) \right| \\ &= \left| \int_{\mathbb{K}^n \times \mathbb{K}^n} \varphi(x) - \varphi(y) \; \mathrm{d}\pi(x, y) \right| \leq \int_{\mathbb{K}^n \times \mathbb{K}^n} |\varphi(x) - \varphi(y)| \, \mathrm{d}\pi(x, y) \\ &\leq K \int_{\mathbb{K}^n \times \mathbb{K}^n} |x - y| \, \mathrm{d}\pi(x, y) \, . \end{aligned}$$

Taking the infimum over all couplings $\pi \in \mathcal{P}(\mu, \nu)$ in the previous display and comparing with (D.2) shows

$$|\langle \varphi, \mu \rangle - \langle \varphi, \nu \rangle| \leq K d_1(\mu, \nu),$$

and thus

$$|b(t, x, \mu) - b(t, y, \nu)| + |\sigma(t, x, \mu) - \sigma(t, y, \nu)| \leq c_t(|x - y| + K d_1(\mu, \nu)).$$

This establishes the instantaneous Lipschitz property (Definition 2.22) for all $p \in [1, \infty)$ by Remark 2.23 (a). Finally, using Remark 2.23 (b), Lemma 2.21 (b) and Assumption (a), $\mathbb{G} \otimes \mathbb{H}$-progressive measurability of b and σ follows. □

Remark 2.42 (Interaction of scalar type in the case where φ is not Lipschitz continuous)**.** Lipschitz continuity of the function φ from (2.88) was needed in Lemma 2.41 to ensure that the coefficients defined by (2.89) have the instantaneous Lipschitz property (2.51). Unfortunately, the interesting example

$$\varphi(x) \coloneqq |x|^2, \qquad x \in \mathbb{K}^n,$$

which would allow b and σ to depend on the second moment (and, with some minor modifications, also on the variance) of the solution, is not covered by the results developed so far. Fortunately, assuming that the dependence of b and σ on higher moments of the measure argument is of a specific form, an analogue of Theorem 2.32 does hold and is proven in Subsection 3.2.2, see Theorem 3.19. This leads to a class of SDEs involving variance dampening, i.e. the magnitude of new noise acting on the solution is reduced once the solution breaches a defined level of variance. Such SDEs are studied in Subsection 3.2.2 beginning with Example 3.21, see in particular Examples 3.24 and 3.26.

Remark 2.43 (Polynomial McKean SDEs)**.** Fix the dimension $n = 1$ and let $\mathbb{K} = \mathbb{R}$. If the function φ from (2.88) has the form

$$\varphi(x) = (x, x^2, x^m), \quad x \in \mathbb{R},$$

we call the corresponding SDE with coefficients as in (2.89) a *polynomial McKean SDE*. Existence, uniqueness and further properties of solutions for SDEs of this type have been studied recently in [15].

2.4. Propagation of Chaos for McKean SDEs

This section is split into two main parts. Subsection 2.4.1, which is based on [26, §10.3.2], introduces the notion of a chaotic sequence of probability measures, which makes precise the meaning of propagation of chaos, and describes related properties. Subsequently in Subsection 2.4.2, we introduce, building on [26, §10.2], [33] and [10, §1.3.4], the particle approximation of the McKean SDE and prove its convergence.

2.4.1. Chaotic Sequences of Probability Measures

Definition 2.44 (Empirical measure)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $X_1, \ldots, X_N$ random variables with values in a measurable space $(S, \mathcal{S})$. The *empirical measure* associated to $(X_1, \ldots, X_N)$ is the $\mathcal{P}(S)$-valued map defined by

$$\tilde{\mu}_{X_1,\ldots,X_N}(\omega) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i(\omega)}, \qquad \omega \in \Omega.$$

Appendix D contains several facts about empirical measures for the case where S is a Polish space and $\mathcal{S}$ is its Borel σ-algebra, e.g. that empirical measures of "close" sets of random variables are close in the Wasserstein metric (Example D.3), that the sequence of empirical measures associated with an i.i.d. family $(\xi_i)_{i\in\mathbb{N}}$ of random variables with law $\mathcal{L}[\xi_1] =: \mu \in \mathcal{P}_p(S)$ converges to μ w.r.t. the Wasserstein metric $\mathcal{W}_p$ on $\mathcal{P}_p(S)$, and in addition gives a simple criterion for the measurability of empirical measures w.r.t. the topologies induced by the Wasserstein distances (Lemma D.16). The next definition, which extends [26, Definition 10.2], makes precise the notion of a chaotic sequence of probability measures.

Definition 2.45 (Chaotic sequence of probability measures, [11, Definition 3.18])**.** Let (S, d) be a metric space and denote, for each $N \in \mathbb{N}$, by μ^N a probability measure on the N-fold product space S^N, which is equipped with the product σ-algebra $\mathcal{B}_S^{\otimes N}$.

(a) Let $\mu \in \mathcal{P}(S)$. The sequence $(\mu^N)_{N\in\mathbb{N}}$ is called *μ-chaotic*, if for each fixed $k \in \mathbb{N}$ and all $\varphi_1, \ldots, \varphi_k \in C_\mathrm{b}(S)$,

$$\int_{S^N} \varphi_1(x_1)\cdots\varphi_k(x_k)\,\mathrm{d}\mu^N(x_1,\ldots,x_N) \xrightarrow[k\le N\to\infty]{} \prod_{i=1}^{k}\int_S \varphi_i\,\mathrm{d}\mu\,. \tag{2.90}$$

(b) We say that $(\mu^N)_{N\in\mathbb{N}}$ is *chaotic*, if there exists a $\mu \in \mathcal{P}(S)$ such that $(\mu^N)_{N\in\mathbb{N}}$ is μ-chaotic.

Lemma 2.46 (Equivalence of μ-chaoticity to weak convergence of each sequence of k-dimensional marginals to $\mu^{\otimes k}$)**.** *Let (S, d) be a complete metric space, and denote, for all $N \in \mathbb{N}$ and $k \in \{1, \ldots, N\}$, by*

$$\pi_k^N : S^N \to S, \quad \pi_k^N(x_1,\ldots,x_N) := x_k, \tag{2.91}$$

the projection from S^N onto the k-th factor. Furthermore, let

$$\pi^N_{[1,k]}\colon S^N \to S^k, \quad \pi^N_{[1,k]}(x_1,\dots,x_N) \coloneqq (x_1,\dots,x_k), \tag{2.92}$$

be the projection onto the first k factors. A sequence $(\mu^N)_{N\in\mathbb{N}}$ as in Definition 2.45 is μ-chaotic if and only if for each $k\in\mathbb{N}$, $(\mu^N\circ(\pi^N_{[1,k]})^{-1})_{N\geq k}$ converges weakly to $\mu^{\otimes k}$ as $N\to\infty$.

Proof. The reverse direction follows directly from the definition 2.45 of chaoticity.[29] For the direct implication, assume that $(\mu^N)_{N\in\mathbb{N}}$ is μ-chaotic. The idea is to use Prokhorov's theorem to obtain tightness of each family $\{\mu^N\circ(\pi^N_{[1,k]})^{-1}\}_{N\geq k}$ of k-dimensional marginal distributions, infer relative sequential compactness again from Prokhorov's theorem and then verify uniqueness of subsequential limits.

Since $(\mu^N)_{N\in\mathbb{N}}$ is μ-chaotic, using (2.90), the sequence $(\mu^N\circ(\pi^N_k)^{-1})_{N\geq k}$ converges, for every $k\in\mathbb{N}$, weakly to μ. This implies using completeness of S and Prokhorov's theorem ([53, Theorem 8.67]) tightness of the family $\{\mu^N\circ(\pi^N_k)^{-1}\}_{N\geq k}$ of univariate marginal distributions for all $k\in\mathbb{N}$. As a consequence (see [53, Exercise 8.65]), the family $\{\mu^N\circ(\pi^N_{[1,k]})^{-1}\}_{N\geq k}$ of k-variate marginal distributions is tight for each $k\in\mathbb{N}$, and thus relatively sequentially compact by another application of Prokhorov's theorem. Fixing $k\in\mathbb{N}$, we have by (2.90) that every weak subsequential limit $\nu\in\mathcal{P}(S^k)$ of $(\mu^N\circ(\pi^N_{[1,k]})^{-1})_{N\geq k}$ satisfies

$$\int_{S^k}\varphi_1(x_1)\cdots\varphi_k(x_k)\,\mathrm{d}\nu(x_1,\dots,x_k) = \int_S\varphi_1\,\mathrm{d}\mu\cdots\int_S\varphi_k\,\mathrm{d}\mu\,,$$

for all $\varphi_1,\dots,\varphi_k\in C_\mathrm{b}(S)$. By approximating indicator functions of products of closed sets in S by products of bounded continuous functions (see [36, Lemma 13.10]), it follows that

$$\nu(C_1\times\cdots\times C_k) = \mu^{\otimes k}(C_1\times\cdots\times C_k),$$

for all closed sets $C_1,\dots,C_k$ in S. The family of sets of this form is a π-system generating $\mathcal{B}_S^{\otimes k}$, hence $\nu=\mu^{\otimes k}$ follows by [53, Corollary 15.67]. Therefore all weak subsequential limits of $(\mu^N\circ(\pi^N_{[1,k]})^{-1})_{N\geq k}$ agree, which together with the already established relative sequential compactness proves that $\mu^N\circ(\pi^N_{[1,k]})^{-1}$ converges weakly[30] to $\mu^{\otimes k}$ as $N\to\infty$. □

Definition 2.47 (Symmetric probability measure, [36, §12])**.** Let $(\Omega,\mathcal{F})$ be a measurable space, fix $N\in\mathbb{N}$ and let S_N denote the symmetric group of degree N. A probability measure μ on $(\Omega^N,\mathcal{F}^{\otimes N})$ is called symmetric, if

$$\mu\circ T_\pi^{-1} = \mu,$$

for all $\pi\in S_N$, where $T_\pi\colon\Omega^N\to\Omega^N$ is defined by

$$T_\pi(\omega_1,\dots,\omega_N) = (\omega_{\pi(1)},\dots,\omega_{\pi(N)}).$$

For symmetric sequences of probability measures, simplified characterizations of chaoticity are available.

[29] This direction does not require completeness of S.

[30] To be precise, we use that weak convergence is convergence in a topology (the weak topology, see e.g. [4, Theorem 7.20]), and the general fact that a sequence $(x_n)_{n\in\mathbb{N}}$ in a topological space $(X,\mathcal{O})$ converges to a point $x\in X$ if and only if every subsequence of $(x_n)_{n\in\mathbb{N}}$ has a subsequence converging to x.

Lemma 2.48 (Characterization of chaoticity for sequences of symmetric probability measures [26, Theorem 10.2])**.** *Fix a metric space (S, d) and $\mu \in \mathcal{P}(S)$ and let, for each $N \in \mathbb{N}$, μ^N be a symmetric probability measure on S^N. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space carrying, for each $N \in \mathbb{N}$, an S^N-valued random variable $X^N = (X^{1,N}, \dots, X^{N,N})$ with law μ^N, and denote by $\tilde{\mu}^N$ the associated empirical measure (Definition 2.44). The following are equivalent:*

(a) The sequence $(\mu^N)_{N\in\mathbb{N}}$ is μ-chaotic (Definition 2.45).

(b) For all $\varphi_1, \varphi_2 \in C_\mathrm{b}(S)$, (2.90) holds with $k = 2$, i.e.

$$\int_{S^N} \varphi_1(x_1)\varphi_2(x_2)\,\mathrm{d}\mu^N(x_1, \dots, x_N) \xrightarrow[2\le N\to\infty]{} \int_S \varphi_1\,\mathrm{d}\mu \int_S \varphi_2\,\mathrm{d}\mu\,.$$

(c) For all $\varphi \in C_\mathrm{b}(S)$,

$$\int_S \varphi\,\mathrm{d}\tilde{\mu}^N \xrightarrow[N\to\infty]{L^1(\mathbb{P})} \int_S \varphi\,\mathrm{d}\mu\,, \tag{2.93}$$

where $\int_S \varphi\,\mathrm{d}\tilde{\mu}^N = \frac{1}{N}\sum_{i=1}^N \varphi(X^{i,N})$ (Definition 2.44).

Proof. (a) $\Rightarrow$ (b): Clear.

(b) $\Rightarrow$ (c): We show that we even get $L^2(\mathbb{P})$-convergence[31] of $\int_S \varphi\,\mathrm{d}\tilde{\mu}^N$ to $\int_S \varphi\,\mathrm{d}\mu$. Using symmetry of μ^N (Definition 2.47) in the second equality, we have

$$\begin{aligned}
&\mathbb{E}\Bigg[\bigg(\frac{1}{N}\sum_{i=1}^N \varphi(X^{i,N}) - \int_S \varphi\,\mathrm{d}\mu\bigg)^2\Bigg] \\
&= \mathbb{E}\Bigg[\frac{1}{N^2}\sum_{i,j=1}^N \varphi(X^{i,N})\varphi(X^{j,N}) - 2\int_S \varphi\,\mathrm{d}\mu \cdot \frac{1}{N}\sum_{i=1}^N \varphi(X^{i,N}) + \bigg(\int_S \varphi\,\mathrm{d}\mu\bigg)^2\Bigg] \\
&= \frac{1}{N}\mathbb{E}[\varphi(X^{1,N})^2] + \frac{N-1}{N}\mathbb{E}\left[\varphi(X^{1,N})\varphi(X^{2,N})\right] \\
&\qquad\qquad - 2\int_S \varphi\,\mathrm{d}\mu \cdot \mathbb{E}\left[\varphi(X^{1,N})\right] + \bigg(\int_S \varphi\,\mathrm{d}\mu\bigg)^2.
\end{aligned}$$

As $N \to \infty$, the first term on the right-hand side above goes to zero by boundedness of φ. The second and third term each go by assumption (b) to $\big(\int_S \varphi\,\mathrm{d}\mu\big)^2$. Therefore, $\int_S \varphi\,\mathrm{d}\tilde{\mu}^N \xrightarrow[N\to\infty]{} \int_S \varphi\,\mathrm{d}\mu$ in $L^2(\mathbb{P})$, which proves (c).

(c) $\Rightarrow$ (a): Fix $\varphi_1, \dots, \varphi_k \in C_\mathrm{b}(S)$. We seek to establish (2.90), or equivalently

$$\left|\mathbb{E}[\varphi_1(X^{1,N})\cdots\varphi_k(X^{k,N})] - \int \varphi_1\,\mathrm{d}\mu \cdots \int \varphi_k\,\mathrm{d}\mu\right| \xrightarrow[k\le N\to\infty]{} 0.$$

[31] By boundedness of φ and Vitali's convergence theorem [53, Theorem 4.60], convergence in probability and L^p, $p \ge 1$ are all equivalent in the present case.

By the triangle inequality,

$$\begin{aligned}\Big|\mathbb{E}[\varphi_1(X^{1,N})\cdots\varphi_k(X^{k,N})] - \int\varphi_1\,\mathrm{d}\mu\cdots\int\varphi_k\,\mathrm{d}\mu\Big| \\ \leq \Big|\mathbb{E}[\varphi_1(X^{1,N})\cdots\varphi_k(X^{k,N})] - \mathbb{E}\Big[\int\varphi_1\,\mathrm{d}\tilde{\mu}^N\cdots\int\varphi_k\,\mathrm{d}\tilde{\mu}^N\Big]\Big| \\ + \Big|\mathbb{E}\Big[\int\varphi_1\,\mathrm{d}\tilde{\mu}^N\cdots\int\varphi_k\,\mathrm{d}\tilde{\mu}^N\Big] - \int_S\varphi_1\,\mathrm{d}\mu\cdots\int_S\varphi_k\,\mathrm{d}\mu\Big|.\end{aligned} \tag{2.94}$$

We show that both terms of the right-hand side of (2.94) converge to zero. For the second term, we have by assumption (c) that for all $i \in \{1,\dots,k\}$, $\int\varphi_i\,\mathrm{d}\tilde{\mu}^N$ converges to $\int\varphi\,\mathrm{d}\mu$ in $L^1(\mathbb{P})$, and thus also in probability. By the continuous mapping theorem for convergence in probability [58, Theorem 2.3],

$$\int\varphi_1\,\mathrm{d}\tilde{\mu}^N\cdots\int\varphi_k\,\mathrm{d}\tilde{\mu}^N\,\mathrm{d}\mu \xrightarrow[N\to\infty]{\mathbb{P}} \int_S\varphi_1\,\mathrm{d}\mu\cdots\int_S\varphi_k\,\mathrm{d}\mu\,.$$

Since $\varphi_1,\dots,\varphi_k$ are bounded, this implies convergence in $L^1(\mathbb{P})$ by Vitali's convergence theorem [53, Theorem 4.60]. The first term on the right-hand side of (2.94) can be written as

$$\begin{aligned}&\Big|\mathbb{E}[\varphi_1(X^{1,N})\cdots\varphi_k(X^{k,N})] - \frac{1}{N^k}\sum_{i_1,\dots,i_k=1}^N\mathbb{E}[\varphi_1(X^{i_1,N})\cdots\varphi_k(X^{i_k,N})]\Big|. \\ &= \Big|\frac{1}{N^k}\sum_{i_1,\dots,i_k=1}^N\mathbb{E}\left[\varphi_1(X^{1,N})\cdots\varphi_k(X^{k,N}) - \varphi_1(X^{i_1,N})\cdots\varphi_k(X^{i_k,N})\right]\Big|.\end{aligned}$$

In the sum in the last line above, there are $\frac{N!}{(N-k)!}$ terms for which the indices $i_1,\dots,i_k$ are distinct. By symmetry of μ^N, these terms vanish. Each of the remaining $N^k - \frac{N!}{(N-k)!}$ terms is bounded by $2M$, where $M := \prod_{j=1}^k\|\varphi_j\|_\infty$ is finite. Hence the first term on the right-hand side in (2.94) is bounded by

$$2M\left(1 - \frac{1}{N^k}\frac{N!}{(N-k)!}\right) \xrightarrow[N\to\infty]{} 0. \qquad\square$$

We will revisit chaotic sequences of probability measures at the end of the next section, where we will show (see Theorem 2.60) that under the conditions of the strong existence and pathwise uniqueness theorem 2.32 (see also Assumption 2.49), the one-dimensional distributions of the solutions of the N-dimensional particle approximation SDEs (Definition 2.51) are chaotic w.r.t. the one-dimensional marginal distributions of the solution of the McKean SDE.

2.4.2. The Particle Approximation

We consider in this section the McKean SDE (2.36) with nonrandom dependence.[32] By Step 1 of the proof of Theorem 2.32, (2.36) is under the conditions of Theorem 2.32 equivalent

[32] We restrict to the case of nonrandom dependence in order to ensure using Theorem 2.32 uniqueness in law of solutions of (2.95) below.

to the classical SDE (2.70), where $\mathbb{Q}^*$ denotes the law of the solution process. Hence, if $\mathbb{Q}^*$ were known, we could obtain approximate solutions of (2.37) by numerical methods for classical SDEs. Unfortunately, the law of the solution is unknown a priori. While Corollary 2.34 provides one way of obtaining solutions to the McKean SDE as a limit of solutions of classical SDEs, another option is given by the particle method introduced below. This method works by approximating a system of N independent McKean SDEs by a system of coupled classical SDEs, where the coupling occurs by replacing the law of the solution in (2.36) by the empirical measure of component solutions.

Assumption 2.49 (Set-up for the propagation of chaos)**.** Until the end of this section, fix $p \in [2,\infty)$, a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ (Definition C.5 (a)) with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$, and let $\mu \in \mathcal{P}_p(\mathbb{K}^n)$ be an initial distribution. Let $(\xi^i, B^i)_{i\in\mathbb{N}}$ be i.i.d., where for each $i \in \mathbb{N}$, $\xi^i\colon \Omega \to \mathbb{K}^n$ is $\mathcal{F}_0$-measurable with $\mathcal{L}[\xi^i] = \mu$ and B^i is an $\mathbb{R}^d$- or $\mathbb{K}^d$-valued $(\mathbb{F}, \mathbb{P})$-standard Brownian motion. Let b and σ be $\mathbb{G}\otimes\mathbb{H}$-progressive coefficients as in (2.34) which have the Lipschitz property[33] and satisfy the bounded growth condition (Definitions 2.19 and 2.20). For each $i \in \mathbb{N}$, let $X^i = (X^i_t)_{t\in\mathbb{R}_+}$ be a strong solution of the associated McKean SDE

$$\mathrm{d}X^i_t = b(t, X^i, \mathcal{L}[X^i])\,\mathrm{d}t + \sigma(t, X^i, \mathcal{L}[X^i])\,\mathrm{d}B^i_t\,, \qquad X^i_0 \overset{\text{a.s.}}{=} \xi^i, \tag{2.95}$$

which exists by Theorem 2.32.

Remark 2.50 (The processes $(X^i)_{i\in\mathbb{N}}$ are i.i.d. with law $\mathcal{L}[X^1] \in \widehat{\mathcal{P}}_p(W_n)$)**.** For later reference, we note the following consequences of Theorem 2.32.

(a) There is uniqueness in law of solutions of (2.95), hence the family $(X^i)_{i\in\mathbb{N}}$ of W_n-valued random variables is identically distributed.

(b) For each $i \in \mathbb{N}$, X^i is measurable w.r.t. the filtration generated by ξ^i and B^i, augmented[34] by $\mathbb{P}$-null sets in $\sigma(\xi^i, (B^i_t)_{t\in\mathbb{R}_+})$. Since $(\xi^i, B^i)_{i\in\mathbb{N}}$ is an independent family of random variables and independence is preserved under augmentation by null sets (see [53, Exercise 3.20 (d)]), $(X^i)_{i\in\mathbb{N}}$ is an independent family of W_n-valued random variables.

(c) For each $i \in \mathbb{N}$ and $t \in \mathbb{R}_+$, the seminorm $p_t(X^i)$ is p-integrable.

Therefore, the processes $(X^i)_{i\in\mathbb{N}}$ are i.i.d. with law $\mathcal{L}[X^1] \in \widehat{\mathcal{P}}_p(W_n)$.

Definition 2.51 (Particle approximation of McKean SDEs, [10, §1.3.4] and [26, §10.2])**.** Consider the setting of Assumption 2.49 and let $N \in \mathbb{N}$. The N-dimensional *particle approximation* of the McKean SDE (2.36) is the SDE

$$\mathrm{d}X^{i,N}_t = b\big(t, X^{i,N}, \tilde{\mu}^N\big)\,\mathrm{d}t + \sigma\big(t, X^{i,N}, \tilde{\mu}^N\big)\,\mathrm{d}B^i_t\,, \qquad i \in \{1, \dots, N\} \tag{2.96}$$

[33] By Lemma 2.21 (b), the Lipschitz property implies $\mathbb{G}\otimes\mathbb{H}$-progressive measurability as soon as b and σ are Borel measurable in the time argument.

[34] The augmentation is in the sense of Definition B.3.

with initial condition $X_0^N \overset{\text{a.s.}}{=} (\xi^1, \dots, \xi^N)$ for a $\mathbb{K}^{nN}$-valued solution process[35] $X^N = (X^{1,N}, \dots, X^{N,N})$ with associated empirical measure

$$\tilde{\mu}^N := \frac{1}{N} \sum_{i=1}^N \delta_{X^{i,N}}. \tag{2.97}$$

Remark 2.52 (Well-posedness of the particle approximation)**.** The particle approximation (2.96) resembles the SDE (2.45) for which there is strong existence and pathwise uniqueness by Remark 2.24. However (2.96) is not exactly of this type, since $\tilde{\mu}^N$ depends on the solution process X^N. Since existence of this process is, at this point, in question, we cannot treat the dependence on $\tilde{\mu}^N$ simply as a random dependence in the sense of classical SDEs (see Definition C.8) in the coefficients of (2.96). In fact, (2.96) is not obviously a type of SDE we have considered so far. This complication does not arise in the particular case of the McKean–Vlasov SDE, where the coefficients in (2.96) result (see (2.83)) from integration against $\tilde{\mu}^N$, and thus, in this case, it is clear that these coefficients are in fact progressively measurable functions of time and X^N. In Remark 2.53 below, we establish the analogous fact in the general setting.

Remark 2.53 (The particle approximation as a classical SDE)**.** Fix $N \in \mathbb{N}$.

(a) Consider the setting of Definition 2.51. The particle approximation SDE (2.96) can be written as

$$\mathrm{d}X_t^N = b^N(t, X^N)\,\mathrm{d}t + \sigma^N(t, X^N)\,\mathrm{d}B_t^{(N)}, \tag{2.98}$$

where $B^{(N)} = (B^1, \dots, B^N)$ is a $\mathbb{K}^{dN}$-valued standard Brownian motion, and where the coefficients $b^N : \mathbb{R}_+ \times W_{nN} \to \mathbb{K}^{nN}$ and $\sigma^N : \mathbb{R}_+ \times W_{nN} \to \mathbb{K}^{nN \times dN}$ are given by

$$b^N(t, w) := \begin{pmatrix} b^{1,N}(t, w) \\ \vdots \\ b^{N,N}(t, w) \end{pmatrix}, \qquad \sigma^N(t, w) := \begin{pmatrix} \sigma^{1,N}(t, w) & & \\ & \ddots & \\ & & \sigma^{N,N}(t, w) \end{pmatrix}, \tag{2.99}$$

for all $t \in \mathbb{R}_+$ and $w = (w^1, \dots, w^N) \in W_{nN}$, where

$$b^{i,N}(t, w) := b(t, w^i, \hat{\mu}(w)) \qquad \text{and} \qquad \sigma^{i,N}(t, w) := \sigma(t, w^i, \hat{\mu}(w)), \tag{2.100}$$

with $\hat{\mu}$ defined in (D.28). By Lemma D.18, b^N and σ^N are $\mathbb{G}^{nN}$-progressive.

(b) By Assumption 2.49, b and σ have the Lipschitz property and are of linearly bounded growth in the sense of McKean SDEs (Definitions 2.19 and 2.20). We check that this implies that b^N and σ^N have the Lipschitz property and are of linearly bounded growth in the sense of classical SDEs (Definitions C.25 and C.27). We begin with the Lipschitz property. Fix $t \in \mathbb{R}_+$ and $w = (w^1, \dots, w^N), \tilde{w} = (\tilde{w}^1, \dots, \tilde{w}^N) \in W_{nN}$. Using (2.99)

[35] The notation X^N for the $\mathbb{K}^{nN}$-valued solution of (2.96) is strictly speaking an abuse of notation, since it could also denote the $\mathbb{K}^n$-valued solution of (2.95) for $i = N$. This will be unproblematic until Theorem 2.60, where we will denote $(X^{1,N}, \dots, X^{N,N})$ instead by $X^{(N)}$ to avoid this ambiguity. See also Footnote 45.

and subadditivity of the square root to bound the Euclidean (resp. Frobenius) norm by Euclidean (resp. Frobenius) norms of components,

$$
\begin{aligned}
&|b^N(t,w) - b^N(t,\tilde{w})| + |\sigma^N(t,w) - \sigma^N(t,\tilde{w})| \\
&\quad \le \sum_{i=1}^N |b^{i,N}(t,w) - b^{i,N}(t,\tilde{w})| + |\sigma^{i,N}(t,w) - \sigma^{i,N}(t,\tilde{w})| \\
&\quad \overset{(2.100)}{=} \sum_{i=1}^N |b(t,w^i,\hat{\mu}(w)) - b(t,\tilde{w}^i,\hat{\mu}(\tilde{w}))| + |\sigma(t,w^i,\hat{\mu}(w)) - \sigma(t,\tilde{w}^i,\hat{\mu}(\tilde{w}))|.
\end{aligned}
\tag{2.101}
$$

From (2.101) and the Lipschitz property (see (2.49)) of b and σ,

$$
\begin{aligned}
&|b^N(t,w) - b^N(t,\tilde{w})| + |\sigma^N(t,w) - \sigma^N(t,\tilde{w})| \\
&\qquad \le \sum_{i=1}^N c_t\big(p_t(w^i - \tilde{w}^i) + \widehat{\mathcal{W}}_{t,p}(\hat{\mu}(w),\hat{\mu}(\tilde{w}))\big)
\end{aligned}
\tag{2.102}
$$

From the definition (2.21) of $\widehat{\mathcal{W}}_{t,p}$ and the notation (2.17) for the first and (D.30) for the second equality,

$$
\widehat{\mathcal{W}}_{t,p}(\hat{\mu}(w),\hat{\mu}(\tilde{w})) = \mathcal{W}_{t,p}(\hat{\mu}(w)_{[0,t]},\hat{\mu}(\tilde{w})_{[0,t]}) = \mathcal{W}_{t,p}\Big(\frac{1}{N}\sum_{i=1}^N \delta_{w^i_{[0,t]}}, \frac{1}{N}\sum_{i=1}^N \delta_{\tilde{w}^i_{[0,t]}}\Big).
$$

Combining the previous display with the inequality (D.3) from Example D.3 and using that by Definition 2.5, $\mathcal{W}_{t,p}$ is the p-th order Wasserstein metric on (W_n^t, p_t),

$$
\widehat{\mathcal{W}}_{t,p}(\hat{\mu}(w),\hat{\mu}(\tilde{w})) \le \Big(\frac{1}{N}\sum_{i=1}^N p_t^p(w^i - \tilde{w}^i)\Big)^{1/p} \le N^{-1/p}\sum_{i=1}^N p_t(w^i - \tilde{w}^i), \tag{2.103}
$$

where the last inequality follows from subadditivity of $\mathbb{R}_+ \ni x \mapsto x^{1/p}$. For each $s \in \mathbb{R}_+$, let $\hat{p}_s$ denote the usual seminorm on W_{nN}, see (2.1). For all $\bar{w} = (\bar{w}^1,\dots,\bar{w}^N) \in W_{nN}$ and $i \in \{1,\dots,N\}$, we have $p_t(\bar{w}^i) \le \hat{p}_t(\bar{w})$, and thus[36]

$$
\sum_{i=1}^N p_t(\bar{w}^i) \le N\hat{p}_t(\bar{w}).
$$

Combining (2.102), (2.103) and the previous display, we obtain

$$
|b^N(t,w) - b^N(t,\tilde{w})| + |\sigma^N(t,w) - \sigma^N(t,\tilde{w})| \le c_t(N + N^{2-1/p})\hat{p}_t(w - \tilde{w}). \tag{2.104}
$$

Hence b^N and σ^N have the Lipschitz property (C.15) of coefficients of classical SDEs. Similarly, using (2.99), once more subadditivity of the square root to bound the

[36] This bound is sharp e.g. for all $\bar{w} \in W_{nN}$ where the components $(\bar{w}^i)_{i=1,\dots,N}$ are supported on disjoint subsets $I_1,\dots,I_N$ of $[0,t]$ such that $p_t(\bar{w}^1) = \cdots = p_t(\bar{w}^N)$.

Euclidean (resp. Frobenius) norm by Euclidean (resp. Frobenius) norms of components, and the fact that b and σ are of linearly bounded growth (Definition 2.20), we have

$$\begin{aligned}|b^N(t,w)| + |\sigma^N(t,w)| &\le \sum_{i=1}^N |b(t,w^i,\hat{\mu}(w))| + |\sigma(t,w^i,\hat{\mu}(w))| \\ &\le \sum_{i=1}^N D_t(1+p_t(w^i)) \le ND_t(1+\hat{p}_t(w)).\end{aligned}$$

Hence b^N and σ^N are of linearly bounded growth in the sense of classical SDEs (Definition C.27).

(c) By parts (a) and (b) above, b^N and σ^N are $\mathbb{G}^{nN}$-progressive, have the Lipschitz property and are of linearly bounded growth in the sense of classical SDEs. Consequently, by Theorem C.34, there is pathwise uniqueness and strong existence of solutions of the particle approximation (2.96). Also by Theorem C.34, using that the initial conditions $\xi^1,\dots,\xi^N$ are by Assumption 2.49 p-integrable, the seminorm $p_t(X^{i,N})$ is p-integrable for all $t \in \mathbb{R}_+$ and all $i \in \{1,\dots,N\}$.

Remark 2.54 (Usefulness of the notion of $\mathbb{G}\otimes\mathbb{H}$-progressive measurability)**.** In Remark 2.53, the results about the measurable structure on $\widehat{\mathcal{P}}_p(W_n)$ from Section 2.1 are helpful. Without them, it would be unclear how to make sense of (2.96). Furthermore, without ensuring through the requirement of $\mathbb{G}\otimes\mathbb{H}$-progressive measurability that the coefficients b and σ in (2.95) are nonanticipative in the measure argument, see Remark 2.17, nonanticipativity (i.e. $\mathbb{G}$-progressive measurability) of the coefficients b^N and σ^N in the particle approximation (2.98) would be in question.

According to Theorem 2.56 below, as $N\to\infty$, the solutions $(X^{i,N})_{i\in\{1,\dots,N\}}$ of the particle approximation (2.96) converge in an L^p-sense to the i.i.d. solutions $(X^i)_{i\in\{1,\dots,N\}}$ of the identical McKean SDEs from (2.95). To prove it, we will use that $((X_i,X^{i,N}))_{i=1,\dots,N}$ is exchangeable, which means that its law is symmetric in the sense of Definition 2.47. That is,

$$\mathcal{L}_{\mathbb{P}}\big[((X_i,X^{i,N}))_{i=1,\dots,N}\big] = \mathcal{L}_{\mathbb{P}}\big[((X_{\varrho(i)},X^{(\varrho(i)),N}))_{i=1,\dots,N}\big]$$

for every bijection ϱ of $\{1,\dots,N\}$.

Lemma 2.55. *The family $((X_i,X^{i,N}))_{i=1,\dots,N}$ is exchangeable.*[37]

Proof. By Assumption 2.49 and Remark 2.53(a), see also the discussion preceding (2.70), $((X_i,X^{i,N}))_{i=1,\dots,N}$ solves the system of classical SDEs

$$\mathrm{d}X^i_t = b(t,X^i,\mathcal{L}[X^i])\,\mathrm{d}t + \sigma(t,X^i,\mathcal{L}[X^i])\,\mathrm{d}B^i_t\,, \qquad X^i_0 \overset{\text{a.s.}}{=} \xi^i, \tag{2.105}$$

$$\mathrm{d}X^{i,N}_t = b^{i,N}(t,X^N)\,\mathrm{d}t + \sigma^{i,N}(t,X^N)\,\mathrm{d}B^i_t\,, \qquad X^{i,N}_0 \overset{\text{a.s.}}{=} \xi^i, \tag{2.106}$$

for $i=1,\dots,N$. Fix a bijection ϱ of $\{1,\dots,N\}$ and let $w^\varrho := (w^{\varrho(1)},\dots,w^{\varrho(N)})$ for each $w=(w^1,\dots,w^N)\in W_{nN}$. The aim is to show that $((X_{\varrho(i)},X^{(\varrho(i)),N}))_{i=1,\dots,N}$ solves the same

[37] This fact is stated without proof in [10, §1.3.4] and [33, §1.1].

system (2.105) and (2.106) of SDEs as $((X_i, X^{i,N}))_{i=1,\dots,N}$ when the $\mathbb{K}^{dN}$-valued standard Brownian motion $B^{(N)} = (B^i)_{i=1,\dots,N}$ is replaced by the $\mathbb{K}^{dN}$-valued standard Brownian motion $B^{(N),\varrho} := (B^{\varrho(i)})_{i=1,\dots,N}$ and then conclude by using that there is uniqueness in law for this system of SDEs.

Since the processes $(X^i)_{i\in\mathbb{N}}$ are i.i.d. by Remark 2.50, we can replace each $\mathcal{L}[X^i]$ in (2.105) by $\mathcal{L}[X^1]$. Then,

$$\begin{aligned} b(t, X^{\varrho(i)}, \mathcal{L}[X^{\varrho(i)}]) &= b(t, (X^{\varrho})^i, \mathcal{L}[X^1]), \quad \text{and} \\ \sigma(t, X^{\varrho(i)}, \mathcal{L}[X^{\varrho(i)}]) &= \sigma(t, (X^{\varrho})^i, \mathcal{L}[X^1]), \end{aligned}$$

for all $i \in \{1, \dots, N\}$, hence the first components $(X_{\varrho(i)})_{i=1,\dots,N}$ of $((X_{\varrho(i)}, X^{(\varrho(i)),N}))_{i=1,\dots,N}$ satisfy (2.105) w.r.t. the Brownian motion $B^{(N),\varrho}$. Furthermore, for each $i \in \{1, \dots, N\}$ and all $w \in W_{nN}$, using (2.100) and the definition (D.28) of $\hat{\mu}$,

$$\begin{aligned} b^{\varrho(i),N}(t, w) &= b(t, w^{\varrho(i)}, \hat{\mu}(w)) = b(t, (w^{\varrho})^i, \hat{\mu}(w^{\varrho})) = b^{i,N}(t, w^{\varrho}), \\ \sigma^{\varrho(i),N}(t, w) &= \sigma(t, w^{\varrho(i)}, \hat{\mu}(w)) = \sigma(t, (w^{\varrho})^i, \hat{\mu}(w^{\varrho})) = \sigma^{i,N}(t, w^{\varrho}). \end{aligned}$$

Hence the second components $(X^{(\varrho(i)),N})_{i=1,\dots,N}$ of $((X_{\varrho(i)}, X^{(\varrho(i)),N}))_{i=1,\dots,N}$ satisfy (2.106) w.r.t. the Brownian motion $B^{(N),\varrho}$. Consequently, the family $((X^{\varrho(i)}, X^{\varrho(i),N}))_{i=1,\dots,N}$ is a solution of the system (2.105), (2.106) of SDEs. Since b and σ are $\mathbb{G}\otimes\mathbb{H}$-progressive, have the Lipschitz property and are of linearly bounded growth (Definitions 2.19 and 2.20) by Assumption 2.49, the same is true for the coefficients of this system by arguments essentially analogous to those in Remark 2.53(a). By Theorem 2.32, there is uniqueness in law of solutions of this system, which proves the claim. □

The following theorem extends [10, Theorem 1.10] and [33, Theorem 3] (and their proofs) to coefficients with pathwise dependence.

Theorem 2.56 (Convergence of the particle approximation for McKean SDEs)**.** *Let b and σ be coefficients and $(X^i)_{i\in\mathbb{N}}$ an associated family of independent solutions of the McKean SDE (2.95) as defined in Assumption 2.49. Let $(X^N)_{N\in\mathbb{N}}$ be strong solutions of the associated particle approximations (Definition 2.51), see also Remark 2.53(c). Then for all $t \in \mathbb{R}_+$,*

$$\lim_{N\to\infty} \sup_{1\le i\le N} \mathbb{E}[p_t^p(X^{i,N} - X^i)] = 0. \tag{2.107}$$

Proof. Fix for now $N \in \mathbb{N}$ and define

$$h_N(t) = \mathbb{E}[p_t^p(X^{1,N} - X^1)], \qquad t \in \mathbb{R}_+, \tag{2.108}$$

which is increasing and finite for all $t \in \mathbb{R}_+$ by p-integrability of $p_t(X^{1,N})$ and $p_t(X^1)$ (see Remarks 2.50(c) and 2.53(c)). Since $((X_i, X^{i,N}))_{i=1,\dots,n}$ is exchangeable by Lemma 2.55,

$$\sup_{1\le i\le N} \mathbb{E}[p_t^p(X^{i,N} - X^i)] = \mathbb{E}[p_t^p(X^{1,N} - X^1)] = h_N(t), \qquad t \in \mathbb{R}_+. \tag{2.109}$$

In analogy to (2.97), denote by

$$\mu^N := \frac{1}{N}\sum_{j=1}^{N} \delta_{X^j} \tag{2.110}$$

the empirical measure associated with the independent solutions of (2.95), and define

$$\tilde{X}^1 = \xi^1 + \int_0^\cdot b(s, X^1, \mu^N)\,\mathrm{d}s + \int_0^\cdot \sigma(s, X^1, \mu^N)\,\mathrm{d}B_s^1\,. \tag{2.111}$$

Note that the stochastic integrals in the previous display exist since the integrands are $\mathbb{F}$-progressive by Lemmas D.18 and C.7 (b) and moreover pathwise locally bounded by continuity of X^1 and the bounded growth assumption on b and σ. By the triangle inequality for the seminorm p_t and Jensen's inequality for $\mathbb{R}_+ \ni x \mapsto x^p$ (see (C.18)), we have

$$h_N(t) \le 2^{p-1}\big(\mathbb{E}[p_t^p(X^{1,N} - \tilde{X}^1)] + \mathbb{E}[p_t^p(\tilde{X}^1 - X^1)]\big). \tag{2.112}$$

Recall that $X^{1,N}$ satisfies the SDE (2.96) (for $i = 1$) with initial condition $X_0^{1,N} \overset{\text{a.s.}}{=} \xi^1$. From (2.97) and Lemma D.17 (b), the empirical measure $\tilde{\mu}^N$ associated with $(X^{i,N})_{i\in\{1,\ldots,N\}}$ appearing in (2.96) is $\mathcal{F}_t$-$\mathcal{H}_t$-measurable for each $t \in \mathbb{R}_+$. Hence Lemma 2.31 (a) applied[38] with $\tilde{X} := X^{1,N}$ and $\tilde{Y} := \tilde{X}^1$ yields

$$\mathbb{E}[p_t^p(X^{1,N} - \tilde{X}^1)] \le \beta(t) \int_0^t c_s^p \Big(\underbrace{\mathbb{E}[p_s^p(X^{1,N} - X^1)]}_{=h_N(s)} + \mathbb{E}\big[\widehat{\mathcal{W}}_{s,p}^p(\tilde{\mu}^N, \mu^N)\big], \tag{2.113}$$

with β defined in (C.22). With the map $\hat{\mu}\colon W_{nN} \to \widehat{\mathcal{P}}_p(W_n)$ from (D.28), we have, using (2.97), $\tilde{\mu}^N = \hat{\mu} \circ (X^{1,N}, \ldots, X^{N,N})$, and similarly from (2.110), $\mu^N = \hat{\mu} \circ (X^1, \ldots, X^N)$. With (2.103) from Remark 2.53 (b), it follows that

$$\mathbb{E}\Big[\widehat{\mathcal{W}}_{s,p}^p(\tilde{\mu}^N, \mu^N)\Big] \le \mathbb{E}\Bigg[\frac{1}{N}\sum_{j=1}^N p_s^p(X^{j,N} - X^j)\Bigg] = \mathbb{E}[p_s^p(X^{1,N} - X^1)] = h_N(s), \tag{2.114}$$

where we have used Lemma 2.55 for the first equality. From (2.113) and (2.114),

$$\mathbb{E}[p_t^p(X^{1,N} - \tilde{X}^1)] \le 2\beta(t) \int_0^t c_s^p\, h_N(s)\,\mathrm{d}s \tag{2.115}$$

To bound the second term on the right-hand side in (2.112), we use again Lemma 2.31 (a), this time with $\tilde{X} := \tilde{X}^1$ and $\tilde{Y} := X^1$. Recall that by construction, $\tilde{Y}$ satisfies the McKean SDE (2.95) with initial condition $\tilde{Y}_0 \overset{\text{a.s.}}{=} \xi^1$. Define[39]

$$\gamma_s^N(\omega) = \begin{cases} d_p(\mu_s^N(\omega), \mathcal{L}[X_s^1]) & \text{if } b \text{ and } \sigma \text{ have the inst. Lipschitz property (2.51)},\\ \widehat{\mathcal{W}}_{s,p}(\mu^N(\omega), \mathcal{L}[X^1]) & \text{otherwise}\ , \end{cases} \tag{2.116}$$

for all $N \in \mathbb{N}$, $s \in \mathbb{R}_+$ and $\omega \in \Omega$. We obtain using (2.58)

$$\mathbb{E}[p_t^p(\tilde{X}^1 - X^1)] \le \beta(t) \int_0^t c_s^p\, \mathbb{E}[(\gamma_s^N)^p]. \tag{2.117}$$

[38] Here, we use only case (i) of Lemma 2.31 (a).

[39] We consider separately the case where the coefficients b and σ have not only the Lipschitz property (Definition 2.19) but also the stronger instantaneous Lipschitz property (Definition 2.22), because the bound (2.120) obtained below will help us obtain a sharper rate of convergence (see Theorem 2.58) in the latter case.

Combining the bounds (2.115) and (2.117),

$$h_N(t) \le \underbrace{2^p\beta(t)\int_0^t c_s^p\,\mathbb{E}\big[(\gamma_s^N)^p\big]\,\mathrm{d}s}_{=:\tilde\alpha(t)} + \underbrace{2^p\beta(t)}_{=:\tilde\beta(t)}\int_0^t c_s^p\,h_N(s)\,\mathrm{d}s\,. \tag{2.118}$$

Since h_N is $\mathbb{R}_+$-valued and increasing, it is locally integrable against the measure μ on $(\mathbb{R}_+,\mathcal{B}_{\mathbb{R}_+})$ with density c^p w.r.t. the Lebesgue measure. In addition, (2.118) shows that h satisfies the integral inequality needed for the application of the extended Grönwall inequality [53, Theorem 8.21] with $I=\mathbb{R}_+$,

$$\tilde\beta(t) = 2^p\beta(t) \qquad \text{and} \qquad \tilde\alpha(t) = \tilde\beta(t)\int_0^t c_s^p\,\mathbb{E}[(\gamma_s^N)^p]\,\mathrm{d}s\,, \qquad t\in\mathbb{R}_+. \tag{2.119}$$

Since $\tilde\alpha$ is increasing and both $\tilde\beta$ and the density c^p of μ are bounded on each interval $[0,t]$, $t\in\mathbb{R}_+$, by $\tilde\beta(t)\vee c_t^p$, we have by [53, Remark 8.23 (c)]

$$\begin{aligned} h_N(t) &\le \tilde\beta(t)\exp\big(t(\tilde\beta(t)\vee c_t^p)^2\big)\int_0^t c_s^p\,\mathbb{E}\left[(\gamma_s^N)^p\right]\mathrm{d}s \\ &\le \tilde\beta(t)\exp\big(t(\tilde\beta(t)\vee c_t^p)^2\big)\,tc_t^p\,\mathbb{E}\Big[\widehat{\mathcal{W}}_{t,p}^p(\mu^N,\mathcal{L}[X^1])\Big], \end{aligned} \tag{2.120}$$

where the last inequality follows from the monotonicity of $s\mapsto\widehat{\mathcal{W}}_{s,p}$, see Remark 2.6 (e), and the inequality (2.25). By definition (see (2.109)), μ^N is the empirical measure of the i.i.d. solutions $X^1,\dots,X^N$ of (2.95), see also Remark 2.50. Using (2.21), the fact that $\mu^N_{[0,t]}=\frac{1}{N}\sum_{i=1}^N\delta_{X^i\restriction_{[0,t]}}$ by (D.30) and finally (D.8) from Theorem D.9,

$$\mathbb{E}\big[\widehat{\mathcal{W}}_{t,p}^p(\mu^N,\mathcal{L}[X^1])\big] = \mathbb{E}\big[\mathcal{W}_{t,p}^p(\mu^N_{[0,t]},\mathcal{L}[X^1]_{[0,t]})\big] \xrightarrow[N\to\infty]{} 0. \tag{2.121}$$

From (2.120) and (2.121), $h_N(t)\xrightarrow[N\to\infty]{}0$, for each $t\in\mathbb{R}_+$. The claim now follows from (2.109). □

Remark 2.57 (Rate of convergence of the particle approximation)**.** By (2.120), every bound for the rate of convergence of $\mathbb{E}[(\gamma_s^N)^p]$ to zero[40] yields a bound for rate of convergence in (2.107) and thus a bound for the convergence of the particle approximation. For coefficients b and σ with the instantaneous Lipschitz property (Definition 2.22), Theorems D.10 and D.11 provide the required rates of convergence.[41] This leads to Theorem 2.58 below, which extends [33, Theorem 3], where a similar result is stated for the case $\mathbb{K}=\mathbb{R}$, a finite time horizon, $p=2$ and with a less explicit bound. For a recent related result regarding the convergence of an interpolated Euler scheme for the numerical solution of McKean SDEs depending on the past trajectory and the curve of past one-dimensional marginal distributions of the solution, see [3, Theorem 1.7].

[40] That $\lim_{N\to\infty}\mathbb{E}[(\gamma_s^N)^p]=0$ follows as in the proof of Theorem 2.56 from Theorem D.9.

[41] We have so far found no sufficiently general bounds for the rate of convergence in Theorem D.9 when the underlying Polish space is $C([0,t],\mathbb{K}^n)$.

Theorem 2.58 (Rate of convergence of the particle approximation for coefficients with the instantaneous Lipschitz property)**.** *Consider the setting of Theorem 2.56. Assume that the coefficients b and σ have the instantaneous Lipschitz property (Definition 2.22).*

(a) Assume that the initial random vector ξ^1 of X^1 (see (2.95)) satisfies

$$\mathbb{E}[|\xi^1|^q] < \infty \tag{2.122}$$

for some $q > p$ satisfying one of the cases in (D.14). For all $N \in \mathbb{N}$ and $t \in \mathbb{R}_+$,

$$\sup_{1\le i\le N} \mathbb{E}[p_t^p(X^{i,N} - X^i)] \le C L_t \alpha_{d_{\mathbb{K}},p,q}(N), \tag{2.123}$$

where the constant $C \in \mathbb{R}_+$ and the function $\alpha_{d_{\mathbb{K}},p,q} \colon \mathbb{N} \to \mathbb{R}_+$ come from Theorem D.10 and

$$L_t := \tilde{\beta}(t) \exp\bigl(t(\tilde{\beta}(t) \vee c_t^p)^2\bigr) t\, c_t^p \biggl(\|\xi^1\|_{L^q} + \sum_{k=1}^{\infty} \Bigl(\frac{(K_t t)^k}{k!} \Bigr)^{1/q} \biggr)^p, \tag{2.124}$$

with $\tilde{\beta} \colon \mathbb{R}_+ \to \mathbb{R}_+$ is defined in (2.119) and (C.22), the function $c \colon \mathbb{R}_+ \to \mathbb{R}_+$ is from Definition 2.22 of the instantaneous Lipschitz property, and the function $K \colon \mathbb{R}_+ \to \mathbb{R}_+$ originates from (2.82) (see also (C.42), replacing ξ with ξ^1).

(b) Assume that (2.122) holds for some $q > p$ satisfying one of the cases in (D.20). Then for all $N \in \mathbb{N}$ and $t \in \mathbb{R}_+$,

$$\sup_{1\le i\le N} \mathbb{E}[p_t^p(X^{i,N} - X^i)] \le L_t \tilde{\alpha}_{d_{\mathbb{K}},p,q}(N), \tag{2.125}$$

with L as in (2.124) and the function $\tilde{\alpha}_{d_{\mathbb{K}},p,q}$ from Theorem D.11.

Proof. The proof incorporates ideas from [33, Theorem 3], [23, §7.3] and [22, Theorem 1].

(a) Fix $N \in \mathbb{N}$ and $t \in \mathbb{R}_+$. By (2.109), (2.120) and (2.116),

$$\sup_{1\le i\le N} \mathbb{E}[p_t^p(X^{i,N} - X^i)] \le \tilde{\beta}(t) \exp\bigl(t(\tilde{\beta}(t) \vee c_t^p)^2\bigr) \int_0^t c_s^p\, \mathbb{E}\left[d_p^p(\mu_s^N, \mathcal{L}[X_s^1])\right]. \tag{2.126}$$

For all $s \in [0,t]$, we have using (D.13) from Theorem D.10 and the inequality $|X_s^1| \le p_t(X^1)$,

$$\mathbb{E}\left[d_p^p(\mu_s^N, \mathcal{L}[X_s^1])\right] \le C\, \|X_s^1\|_{L^q(\mathbb{P})}^p \alpha_{d_{\mathbb{K}},p,q}(N) \le C\, \|p_t(X^1)\|_{L^q(\mathbb{P})}^p\, \alpha_{d_{\mathbb{K}},p,q}(N). \tag{2.127}$$

Applying the triangle inequality once for the seminorm p_t and once for $\|\cdot\|_{L^q}(\mathbb{P})$ together with the moment bound (2.82) from Remark 2.36,[42]

$$\|p_t(X^1)\|_{L^q(\mathbb{P})} \le \|\xi^1\|_{L^q(\mathbb{P})} + \|p_t(X^1 - \xi^1)\|_{L^q(\mathbb{P})} \le \|\xi^1\|_{L^q(\mathbb{P})} + \sum_{k=1}^{\infty} \Bigl(\frac{(K_t t)^k}{k!} \Bigr)^{1/q}. \tag{2.128}$$

[42] The application of Remark 2.36 is valid, since the Lipschitz property (Definition 2.19) for the Wasserstein order p implies the Lipschitz property for all Wasserstein orders $q \in [p, \infty)$ by Remark 2.23(a).

From (2.126) and (2.127) for the first, (2.128) for the second and the definition (2.124) of L for the third line,

$$\begin{aligned}\sup_{1\le i\le N} \mathbb{E}[p_t^p(X^{i,N}-X^i)] &\le \tilde\beta(t)\exp\big(t(\tilde\beta(t)\vee c_t^p)^2\big)tc_t^p C\,\|p_t(X^1)\|^p_{L^q(\mathbb{P})}\,\alpha_{d_{\mathbb{K}},p,q}(N)\\ &= C\tilde\beta(t)\exp\big(t(\tilde\beta(t)\vee c_t^p)^2\big)t\,c_t^p\left(\|\xi^1\|_{L^q(\mathbb{P})}+\sum_{k=1}^{\infty}\Big(\frac{(K_t t)^k}{k!}\Big)^{1/q}\right)^p\alpha_{d_{\mathbb{K}},p,q}(N)\\ &= CL_t\alpha_{d_{\mathbb{K}},p,q}(N).\end{aligned}$$ □

(b) By assumption, p and q are such that (2.122) holds for one of the cases in (D.20). The proof now proceeds exactly as in part (a), but instead of (2.127), we obtain using (D.19) from Theorem D.11 the bound

$$\mathbb{E}\left[d_p^p(\mu_s^N,\mathcal{L}[X_s^1])\right] \le \|X_s^1\|^p_{L^q(\mathbb{P})}\tilde\alpha_{d_{\mathbb{K}},p,q}(N) \le \|p_t(X^1)\|^p_{L^q(\mathbb{P})}\,\tilde\alpha_{d_{\mathbb{K}},p,q}(N). \tag{2.129}$$

Proceeding as before concludes the proof.

Remark 2.59 (Quality of the rate of convergence in Theorem 2.58)**.** While the constants appearing in the function $\tilde\alpha_{d_{\mathbb{K}},p,q}(N)$ from (2.125) are not overly large for reasonable parameter combinations (see Table D.1), the function L defined in (2.124) grows rapidly. This shown in Figure 2.1 for $d=1$, $(p,q)\in\{(2,5),(3,7),(4,9)\}$, $C_p := 4$, $\|\xi^1\|_{L^p}=1$, $\|\xi^1\|_{L^q}=2$, and assuming[43] that $c_t\vee D_t\le 1$ for all $t\le 1$.

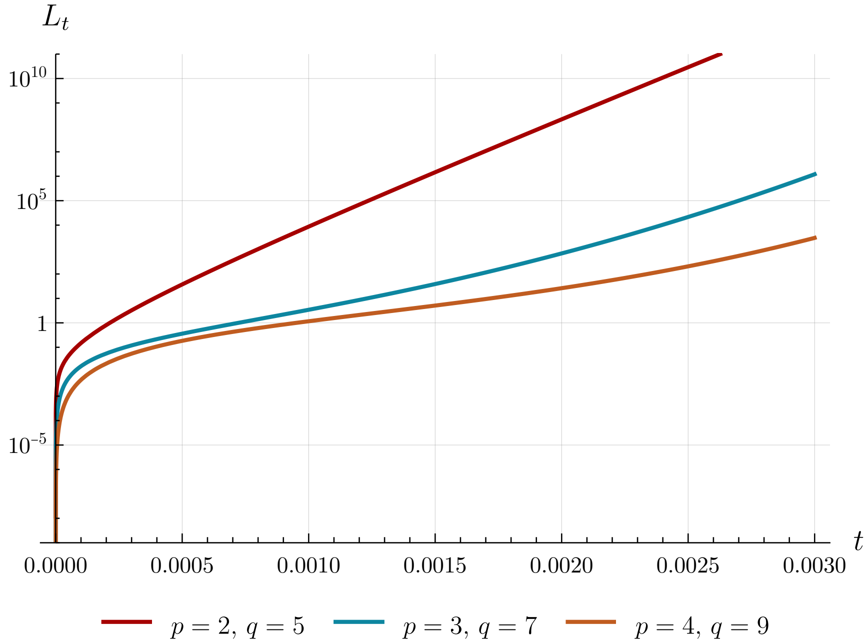

Figure 2.1. The function L defined in (2.124), for $d=1$, $(p,q)\in\{(2,5),(3,7),(4,9)\}$, $C_p=4$, $\|\xi^1\|_{L^p}=1$, $\|\xi^1\|_{L^q}=2$, and assuming that $c_t\vee D_t\le 1$ for all $t\le 1$.

[43] Note that C_p in (2.124) (see also (2.119) and (C.22)) originates from the Burkholder–Davis–Gundy inequality [53, Theorem 6.32]. If $p=2$, we may assume that $C_2=4$, but different values are appropriate when $p\neq 2$.

The next theorem establishes, as a corollary of Theorem 2.56, a classical propagation of chaos result, which states that under the conditions of Assumption 2.49, the sequence (of restrictions to finite time intervals) of probability distributions associated with the solutions X^N, $N \in \mathbb{N}$, of the particle approximation (Definition 2.51) is chaotic (Definition 2.45) w.r.t. the corresponding probability distribution of the solution of the McKean SDE (2.36). Theorem 2.60 (though not its proof) extends [26, Theorem 10.3], where a pointwise[44] propagation of chaos result is proved.

Theorem 2.60 (Pathwise propagation of chaos for the McKean SDE, [26, Theorem 10.3])**.** *Let b and σ be coefficients of the McKean SDE (2.36) as defined in Assumption 2.49 and $(X^i)_{i\in\mathbb{N}}$ the associated family of i.i.d. solutions, see also Remark 2.50. For each $N \in \mathbb{N}$, let $X^{(N)} = (X^{1,N}, \dots, X^{N,N})$ denote[45] the solution of the N-dimensional particle approximation SDE (2.96) from Definition 2.51. Define*

$$\mu = \mathcal{L}[X^1] \qquad \textit{and} \qquad \mu^N = \mathcal{L}[X^{(N)}], \quad N \in \mathbb{N},$$

fix $t \in \mathbb{R}_+$ and denote by $\mu_{[0,t]}$ and $\mu^N_{[0,t]}$ the distributions of the restrictions of the respective processes to the time interval $[0,t]$ (see (2.17)). Then $(\mu^N_{[0,t]})_{N\in\mathbb{N}}$ is $\mu_{[0,t]}$-chaotic (Definition 2.45).

Proof. By Lemma 2.46, it suffices to show that for each $k \in \mathbb{N}$, the sequence $(\mu^N_{[0,t]} \circ (\pi^N_{[1,k]})^{-1})_{N\geq k}$ of k-dimensional marginal distributions (see (2.92)) converges weakly to $\mu^{\otimes k}_{[0,t]}$ as $N \to \infty$. Fix $k \in \mathbb{N}$. By the definition of μ^N and Remark 2.50,

$$\mu^N_{[0,t]} \circ (\pi^N_{[1,k]})^{-1} = \mathcal{L}[X^{1,N}{\restriction_{[0,t]}}, \dots, X^{k,N}{\restriction_{[0,t]}}] \quad \text{and} \quad \mu^{\otimes k}_{[0,t]} = \mathcal{L}[X^1{\restriction_{[0,t]}}, \dots, X^k{\restriction_{[0,t]}}].$$

From (2.107) in Theorem 2.56, we have

$$\lim_{k\leq N\to\infty} \mathbb{E}\left[\sum_{i=1}^k p_t^2(X^{i,N} - X^i)\right] = 0. \tag{2.130}$$

Let $\tilde{p}_t \colon W^t_{kn} \to \mathbb{R}_+$ denote the norm of uniform convergence on W^t_{kn} defined in analogy to (2.7). For all $w = (w^1, \dots, w^k) \in W^t_{kn}$ and all $s \in [0,t]$, we have $|w_s| \leq \sum_{i=1}^k |w^i_s|$ by subadditivity of the square root and thus

$$\begin{aligned} \tilde{p}_t(w) &= \sup_{s\in[0,t]} |w_s| \leq \sup_{s\in[0,t]} \sum_{i=1}^k |w^i_s| \\ &\leq \sum_{i=1}^k \sup_{s\in[0,t]} |w^i_s| = \sum_{i=1}^k p'_t(w^i), \end{aligned}$$

[44] By pointwise propagation of chaos, we mean the analogue of Theorem 2.60 with $\mu_{[0,t]}$ replaced by the one-dimensional marginal distribution μ_t and, correspondingly, $\mu^N_{[0,t]}$ replaced for each $N \in \mathbb{N}$ by μ^N_t, see also (2.17). This terminology is from [11, Definition 3.24].

[45] We switch to the notation $X^{(N)}$ compared to X^N in Definition 2.51 to avoid the ambiguity described in Footnote 35.

with p_t' defined in (2.7). From the previous display, Jensen's inequality for $\mathbb{R}_+ \ni x \mapsto x^2$ (see (C.18)) and (2.130), it follows that

$$\mathbb{E}\big[\tilde{p}_t^2\big((X^{1,N}\restriction_{[0,t]}, \dots, X^{k,N}\restriction_{[0,t]}) - (X^1\restriction_{[0,t]}, \dots, X^k\restriction_{[0,t]})\big)\big] \xrightarrow[k \leq N \to \infty]{} 0.$$

Thus, by Markov's inequality for random vectors ([53, Lemma 16.1 and Remark 16.2(c)]), $(X^{1,N}\restriction_{[0,t]}, \dots, X^{k,N}\restriction_{[0,t]})$ converges to $(X^1\restriction_{[0,t]}, \dots, X^k\restriction_{[0,t]})$ in probability in the sense that for every $\varepsilon > 0$,

$$\mathbb{P}[\tilde{p}_t((X^{1,N}\restriction_{[0,t]}, \dots, X^{k,N}\restriction_{[0,t]}) - (X^1\restriction_{[0,t]}, \dots, X^k\restriction_{[0,t]})) > \varepsilon] \xrightarrow[k \leq N \to \infty]{} 0.$$

Since convergence in probability implies weak convergence ([36, Corollary 13.19]), the claim follows. $\square$

3. Applications of McKean SDEs in Mathematical Finance

In this chapter, we present two applications of McKean SDEs to finance. The first (Section 3.1) is the calibration of the local stochastic volatility model to market smiles in the context of option pricing. The second (Section 3.2) is a mean-field LIBOR market model in the context of interest rate modelling.

For the remainder of this chapter, fix a time horizon $\overline{T} \in \mathbb{R}_+$. By the extension procedure described in Remark 2.30, we may consider the case of coefficients defined on a finite time horizon as a special case of coefficients defined on the infinite time horizon, i.e. the setting of Appendix C and Chapter 2.

3.1. Calibration of Local Stochastic Volatility Models

This section introduces local stochastic volatility models and a method for their calibration using an analogue of the particle approximation method for McKean SDEs from Subsection 2.4.2. The presentation is based mainly on [26, §11]. We reduce considerably the level of mathematical rigour upheld in the other parts of this thesis, focusing on the context and ideas behind the derivations and methods presented, at the expense of giving complete proofs. This means that several ad-hoc assumptions will be made (especially in the proof of Proposition 3.5) and not all steps will be justified completely.[1] We aim to highlight clearly all such departures from rigour.

In Subsection 3.1.1, we introduce the context and motivation behind the local stochastic volatility model (LSVM) beginning with two precursors, the local volatility and the stochastic volatility models. In Subsection 3.1.2, we describe the local stochastic volatility model and consider some examples. Subsections 3.1.3 and 3.1.4 treat the calibration of LSVMs to option prices observed in the market by means of a particle method. For simplicity, we will consider only the case of a single risky asset.

3.1.1. Background and Motivation for Local Stochastic Volatility Models

[2]The classical Black–Scholes model assumes the volatility of the risky asset and the risk-free interest rate to be constant. While this provides a parsimonious, analytically tractable mathematical model for option pricing, it fails to capture important empirically observed features of financial markets. In practice, both the volatility and interest rate can vary with time in a not necessarily deterministic fashion, as seen in Figure 3.1.

[1] In this, we follow our main source, [26]. We have not been able to find or produce a rigorous proof e.g. of Proposition 3.5.

[2] This section is loosely based on [1, §6 and §7] and [25, §1].

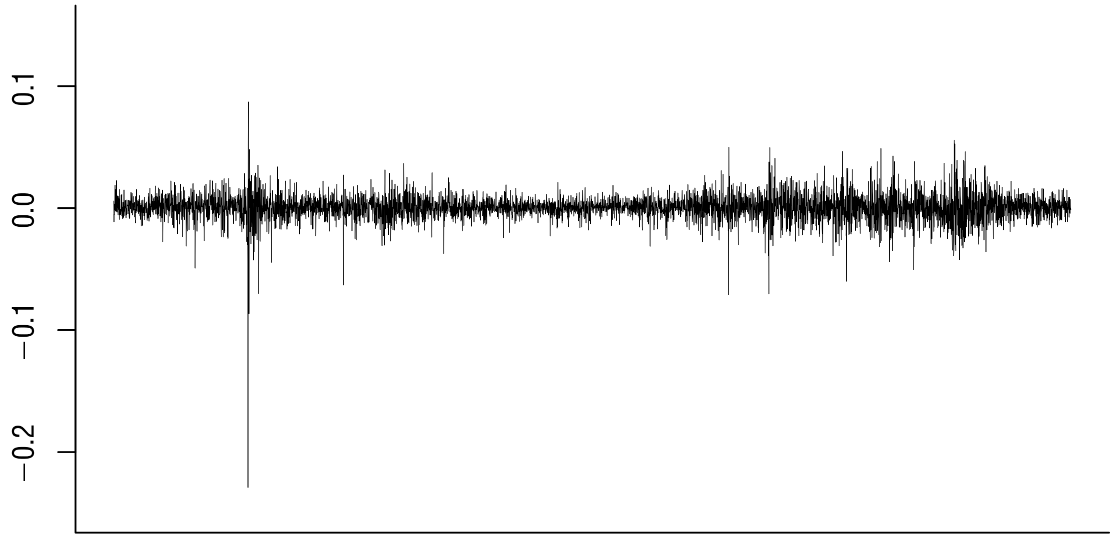

Figure 3.1. SP500 daily log returns from December 31, 1984, to December 31, 2004. We can observe volatility clustering (i.e. periods with high volatility tend to be followed by periods with high volatility) and mean reversion (i.e. high levels of volatility tend to dissipate eventually) [25, §1].

In the context of option pricing, the violation of the assumption of a constant volatility in the Black–Scholes model manifests in the smile or skew of the volatility surface, which is defined as the surface

$$[0,\bar{T}] \times \mathbb{R}_+ \ni (T,K) \mapsto \sigma_{\mathrm{impl}}(T,K) \in \mathbb{R}_+$$

of implied volatilities[3] of vanilla[4] options indexed by their maturities $T \in [0,\bar{T}]$ and strikes $K \in \mathbb{R}_+$.[5] If the volatility of a stock were constant throughout time, then the implied volatility $\sigma_{\mathrm{impl}}(T,K)$ would depend neither on the maturity nor on the strike price of the option. Instead, a considerable dependence is observed in practice. For example, for equities, a common case is that for a fixed maturity $T \in [0,\bar{T}]$, the implied volatility $\sigma_{\mathrm{impl}}(T,\,\cdot\,)$ is decreasing in the strike, with strikes far below the current stock price typically having the highest implied volatilities.[6] Figure 3.2 contains a plot of a typical volatility surface for the SP500. In foreign exchange markets, the shape of the curve $\sigma_{\mathrm{impl}}(T,\,\cdot\,)$ is often symmetrical around the at-the-money strike, and this phenomenon is called the volatility smile.[7]

[3] The implied volatility determined by the Black–Scholes model for a given price, strike and maturity of a call option is that volatility, for which the Black–Scholes pricing formula produces exactly the given price. Here, the risk-free rate is assumed to be constant.

[4] A vanilla option always refers to a call or a put. Options with more complicated payoffs are called exotic.

[5] The volatility smile or skew cannot be produced by a deterministic time-dependent volatility, since in this case all vanilla options with the same maturity are priced with the same volatility, hence there can be no dependence of the implied volatility on the strike price.

[6] This is economically plausible, since strike prices far out-of-the-money correspond to large fluctuations of the price of the underlying, especially for downward moves.

[7] This is also visible in Figure 3.2 for short maturities.

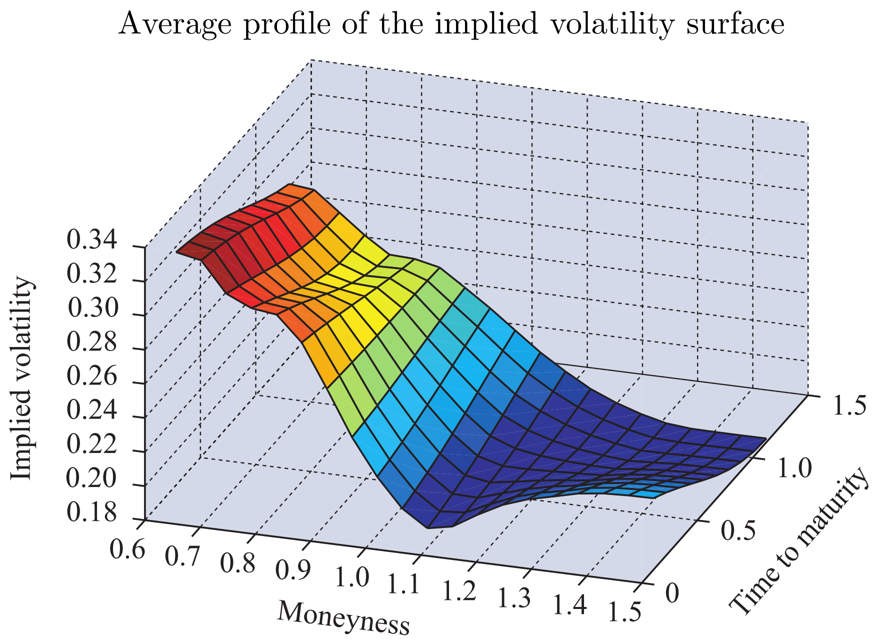

Figure 3.2. A typical profile of the implied volatility of vanilla options on the SP500 as a function of time to maturity and moneyness (defined as the quotient K/S of the strike price K and the current stock price S), for March 1999, [12, §2.3].

To deal with the shortcomings of the Black–Scholes model, several extensions have been developed. The *local volatility model*, introduced by Bruno Dupire in [20], replaces the constant volatility of the Black–Scholes model by a locally bounded measurable function

$$\sigma\colon [0,\overline{T}] \times \mathbb{R}_+ \to \mathbb{R}, \tag{3.1}$$

called the *local volatility*. The risky asset price process $S = (S_t)_{t\in[0,\overline{T}]}$ is then defined, w.r.t. a risk-neutral measure $\mathbb{Q}$, by the SDE

$$\mathrm{d}S_t = S_t(r_t\,\mathrm{d}t + \sigma(t,S_t)\,\mathrm{d}W_t), \qquad t \in [0,\overline{T}], \tag{3.2}$$

where the risk-free rate $r\colon [0,\overline{T}] \to \mathbb{R}_+$ is a measurable function of time. Dupire derived in [20] a calibration condition which is a special case of (3.13) below,[8] whereby the local volatility is expressed in terms of the partial derivatives of the surface

$$[0,\overline{T}] \times \mathbb{R}_+ \ni (T,K) \mapsto \mathcal{C}_{\mathbb{Q}}(T,K) \coloneqq \mathbb{E}_{\mathbb{Q}}\Big[\mathrm{e}^{-\int_0^T r_u \mathrm{d}u}(S_T-K)^+\Big]$$

of model implied call option prices. Then, to price and hedge exotic options, the idea is to start with the surface $\mathcal{C}_{\mathrm{mkt}}\colon [0,\overline{T}] \times \mathbb{R}_+ \to \mathbb{R}_+$ of call option prices observed[9] in the market and determine the local volatility by Dupire's calibration condition. This gives the law of

[8] Some of the ideas discussed here are described in more generality and detail below, see Proposition 3.5 and Remark 3.6.

[9] Since we can only ever observe finitely many call option prices, some form of interpolation is required to obtain a sufficiently differentiable surface, see Remark 3.6 for details.

the risky asset S under the risk-neutral measure $\mathbb{Q}$, which can in turn be used to price exotics by the usual risk-neutral pricing framework.[10] The attractive feature of the local volatility model obtained in this way is that it reproduces, by construction, exactly the observed market prices of call options over the complete range of strikes and maturities.[11] This also means that such a model reproduces, at least at the time of its construction, the volatility surface derived from market option prices.

Unfortunately, the local volatility model has some significant drawbacks. A priori, the dynamics specified by (3.2) are unrealistic, since the volatility depends exclusively on time and the stock price, i.e. there is no way to account for added randomness that is not connected to movements of the stock. A related inflexibility of the model is that there is no way to adjust the volatility of volatility. Moreover, the local volatility model produces an evolution of the volatility surface through time which does not match what is observed in practice. This can lead to delta hedging by means of the local volatility model being inferior even to delta hedging using the Black–Scholes model [27].

The next important class of models extending the Black–Scholes model are the *stochastic volatility models*. Here, in a rather general formulation, the SDE (3.2) is replaced by

$$\mathrm{d}S_t = S_t(r_t\,\mathrm{d}t + a_t\,\mathrm{d}W_t), \qquad t \in [0, \overline{T}], \tag{3.3}$$

where the *stochastic volatility* $a = (a_t)_{t\in[0,\overline{T}]}$ is an $\mathbb{R}$-valued Itô process. If the local volatility from (3.1) is such that Itô's formula is applicable to the process $(\sigma(t, S_t))_{t\in[0,\overline{T}]}$ (in particular if σ is $C^{1,2}$, i.e. continuously differentiable once in the time and twice in the space variable), then the stochastic volatility model contains the local volatility model as a special case. However, the stochastic volatility models found in the literature differ essentially from the local volatility model in that the stochastic volatility process a is most frequently specified by an SDE defined by a small number of parameters (often constants) and not involving the risky asset S. A popular choice is to take as the stochastic volatility a CIR process, see Example 3.3 (b) below.[12]

Like the local volatility model, the stochastic volatility model is able to produce a smile or skew of implied volatilities. However, as noted in [1, §7.8], the small number of parameters involved in common stochastic volatility models means that such models, unlike the local volatility model, cannot be calibrated exactly to the complete surface of vanilla option prices over the range of strikes and maturities observed in the market.

3.1.2. Local Stochastic Volatility Models with Random Interest Rates

The local stochastic volatility model (LSVM) combines the approaches of the local and stochastic volatility models by taking as the effective volatility of the risky asset the product of a local and a stochastic volatility term (see (3.5) below). Due to its ability to combine the strengths of the local and stochastic volatility models, it is called in [1, §9.4] the "triumph of

[10] This goes beyond the Black–Scholes setting, but the basic ideas underlying the risk-neutral pricing framework are similar. See [55, §5] for details.

[11] Under the condition that those prices can at all be explained by a local volatility model, see Remark 3.6.

[12] For more examples, see [29].

modern derivatives pricing". For added generality, we consider also the case of stochastic interest rates.

Definition 3.1 (Local stochastic volatility model)**.** A local stochastic volatility model over a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{Q})$ with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,\overline{T}]}$ satisfying the usual completeness assumption (Definition C.3) consists of

(a) an $\mathbb{R}$-valued $(\mathbb{F}, \mathbb{Q})$-standard Brownian motion $W = (W_t)_{t\in[0,\overline{T}]}$,

(b) an $\mathbb{R}$-valued Itô process $r = (r_t)_{t\in[0,\overline{T}]}$ modelling the short-term risk-free interest rate,

(c) an $\mathbb{R}$-valued Itô process $a = (a_t)_{t\in[0,\overline{T}]}$ called the stochastic volatility,

(d) a locally bounded measurable map as in (3.1), called the *local volatility*,

(e) a risk-free asset $S^0 = (S^0_t)_{t\in[0,\overline{T}]}$ with

$$S^0_t := \exp\left(\int_0^t r_u \,\mathrm{d}u\right), \qquad t \in [0, \overline{T}], \tag{3.4}$$

and

(f) a risky asset $S = (S_t)_{t\in[0,\overline{T}]}$ with a given initial value $S_0 \in (0, \infty)$, described by the SDE

$$\mathrm{d}S_t = S_t\big(r_t \,\mathrm{d}t + \sigma(t, S_t)a_t \,\mathrm{d}W_t\big), \qquad t \in [0, \overline{T}], \tag{3.5}$$

where the interest rate, stochastic volatility and local volatility are assumed to be such that there exists an integrable strong solution[13] of (3.5) with the added property that the discounted asset price process S/S^0 is an $(\mathbb{F}, \mathbb{Q})$-martingale.

Remark 3.2. (a) Local, stochastic and local stochastic volatility models are often discussed restricting to deterministic interest rates. As stated in [26, §11.1], an advantage of the particle method described in Subsection 3.1.4 below is that it can deal with stochastic interest rates efficiently compared to PDE methods, which suffer from the curse of dimensionality.

(b) The probability measure $\mathbb{Q}$ in Definition 3.1 is interpreted as the risk-neutral measure for a financial market, meaning that prices of derivatives can be obtained as (conditional) expectations of the associated discounted payoffs w.r.t. $\mathbb{Q}$. Consequently, the drift term of the risky asset is given by $r_t S_t \,\mathrm{d}t$.

The risk-free rate and stochastic volatility processes in Definition 3.1 are usually specified by SDEs. Two popular examples are the Hull–White and Cox–Ingersoll–Ross models, which we discuss next.

Example 3.3 (Models for the risk-free rate and stochastic volatility)**.** Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic base (Definition C.5) carrying an $\mathbb{R}$-valued standard $(\mathbb{F}, \mathbb{P})$-Brownian motion $W = (W_t)_{t\in[0,\overline{T}]}$.

[13] I.e. there exists a strong solution $S = (S_t)_{t\in[0,\overline{T}]}$ of (3.5) with $S_t \in \mathcal{L}^1(\mathbb{Q})$ for all $t \in [0, \overline{T}]$. Integrability is needed in (3.11) below.

(a) (Hull–White model, [31], [55, Example 6.5.1]) In the Hull–White model,[14] the risk-free interest rate $r = (r_t)_{t\in[0,\overline{T}]}$ follows the SDE

$$\mathrm{d}r_t = \kappa(t)\big(\mu(t) - r_t\big)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W_t\,, \tag{3.6}$$

where the initial value $r_0 \in \mathbb{R}_+$ is given and

$$\kappa\colon [0,\overline{T}] \to \mathbb{R}_+, \quad \mu\colon [0,\overline{T}] \to \mathbb{R}_+ \quad \text{and} \quad \sigma\colon [0,\overline{T}] \to \mathbb{R}$$

are measurable functions such that there is strong existence and pathwise uniqueness of solutions of (3.6). Note that for each $t \in [0,\overline{T}]$, the drift $\kappa(t)(\mu(t) - r_t)$ is, assuming that $\kappa(t) > 0$, strictly positive if $r_t < \mu(t)$ and strictly negative if $r_t > \mu(t)$. Hence r_t is pushed towards $\mu(t)$. In particular, if μ has a constant value μ_0, then r has the tendency to move towards this constant. This feature, which is called *mean reversion*, is desirable, since economic arguments suggest that interest rates and volatilities cannot remain at extreme levels indefinitely. The Hull–White model has several other desirable features, e.g. there is a closed-form expression for r (an Ornstein–Uhlenbeck process, see [53, Example 8.19]), and, in the context of interest rate models, the fair prices of zero-coupon bonds can be computed analytically, see [55, Example 6.5.1]. A potential disadvantage of the Hull–White model is that the process r can take on negative values.[15]

(b) (Cox–Ingersoll–Ross model, [55, Example 6.5.2], [14] (for interest rate modelling) and [28] (for stochastic volatility)) In the Cox–Ingersoll–Ross (CIR) model, the process Y, which can be either a risk-free interest rate or a stochastic volatility process, follows the SDE

$$\mathrm{d}Y_t = \kappa\big(\mu - Y_t\big)\,\mathrm{d}t + \sigma\sqrt{Y_t}\,\mathrm{d}W_t\,, \tag{3.7}$$

where the initial value $Y_0 \in \mathbb{R}_+$ is given and κ, μ and σ are nonnegative constants. Compared to the Hull–White model, the drift is unchanged and has the same interpretation in terms of mean reversion. The difference is in the volatility factor $\sigma\sqrt{Y_t}\,\mathrm{d}W_t$, where the factor $\sqrt{Y_t}$ now prevents the process from taking negative values (see [32, §6.3.1]), thus remedying a disadvantage of the Hull–White model. In the context of interest rate models, the fair prices of zero-coupon bonds can be computed analytically, as in the case of the Hull–White model (see [55, Example 6.5.2]).

Note that in the Hull–White model, r appears linearly on the right-hand side of the SDE (3.6), hence strong existence and pathwise uniqueness follow from Theorem C.34 as soon as the functions κ, μ and σ are locally bounded.[16] For the CIR model, the square root in (3.7) means that the volatility coefficient of this SDE does not have the Lipschitz property (Definition C.25). Despite this, pathwise uniqueness and strong existence can be obtained from results specifically for one-dimensional SDEs, see [32, Theorem 1.5.5.1].

[14] If the functions κ, μ and σ in (3.6) are constant, the Hull–White model is also called the Vasicek model.
[15] Depending on the interest rate environment, this disadvantage can also be considered as an advantage.
[16] In fact, Theorem 2.32 is not required, since strong existence and pathwise uniqueness as well as the explicit form of the solution follow more elementarily from [53, Lemma 8.21].

Definition 3.4 (Auxiliary quantities in the LSVM)**.**

(a) Let $D \coloneqq (S^0)^{-1}$ denote the discount process. Explicitly,

$$D_t = \exp\Big(-\int_0^t r_u \,\mathrm{d}u\Big), \qquad t \in [0, \overline{T}]. \tag{3.8}$$

In addition, define for all $s \le t$ in $[0, \overline{T}]$,

$$D_{st} = \frac{D_t}{D_s} = \exp\Big(-\int_s^t r_u \,\mathrm{d}u\Big),$$

which is the amount of money which would have to be invested at time s to obtain a risk-free payoff of 1 unit of currency at time t.[17]

(b) For all $s \le t$ in $[0, \overline{T}]$, let

$$P_{st} \coloneqq \mathbb{E}_{\mathbb{Q}}\big[D_{st}|\mathcal{F}_s\big] = \mathbb{E}_{\mathbb{Q}}\Big[\exp\Big(-\int_s^t r_u \,\mathrm{d}u\Big)\Big|\mathcal{F}_s\Big] \tag{3.9}$$

denote the fair value of the zero-coupon bond with maturity t at time s, and let

$$r_t^0 \coloneqq -\partial_t \ln P_{0t}, \qquad t \in [0, \overline{T}], \tag{3.10}$$

denote the spot rate curve associated with the initial family $(P_{0t})_{t\in[0,\overline{T}]}$ of zero-coupon bond prices, which is assumed to be smooth enough such that r^0 can be defined by (3.10).

(c) For all $(t, K) \in [0, \overline{T}] \times \mathbb{R}_+$, let

$$\mathcal{C}_{\mathbb{Q}}(t, K) \coloneqq \mathbb{E}_{\mathbb{Q}}[D_{0t}(S_t - K)^+], \tag{3.11}$$

denote the fair value of the European call option with maturity t and strike price K w.r.t. the risk-neutral measure $\mathbb{Q}$.

3.1.3. The Calibration Condition

The main result of this section, Proposition 3.5, states that the local volatility of a local stochastic volatility model is determined by the surface $\mathcal{C}_{\mathbb{Q}}$ of model-induced call option prices defined in (3.11) and certain unconditional and conditional expectations w.r.t. the underlying measure $\mathbb{Q}$.

Proposition 3.5 (The calibration condition, [26, Proposition 11.1])**.** *Consider a local stochastic volatility model (Definition 3.1) with stochastic volatility $a = (a_t)_{t\in[0,\overline{T}]}$ and local volatility σ as in (3.1). Assume that the associated surface $\mathcal{C}_{\mathbb{Q}}$ of call option prices defined*

[17] Since the interest rate process r is random, D_{st} is itself random (and not $\mathcal{F}_s$-measurable in general), hence we do not usually "know" this amount before time t.

in (3.11) *is continuously differentiable in* t*, twice continuously differentiable in* K *and that* $\partial_K^2 \mathcal{C}_{\mathbb{Q}}$ *is strictly positive.*[18] *Then*

$$\sigma(t,K)^2 \frac{\mathbb{E}_{\mathbb{Q}}[D_{0t}a_t^2 | S_t = K]}{\mathbb{E}_{\mathbb{Q}}[D_{0t} | S_t = K]} = \sigma_{\mathrm{Dup}}(t,K)^2 - \frac{\mathbb{E}_{\mathbb{Q}}[D_{0t}(r_t - r_t^0)\,\mathbb{1}_{\{S_t > K\}}]}{\frac{1}{2}K\partial_K^2 \mathcal{C}_{\mathbb{Q}}(t,K)}, \tag{3.12}$$

for all $(t,K) \in [0,\overline{T}] \times \mathbb{R}_+$*, where we use the notations from Definition 3.4 and*

$$\sigma_{\mathrm{Dup}}(t,K)^2 := \frac{\partial_t \mathcal{C}_{\mathbb{Q}}(t,K) + r_t^0 K \partial_K \mathcal{C}_{\mathbb{Q}}(t,K)}{\frac{1}{2}K^2 \partial_K^2 \mathcal{C}_{\mathbb{Q}}(t,K)}, \qquad (t,K) \in [0,\overline{T}] \times \mathbb{R}_+, \tag{3.13}$$

denotes Dupire's local volatility *function.*

Proof sketch. We follow the heuristic line of argument presented in the proof of [26, Proposition 11.1]. For each $t \in [0,\overline{T}]$, let $\mathcal{P}_t := D_{0t}(S_t - K)^+$ be the discounted payoff of the European call option with maturity t and strike K. Applying the product rule for continuous semimartingales together with the Itô–Tanaka formula[19] for the term $(S_t - K)^+$,

$$\mathrm{d}\mathcal{P}_t = -D_{0t}(S_t - K)^+ r_t \,\mathrm{d}t + D_{0t}\,\mathbb{1}_{\{S_t > K\}}\,\mathrm{d}S_t + \frac{1}{2} D_{0t}\delta(S_t - K)\,\mathrm{d}[S]_t\,, \tag{3.14}$$

where δ denotes the Dirac delta distribution. From (3.5), the second term in (3.14) expands, using also that $S_t\,\mathbb{1}_{\{S_t > K\}} = (S_t - K)^+ + \mathbb{1}_{\{S_t > K\}}\,K$, to

$$\begin{aligned} D_{0t}\,\mathbb{1}_{\{S_t > K\}}\,\mathrm{d}S_t &= D_{0t}\,\mathbb{1}_{\{S_t > K\}}\,S_t\big(r_t\,\mathrm{d}t + \sigma(t,S_t)a_t\,\mathrm{d}W_t\big) \\ &= D_{0t}(S_t - K)^+ r_t\,\mathrm{d}t + D_{0t}\,\mathbb{1}_{\{S_t > K\}}\,K\,r_t\,\mathrm{d}t + D_{0t}\,\mathbb{1}_{\{S_t > K\}}\,S_t\sigma(t,S_t)a_t\,\mathrm{d}W_t\,. \end{aligned} \tag{3.15}$$

The first term on the right-hand side in (3.15) cancels with the first term in (3.14). Moreover, by (3.5), $\mathrm{d}[S]_t = S_t^2\sigma(t,S_t)^2 a_t^2\,\mathrm{d}t$ and $S_t^2\sigma(t,S_t)^2\delta(S_t - K) = K^2\sigma(t,K)^2\delta(S_t - K)$, hence

$$\begin{aligned} \mathrm{d}\mathcal{P}_t = \Big(D_{0t}\,\mathbb{1}_{\{S_t > K\}}\,K\,r_t + \frac{1}{2}D_{0t}K^2\sigma(t,K)^2 a_t^2\delta(S_t - K)\Big)\,\mathrm{d}t \\ + D_{0t}\,\mathbb{1}_{\{S_t > K\}}\,S_t\sigma(t,S_t)a_t\,\mathrm{d}W_t\,. \end{aligned}$$

Rewriting this in integral form,

$$\mathcal{P}_t \overset{\text{a.s.}}{=} (S_0 - K)^+ + \int_0^t H_s\,\mathrm{d}s + M_t, \qquad t \in [0,\overline{T}], \tag{3.16}$$

where

$$H_s := D_{0s}\,\mathbb{1}_{\{S_s > K\}}\,K\,r_s + \frac{1}{2}D_{0s}K^2\sigma(s,K)^2 a_s^2\delta(S_s - K), \qquad s \in [0,\overline{T}], \tag{3.17}$$

[18] The assumption of strict positivity is reasonable, since by the Breeden–Litzenberger formula, see [7], the second derivative $\partial_K^2\mathcal{C}_{\mathbb{Q}}(t,K)$ is, at least in the case of deterministic interest rates, equal to $D_{0t}f_S(t,K)$, where $f_S(t,\cdot)$ denotes the density of S_t under the risk-neutral measure $\mathbb{Q}$, provided that such a density exists, is continuous and is such that S_t is integrable w.r.t. $f_S(t,\cdot)$. This can be seen by twofold differentiation of (3.11) (expressed through $f_S(t,\cdot)$) w.r.t. K.

[19] The Itô–Tanaka formula generalizes Itô's formula to the case of convex (instead of C^2) functions. It is obtained from the theory of local times of continuous semimartingales, for which we refer to [35, §3.6] and [38, §9].

and $M := \int_0^\cdot D_{0s}\,\mathbb{1}_{\{S_s>K\}}\,S_s\sigma(s,S_s)a_s\,\mathrm{d}W_s$, which is a continuous local martingale by local boundedness of the local volatility σ and continuity of sample paths of the interest rate r, the stochastic volatility a and the stock price S. We make the further assumption that M is in fact a true martingale. In this case, taking the expectation w.r.t. $\mathbb{Q}$ of both sides of (3.16), we obtain with the definition (3.11) of $\mathcal{C}_\mathbb{Q}$,

$$\mathcal{C}_\mathbb{Q}(t,K) = \mathbb{E}_\mathbb{Q}[\mathcal{P}_t] = (S_0-K)^+ + \mathbb{E}_\mathbb{Q}\bigg[\int_0^t H_s\,\mathrm{d}s\bigg].$$

Assuming further that H is integrable over $[0,\overline{T}]\times\Omega$ w.r.t. $\lambda\otimes\mathbb{Q}$, Fubini's theorem yields

$$\mathcal{C}_\mathbb{Q}(t,K) = (S_0-K)^+ + \int_0^t \mathbb{E}_\mathbb{Q}[H_s]\,\mathrm{d}s\,, \qquad t\in[0,\overline{T}]. \tag{3.18}$$

If in addition $[0,\overline{T}]\ni s\mapsto \mathbb{E}_\mathbb{Q}[H_s]$ is continuous,[20] (3.18) can be differentiated w.r.t. time to obtain, using also the definition (3.17) of H,

$$\begin{aligned}\partial_t\mathcal{C}_\mathbb{Q}(t,K) &= \mathbb{E}_\mathbb{Q}[H_t]\\ &= K\mathbb{E}_\mathbb{Q}[D_{0t}\,\mathbb{1}_{\{S_t>K\}}\,r_t] + \frac{1}{2}K^2\sigma(t,K)^2\mathbb{E}_\mathbb{Q}[D_{0t}a_t^2\delta(S_t-K)].\end{aligned} \tag{3.19}$$

We interpret expectations involving the Dirac delta term $\delta(S_t-K)$ as expectations conditional[21] on $S_t=K$. Hence the last expectation in the previous display can be written as

$$\mathbb{E}_\mathbb{Q}[D_{0t}a_t^2\delta(S_t-K)] = \mathbb{E}_\mathbb{Q}[D_{0t}a_t^2|S_t=K]. \tag{3.20}$$

Assuming furthermore that the following differentiations under the expectation are justified, we have from (3.11)

$$\partial_K\mathcal{C}_\mathbb{Q}(t,K) = -\mathbb{E}_\mathbb{Q}[D_{0t}\,\mathbb{1}_{\{S_t>K\}}], \quad \text{and} \tag{3.21}$$

$$\partial_K^2\mathcal{C}_\mathbb{Q}(t,K) = \mathbb{E}_\mathbb{Q}[D_{0t}\delta(S_t-K)] = \mathbb{E}_\mathbb{Q}[D_{0t}|S_t=K]. \tag{3.22}$$

The first term on the right-hand side in (3.19) can be written as

$$K\mathbb{E}_\mathbb{Q}[D_{0t}\,\mathbb{1}_{\{S_t>K\}}\,r_t] = K\mathbb{E}_\mathbb{Q}[D_{0t}\,\mathbb{1}_{\{S_t>K\}}(r_t-r_t^0)] + r_t^0K\,\overbrace{\mathbb{E}_\mathbb{Q}[D_{0t}\,\mathbb{1}_{\{S_t>K\}}]}^{\overset{(3.21)}{=}-\partial_K\mathcal{C}_\mathbb{Q}(t,K)} \tag{3.23}$$

with the deterministic initial spot rate process r^0 from (3.10). We can rewrite (3.19), using (3.23) for the first and (3.20) together with (3.22) for the second term, obtaining

$$\begin{aligned}\partial_t\mathcal{C}_\mathbb{Q}(t,K) = K\mathbb{E}_\mathbb{Q}[D_{0t}(r_t-r_t^0)\,\mathbb{1}_{\{S_t>K\}}] - r_t^0K\,\partial_K\mathcal{C}_\mathbb{Q}(t,K)\\ + \frac{1}{2}K^2\sigma(t,K)^2\,\partial_K^2\mathcal{C}_\mathbb{Q}(t,K)\frac{\mathbb{E}_\mathbb{Q}[D_{0t}a_t^2|S_t=K]}{\mathbb{E}_\mathbb{Q}[D_{0t}|S_t=K]}.\end{aligned}$$

[20] From the assumed $(\lambda\otimes\mathbb{Q})$-integrability of $[0,\overline{T}]\times\Omega\ni(s,\omega)\mapsto H_s(\omega)$ and Fubini's theorem, $s\mapsto\mathbb{E}_\mathbb{Q}[H_s]$ is integrable. Therefore, by the Lebesgue differentiation theorem [21, Theorem 3.21], $t\mapsto\int_0^t\mathbb{E}_\mathbb{Q}[H_s]\,\mathrm{d}s$ is at least λ-a.e. differentiable.

[21] An expression of the form $\mathbb{E}[Y|X=x]$, where X and Y are random variables with X taking values in a Polish space S and Y is $\mathbb{K}$-valued and integrable can be understood by viewing $x\mapsto\mathbb{E}[Y|X=x]$ as a factorizing map in the sense of Lemma B.8 of (a version of) the $\sigma(X)$-measurable usual conditional expectation $\mathbb{E}[Y|\sigma(X)]$. We will not go into details about conditional expectations in this text and instead refer to [36, §8.3, in particular Definition 8.24] for details.

Rearranging, the previous display is equivalent to

$$\begin{aligned}\sigma(t,K)^2\frac{\mathbb{E}_{\mathbb{Q}}[D_{0t}a_t^2|S_t=K]}{\mathbb{E}_{\mathbb{Q}}[D_{0t}|S_t=K]}&=\frac{\partial_t\mathcal{C}_{\mathbb{Q}}(t,K)+r_t^0K\,\partial_K\mathcal{C}_{\mathbb{Q}}(t,K)}{\frac{1}{2}K^2\partial_K^2\mathcal{C}_{\mathbb{Q}}(t,K)}-\frac{\mathbb{E}_{\mathbb{Q}}[D_{0t}(r_t-r_t^0)\,\mathbb{1}_{\{S_t>K\}}]}{\frac{1}{2}K\,\partial_K^2\mathcal{C}_{\mathbb{Q}}(t,K)}\\&=\sigma_{\mathrm{Dup}}^2(t,K)-\frac{\mathbb{E}_{\mathbb{Q}}[D_{0t}(r_t-r_t^0)\,\mathbb{1}_{\{S_t>K\}}]}{\frac{1}{2}K\partial_K^2\mathcal{C}_{\mathbb{Q}}(t,K)},\end{aligned}$$

where we have used the assumed strict positivity of $\partial_K^2\mathcal{C}_{\mathbb{Q}}$ and applied (3.13) for the last equality. □

Remark 3.6 (Interpretation and application of the calibration condition (3.12))**.** Denote for all $(t,K)\in[0,\overline{T}]\times\mathbb{R}_+$ by $\mathcal{C}_{\mathrm{mkt}}(t,K)$ the price of the call option with maturity t and strike price K observed in the market.[22] If market prices can at all be explained by an LSVM model, i.e.

$$\mathcal{C}_{\mathrm{mkt}}=\mathcal{C}_{\mathbb{Q}},\tag{3.24}$$

for some LSVM (Definition 3.1) with associated risk-neutral measure $\mathbb{Q}$, and where $\mathcal{C}_{\mathbb{Q}}$ is defined by (3.11), then the local volatility function σ of the LSVM is by the calibration condition (3.12) given by

$$\sigma(t,K)^2=\left(\sigma_{\mathrm{Dup}}^2(t,K)-\frac{\mathbb{E}_{\mathbb{Q}}[D_{0t}(r_t-r_t^0)\,\mathbb{1}_{\{S_t>K\}}]}{\frac{1}{2}K\partial_K^2\mathcal{C}_{\mathrm{mkt}}(t,K)}\right)\frac{\mathbb{E}_{\mathbb{Q}}[D_{0t}|S_t=K]}{\mathbb{E}_{\mathbb{Q}}[D_{0t}a_t^2|S_t=K]},\tag{3.25}$$

at least assuming that the ad-hoc assumptions made in the proof of Proposition 3.5 are valid. Equation (3.25) allows to obtain (or, more realistically, approximate) the local volatility function σ, provided we can evaluate (or approximate) the conditional and unconditional expectations in (3.25). According to [26, §11.6 and §11.7], such an approximation is provided by the particle algorithm described below. We leave open the question which conditions on a surface $\mathcal{C}_{\mathrm{mkt}}$ of call option prices guarantee the existence of a LSVM for which (3.24) holds.[23]

Assuming that (3.24) holds, the SDE (3.5) governing the risky asset dynamics in the LSVM can, by (3.25), be written as

$$\mathrm{d}S_t=S_t\big(r_t\,\mathrm{d}t+\sigma_S(t,S_t,\mathcal{L}[a,r,S])a_t\,\mathrm{d}W_t\big),\qquad t\in[0,\overline{T}],\tag{3.26}$$

[22] The existence of such a function is unrealistic, as we can at any given point of time observe only a finite sample of call option prices. Consequently, whenever a smooth surface $\mathcal{C}_{\mathrm{mkt}}$ of call option prices is required, an interpolation (and, possibly, extrapolation) step is needed, and this brings with it its own challenges. E.g., when interpolating, the resulting surface should satisfy certain no-arbitrage conditions. See [49] for an overview (and detailed investigation) of different approaches to this interpolation problem.

[23] In the literature, see e.g. [26, Proposition 11.1], it seems to be commonly assumed that there is, at least for practical purposes, always some LSVM such that (3.24) is satisfied. [49] and [52] investigate the related question which conditions have to be imposed on $\mathcal{C}_{\mathrm{mkt}}$ in order to guarantee an analogue of (3.24), where S in (3.11) is replaced by a general nonnegative Markov martingale.

where σ_S is defined, cf. the right-hand side of (3.25), by[24]

$$\sigma_S(t,K,\mu)^2 := \left(\sigma^2_{\mathrm{Dup}}(t,K) - \frac{\mathbb{E}_\mu[D_{0t}(r_t - r^0_t)\,\mathbb{1}_{\{S_t>K\}}]}{\frac{1}{2}K\partial^2_K\mathcal{C}_{\mathrm{mkt}}(t,K)}\right)\frac{\mathbb{E}_\mu[D_{0t}|S_t=K]}{\mathbb{E}_\mu[D_{0t}a^2_t|S_t=K]}, \tag{3.27}$$

for all $t\in[0,\overline{T}]$, $K\in\mathbb{R}_+$ and $\mu\in\mathcal{P}_1(C([0,\overline{T}],\mathbb{R})\times C([0,\overline{T}],\mathbb{R})\times C([0,\overline{T}],\mathbb{R}_+))$, with the Dupire local volatility σ_{Dup} from (3.13) and where in (3.27), we understand r, a and S as the canonical projections on $C([0,\overline{T}],\mathbb{R})\times C([0,\overline{T}],\mathbb{R})\times C([0,\overline{T}],\mathbb{R}_+)$, with D_{0t} defined accordingly as in (3.8).[25]

Remark 3.7 (Caveats surrounding the calibration condition (3.12))**.**

(a) While (3.26) has superficially the form of the McKean SDE (2.39), the appearance of conditional expectations in the right-hand side of (3.27) raises the question of $\mathbb{G}\otimes\mathbb{H}$-progressive measurability of coefficients of this SDE (see Definition 2.11). Even if the question of measurability is addressed, the Lipschitz property of coefficients (Definition 2.19) required to infer strong existence and uniqueness of solutions from Theorem 2.32 is not given for σ_S from (3.27) due to Example D.14.

(b) In [26, Remark 11.4], it is stated that the right-hand side of (3.25) can be negative, in which case market prices cannot be explained by a LSVM (meaning that there is no LSVM as in Definition 3.1 and an associated risk-neutral measure $\mathbb{Q}$ such that (3.24) is satisfied). Alternatively, negativity of the right-hand side in (3.25) could result from one of the assumptions made in the proof of Proposition 3.5 not being satisfied.

We do not investigate these questions further at this point. This is also because, due to the introduction of regularizing kernels in a subsequent approximation step described in Subsection 3.1.4 below, the SDE (3.26) is not the one solved in practice when calibrating LSVMs.

Remark 3.8. Very recently, in [18], existence (but not uniqueness) of solutions of the calibrated LSVM have been obtained assuming a specific form of the stochastic volatility, under mild continuity and growth assumptions.

3.1.4. The Particle Method

When calibrating a LSVM to a surface of vanilla (i.e. European call or put) option prices, we assume that the stochastic volatility processes and risk-free rate (Definition 3.1 (c) and (b)) are described by the classical SDEs

$$\mathrm{d}a_t = b_a(t,a)\,\mathrm{d}t + \sigma_a(t,a)\,\mathrm{d}W^a_t\,, \tag{3.28}$$

[24] Due to the presence of the stochastic interest rate process r, the right-hand side of (3.27) depends (through the discount factor D_{0t}) on the law of the entire past path of the market model. That is, we are closer to the situation of (2.37) than (2.39).

[25] Strictly speaking, to go from (3.25) to (3.26), we also need to check that the conditional expectation of one random variable conditioned on another has a functional form (in the sense of Lemma B.8) depending (at least up to null sets) only on the joint distribution of the random variables and not the ambient measure. This is the case, though we omit the proof.

and

$$\mathrm{d}r_t = b_r(t,r)\,\mathrm{d}t + \sigma_r(t,r)\,\mathrm{d}W_t^r \tag{3.29}$$

with deterministic initial values a_0 and r_0, $\mathbb{R}$-valued standard Brownian motions W^a and W^r and $\mathbb{R}$-valued, $\mathbb{G}$-progressive (see (2.5)) coefficients b_a, σ_a, b_r and σ_r, which are assumed to be such that there is pathwise uniqueness and strong existence of solutions.

Remark 3.9. In practice, the coefficients in (3.29) and (3.28) have a specific functional form depending on a chosen model, see Example 3.3 for two concrete possibilities. Depending on the model, there is a number of parameters which need to be chosen to obtain a concrete SDE for which solutions can be found or approximated. This choice usually depends on a prior calibration, e.g., to bond market prices in the case of the risk-free rate, but this will be of no further concern to us at present — instead, we will assume the coefficients in (3.29) and (3.28) to be given.

The market evolution is then described by the triple $X = (a, r, S)$ of the stochastic volatility, interest rate and risky asset price processes. Therefore, we seek to obtain (approximate) solutions of the SDE

$$\mathrm{d}X_t = b(t,X)\,\mathrm{d}t + \sigma_X(t,X,\mathcal{L}[X])\,\mathrm{d}B_t\,, \tag{3.30}$$

where B is an $\mathbb{R}^3$-valued standard Brownian motion, with given initial values $a_0 \in \mathbb{R}$, $r_0 \in \mathbb{R}_+$ and $S_0 \in \mathbb{R}_+$, where

$$b_X(t,(a,r,S)) := \begin{pmatrix} b_a(t,a) \\ b_r(t,r) \\ S_t r_t \end{pmatrix} \quad \text{and} \quad \sigma_X(t,(a,r,S),\mu) := \begin{pmatrix} \sigma_a(t,a) \\ \sigma_r(t,r) \\ S_t\sigma_S(t,S_t,\mu)a_t \end{pmatrix}, \tag{3.31}$$

for all $t \in [0,\overline{T}]$, all $(a,r,S) \in C([0,\overline{T}],\mathbb{R}) \times C([0,\overline{T}],\mathbb{R}) \times C([0,\overline{T}],\mathbb{R}_+)$ and all $\mu \in \mathcal{P}_1(C([0,\overline{T}],\mathbb{R}) \times C([0,\overline{T}],\mathbb{R}) \times C([0,\overline{T}],\mathbb{R}_+))$, see also (3.26) and (3.27). The SDE (3.30) has the form of a McKean SDE (cf. (2.36)), except for the questionable $\mathbb{G}\otimes\mathbb{H}$-progressive measurability of σ_X discussed in Remark 3.7(a).

Remark 3.10 (Correlated driving Brownian motions)**.** In practice, it is sensible to assume that the individual $\mathbb{R}$-valued Brownian motions driving the stock, the risk-free rate and the stochastic volatility are correlated. This means that B in (3.30) has components which are standard Brownian motions, but B is not, as assumed above, a multidimensional standard Brownian motion, whose components are after all independent by definition. We discuss SDEs driven by correlated Brownian motions in some detail in Appendix C.3, see in particular Example C.42. From the discussion therein, adding (reasonably simple) correlations to the model adds mainly notational complexity and will thus be omitted for clarity in this section.

The calibration of the LSVM to market prices then consists of obtaining approximate solutions of (3.30) by a variant of the particle method introduced in Subsection 2.4.2. As in Definition 2.51, the idea is to fix a particle number $N \in \mathbb{N}$ and pass from the SDE (3.30) to the system of N coupled SDEs

$$\mathrm{d}X_t^{i,N} = b_X\big(t,X^{i,N}\big)\,\mathrm{d}t + \sigma_X\big(t,X^{i,N},\tilde{\mu}^N\big)\,\mathrm{d}B_t^i\,, \qquad i \in \{1,\dots,N\}, \tag{3.32}$$

with initial conditions $X_0^{i,N} \overset{\text{a.s.}}{=} (a_0, r_0, S_0)$, $i \in \{1, \dots, N\}$, for a $\mathbb{R}^{3N}$-valued solution process $X^N = (X^{1,N}, \dots, X^{N,N})$ whose components we refer to as particles, where the empirical measure $\tilde{\mu}^N$ is given by (2.97) and B is an $\mathbb{R}^{3N}$-valued standard Brownian motion.

In [26, §11.6.1], it is stated that (3.32) cannot be taken literally, since the conditional expectations in (3.27) are undefined when μ is an empirical measure due to non-existence of a density in this case. Even in the absence of densities, we can interpret conditional expectations w.r.t. empirical measures by using regular conditional distributions. In Remark 3.11 below, we show that this interpretation has, at least under the assumption (3.37), the paradoxical consequence that the particles are independent and follow local volatility instead of local stochastic volatility dynamics (cf. (3.1)).

Remark 3.11 (Effective decoupling and reduction to a local volatility model of (3.32))**.** Fix $N \in \mathbb{N}$ and, for now, $i \in \{1, \dots, N\}$. The third component of the coefficient $\sigma_X(t, X^{i,N}, \tilde{\mu}^N)$ from (3.32) depends, see (3.31), on $\sigma_S(t, X_t^{i,N}, \tilde{\mu}^N)$, which by (3.27) depends on the conditional expectations

$$\mathbb{E}_{\tilde{\mu}^N}[D_{0t} | S_t = S_t^{i,N}] \qquad \text{and} \qquad \mathbb{E}_{\tilde{\mu}^N}[D_{0t} a_t^2 | S_t = S_t^{i,N}], \tag{3.33}$$

where, as in (3.27) and from now on, we understand r, a and S as the canonical projections on $C([0,\overline{T}], \mathbb{R}) \times C([0,\overline{T}], \mathbb{R}) \times C([0,\overline{T}], \mathbb{R}_+)$, with D_{0t} defined as in (3.8). It follows from Lemma B.12 applied with $\pi_X = S$, $\pi_Y = (a, r)$ and $\varphi(S) \coloneqq S_t$ for all $S \in C([0,\overline{T}], \mathbb{R}_+)$ that a regular conditional distribution of (a, r) given S_t w.r.t. the empirical measure $\tilde{\mu}^N$ is given by

$$\mathcal{L}_{\tilde{\mu}^N}[(a,r) | S_t = s] = \begin{cases} \dfrac{1}{|I_s|} \displaystyle\sum_{j \in I_s} \delta_{(a^{j,N}, r^{j,N})} & |I_s| > 0, \\ \delta_{(0,0)} & |I_s| = 0, \end{cases} \tag{3.34}$$

where $I_s \coloneqq \{i \in \{1, \dots, N\} : S_t^{i,N} = s\}$, $s \in \mathbb{R}_+$. Hence the conditional expectations in (3.33) are by [36, Theorem 8.38][26] equal to

$$\mathbb{E}_{\tilde{\mu}^N}[D_{0t} a_t^2 | S_t = S_t^{i,N}] = \frac{1}{|I_{S_t^{i,N}}|} \sum_{\substack{j \in \{1,\dots,N\} \\ S_t^{j,N} = S_t^{i,N}}} D_{0t}^{j,N} (a_t^{j,N})^2 \tag{3.35}$$

and

$$\mathbb{E}_{\tilde{\mu}^N}[D_{0t} | S_t = S_t^{i,N}] = \frac{1}{|I_{S_t^{i,N}}|} \sum_{\substack{j \in \{1,\dots,N\} \\ S_t^{j,N} = S_t^{i,N}}} D_{0t}^{j,N} \tag{3.36}$$

where $D_{0t}^{j,N} \coloneqq \exp(-\int_0^t r_s^{j,N} \, \mathrm{d}s)$ for all $j \in \{1, \dots, N\}$ and $t \in [0, \overline{T}]$. This is problematic, since in the sums in (3.35) and (3.36), we don't expect the case $S_t^{j,N} = S_t^{i,N}$ to occur with positive probability when $i \neq j$, and indeed this is impossible if the random vector

[26] This theorem is stated for integrable random variables, but its proof is equally valid for nonnegative random variables, which is the case with D_{0t} and $D_{0t} a_t^2$ in (3.35) and (3.36).

$(S_t^{1,N}, \dots, S_t^{N,N})$ has a density w.r.t. the Lebesgue measure. But if this is true, i.e. if under the probability measure $\mathbb{Q}^N$ of the solution X^N of (2.96),

$$\mathbb{Q}^N[S_t^{i,N} = S_t^{j,N}] = \mathbb{1}_{i=j}, \qquad t \in [0, \overline{T}],\ i,j \in \{1, \dots, N\}, \tag{3.37}$$

then from (3.35),

$$\mathbb{E}_{\tilde{\mu}^N}[D_{0t} a_t^2 | S_t = S_t^{i,N}] \overset{\text{a.s.}}{=} D_{0t}^{i,N} (a_t^{i,N})^2, \tag{3.38}$$

and similarly from (3.36),

$$\mathbb{E}_{\tilde{\mu}^N}[D_{0t} | S_t = S_t^{i,N}] \overset{\text{a.s.}}{=} D_{0t}^{i,N} \tag{3.39}$$

for all $t \in [0, \overline{T}]$ and $i, j \in \{1, \dots, N\}$. This means that for each $i \in \{1, \dots, N\}$, the conditional expectations in the third component of $\sigma_X(t, X^{i,N}, \tilde{\mu}^N)$ in (3.32) effectively depend only on the i-th particle $X^{i,N}$, and on none of the others. The only true dependence of $\sigma_X(t, X^{i,N}, \tilde{\mu}^N)$ on $\tilde{\mu}^N$ (and thus the only true interaction) is through the unconditional expectation $\mathbb{E}_{\tilde{\mu}^N}[D_{0t}(r_t - r_t^0)\, \mathbb{1}_{\{S_t > S_t^{i,N}\}}]$ in (3.27). Assuming in addition that the interest rate is deterministic (then in particular $r = r^0$, so the second term in the parenthesis in (3.25) vanishes), the particle approximation SDE (3.32) is effectively decoupled, with the (a, S) components of individual particles being independent solutions of the SDE

$$\begin{aligned} \mathrm{d}a_t &= b_a(t, a)\, \mathrm{d}t + \sigma_a(t, a)\, \mathrm{d}W_t^a\,, \\ \mathrm{d}S_t &= S_t\big(r_t\, \mathrm{d}t + \sigma_{\text{eff}}(t, S_t) a_t\, \mathrm{d}W_t^S\big), \end{aligned} \tag{3.40}$$

with (see (3.27), (3.38) and (3.39)) the effective local volatility

$$\sigma_{\text{eff}}(t, s)^2 := \sigma_{\text{Dup}}^2(t, s) \frac{\mathbb{E}_{\mathbb{Q}}[D_{0t} | S_t = s]}{\mathbb{E}_{\mathbb{Q}}[D_{0t} a_t^2 | S_t = s]} = \sigma_{\text{Dup}}^2(t, s) \frac{1}{a_t^2}$$

for all $t \in [0, \overline{T}]$ and $s \in \mathbb{R}_+$. Comparing with (3.40), we see that the particles given by (3.32) follow, assuming additionally that the solutions of (3.28) are positive up to indistinguishability, in fact, independently Dupire's local volatility model, i.e. the original LSVM (3.5) with the stochastic volatility term a set to one. This dynamic is materially different from the one given by (3.26) and (3.27). Hence (3.5) should not be taken literally.

The Regularizing Kernel Approximation

This section describes the approach followed in [26, §11.6.1] to sidestep the issue described in Remark 3.11. The starting point is that (3.35) can be rewritten as

$$\mathbb{E}_{\tilde{\mu}^N}[D_{0t} a_t^2 | S_t = S_t^{i,N}] \overset{\text{a.s.}}{=} \frac{\sum_{j=1}^N D_{0t}^{j,N} (a_t^{j,N})^2\, \mathbb{1}_{\{S_t^{j,N} = S_t^{i,N}\}}}{\sum_{j=1}^N \mathbb{1}_{\{S_t^{j,N} = S_t^{i,N}\}}},$$

and similarly (3.36) as

$$\mathbb{E}_{\tilde{\mu}^N}[D_{0t} | S_t = S_t^{i,N}] \overset{\text{a.s.}}{=} \frac{\sum_{j=1}^N D_{0t}^{j,N}\, \mathbb{1}_{\{S_t^{j,N} = S_t^{i,N}\}}}{\sum_{j=1}^N \mathbb{1}_{\{S_t^{j,N} = S_t^{i,N}\}},}.$$

The idea proposed in [26, §11.6.1] is to introduce a *regularizing kernel* $\delta_{t,N}\colon \mathbb{R}\to\mathbb{R}_+$ and replace, for all $i,j\in\{1,\dots,N\}$, the indicator[27] $\mathbb{1}_{\{S_t^{j,N}=S_t^{i,N}\}}$ in the previous two displays by $\delta_{t,N}(S_t^{j,N}-S_t^{i,N})$. The regularizing kernel is defined as

$$\delta_{t,N}(x)=\frac{1}{h_{t,N}}K\left(\frac{x}{h_{t,N}}\right),\qquad x\in\mathbb{R},$$

with a time-dependent bandwidth[28] $h_{t,N}\in(0,\infty)$ tending, for each fixed $t\in\mathbb{R}_+$, to zero as $N\to\infty$, and where $K\colon\mathbb{R}\to\mathbb{R}_+$ is a fixed, symmetric kernel, e.g. $K(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$ or $K(x)=\frac{15}{16}(1-x^2)^2\,\mathbb{1}_{|x|\le 1}$ for all $x\in\mathbb{R}$. Consequently, we make in (3.32), see also (3.31) and (3.27), for each $i\in\{1,\dots,N\}$, the following replacements:

$$\mathbb{E}_{\tilde\mu^N}\left[D_{0t}\,\middle|\,S_t=S_t^{i,N}\right]\quad\longrightarrow\quad\frac{\sum_{j=1}^N D_{0t}^{j,N}\delta_{t,N}(S_t^{j,N}-S_t^{i,N})}{\sum_{j=1}^N\delta_{t,N}(S_t^{j,N}-S_t^{i,N})},\tag{3.41}$$

$$\mathbb{E}_{\tilde\mu^N}\left[D_{0t}a_t^2\,\middle|\,S_t=S_t^{i,N}\right]\quad\longrightarrow\quad\frac{\sum_{j=1}^N D_{0t}^{j,N}(a_t^{j,N})^2\delta_{t,N}(S_t^{j,N}-S_t^{i,N})}{\sum_{j=1}^N\delta_{t,N}(S_t^{j,N}-S_t^{i,N})}.\tag{3.42}$$

From the definition (2.97) of $\tilde\mu^N$, we obtain for the unconditional expectation entering in (3.32), see also (3.27),

$$\mathbb{E}_{\tilde\mu^N}\left[D_{0t}(r_t-r_t^0)\,\mathbb{1}_{\{S_t>S_t^{i,N}\}}\right]=\frac{1}{N}\sum_{j=1}^N D_{0t}^{j,N}(r_t^{j,N}-r_t^0)\,\mathbb{1}_{\{S_t^{j,N}>S_t^{i,N}\}}.\tag{3.43}$$

With the replacements from (3.41), (3.42) and (3.43), we have for the volatility coefficient $\sigma_X(t,X^{i,N},\tilde\mu^N)$ of the i-th particle from (3.32), see also (3.31), the replacement

$$\sigma_X(t,X^{i,N},\tilde\mu^N)\quad\longrightarrow\quad\sigma_X^{i,N}(t,X^N),\tag{3.44}$$

where $X^N=(X^{1,N},\dots,X^{N,N})$ is the $\mathbb{R}^{3N}$-valued stochastic process describing the trajectories of all particles and where for all $t\in[0,\overline{T}]$ and $X=(a^1,r^1,S^1,\dots,a^N,r^N,S^N)\in(C([0,\overline{T}],\mathbb{R})\times C([0,\overline{T}],\mathbb{R})\times C([0,\overline{T}],\mathbb{R}_+))^N$,

$$\sigma_X^{i,N}(t,X):=\begin{pmatrix}\sigma_r(t,r^i)\\ \sigma_a(t,a^i)\\ S_t^i\sigma_S^{i,N}(t,X)a_t^i\end{pmatrix},\tag{3.45}$$

with, cf. (3.27) and (3.31),

$$\begin{aligned}\sigma_S^{i,N}(t,X)^2:=\left(\sigma_{\mathrm{Dup}}^2(t,S_t^i)-\frac{\frac{1}{N}\sum_{j=1}^N D_{0t}^j(r_t^j-r_t^0)\,\mathbb{1}_{\{S_t^j>S_t^i\}}}{\frac12 S_t^i\partial_K^2\mathcal{C}_{\mathrm{mkt}}(t,S_t^i)}\right)&\\ \times\frac{\sum_{j=1}^N D_{0t}^j\delta_{t,N}(S_t^j-S_t^i)}{\sum_{j=1}^N D_{0t}^j(a_t^j)^2\delta_{t,N}(S_t^j-S_t^i)}&.\end{aligned}\tag{3.46}$$

[27] In [26, §11.6.1], what we call here an indicator function is referred to as a Dirac function.

[28] The precise value of the bandwidth $h_{t,N}$ is determined by practical considerations for which we refer to [26, §11.6.2].

The introduction of the regularizing kernel, which led to (3.46), restores some interaction between the particles, thereby sidestepping the issue described in Remark 3.11. However, especially in the light of Remark 3.11, the question remains in which sense the particle approximation obtained from the replacement (3.44) truly approximates the SDE (3.30) governing the LSVM. We do not investigate this question further at this point. An empirical investigation of the performance of the particle method for the calibration of LSVMs has been carried out in [26, §11.8] for several choices of the risk-free interest rate and stochastic volatility. There, it was found that the particle method achieves very good calibration to the market smile of implied volatilities. Crucially, this was also verified for models driven by multiple Brownian motions and in the presence of stochastic interest rates, where PDE methods are infeasible for performance reasons. On the other hand, in [26, §11.8.3], it is reported that for certain parameter combinations (in particular a high volatility-of-volatility), the particle method does not converge. It is left open whether this is due to numerical issues or the non-existence of a solution.

Remark 3.12 (Option pricing using the particle method, [26, §11.6])**.** To price options using the particle method, the system of regularized SDEs with volatility coefficients defined by (3.45) for each $i \in \{1, \dots, N\}$ is simulated using numerical methods for classical SDEs. This provides a discrete-time sampling $\hat{X} = (\hat{X}_t^N)_{t\in I} = ((\hat{X}_t^{1,N})_{t\in I}, \dots, (\hat{X}_t^{N,N})_{t\in I})$ of (an approximation of) the $\mathbb{R}^{3N}$-valued solution $X^N = (X^{1,N}, \dots, X^{N,N})$ of (3.32), where $I \subseteq [0, \overline{T}]$ is a finite set of times. Assuming that a form of propagation of chaos (see Subsection 2.4.2, in particular Theorem 2.60) applies to the SDE (3.30), we could expect that the empirical distribution of $(\hat{X}^1, \dots, \hat{X}^N)$ to be a good approximation of the law of $(X_t^1)_{t\in I}$. Consequently, the price of every option depending on the triple (a, r, S) of stochastic volatility, risk-free interest rate and risky asset only on the times in I can be approximated as $\frac{1}{N}\sum_{k=1}^N H^{k,N}$, where $H^{k,N}$ is the discounted payout of the option evaluated on the approximated path $\hat{X}^{k,N}$ of the k-th particle.

3.2. A Mean Field LIBOR Market Model

This section discusses a mean-field variant of the LIBOR (London interbank offered rate) market model (LMM) for interest rates introduced in [17]. LMMs are widely used in the pricing of interest rate derivatives due to their ease of calibration and economic interpretability. Their application has extended to the insurance sector, including the valuation of long-term guarantees under the Solvency II regulatory framework. Life insurance portfolios have time horizons of several decades. In this context, a major issue with LMMs is the risk of explosion, which means that in a significant fraction of scenarios (say, more than 1%), the forward interest rates being modelled exceed realistic thresholds (e.g. 50%). This is economically implausible and leads to numerical issues in Monte Carlo simulations due to discount factors becoming vanishingly small.

These challenges have spawned mitigation techniques such as volatility freezing (meaning that the volatility is set to zero for scenarios where forward rates exceed a defined threshold) and capping (meaning that rates are set equal to a pre-defined cap once it is reached). While these techniques prevent explosion, they have technical disadvantages related to martingale properties and implied volatilities, which are undesirable.

In [17], a mean-field version of the LMM is introduced, where the probability of explosion is significantly reduced by allowing for a dependence of the volatilities of the forward rates on their laws (in particular their variances). This extends the classical LMM, where volatilities are deterministic functions of time.

In Subsection 3.2.1 below, we introduce the classical LMM. Subsequently, in Subsection 3.2.2, we consider the matter of existence and uniqueness of solutions for McKean SDEs with coefficients of a specific form (see (3.55)) where the Lipschitz continuity in the measure argument is relaxed from a global to a local form (see (3.58)). In Theorem 3.19, we generalize an existence and uniqueness result for such SDEs from [17], which makes accessible a class of SDEs characterized by variance-based volatility dampening (see Examples 3.21, 3.24 and 3.26). Finally, in Subsection 3.2.3, we describe the mean-field LIBOR market model introduced in [17].

3.2.1. The Classical LIBOR Market Model

This section recapitulates the classical LIBOR market model, following [43, §2 and §4] and [17, §2]. Let $\overline{T} \in \mathbb{R}_+^*$ be a time horizon and denote, for all $t \leq T$ in $[0, \overline{T}]$, by $P(t,T)$ the fair price at time t of a zero coupon bond paying 1 unit of currency at a maturity date T. By engaging in *forward investing*,[29] we can lock in, at time t, a risk-free return for investment in zero coupon bonds between the future times T and $T+\delta$ in $[t, \overline{T}]$, by the following trading strategy:

(a) At time t, short a zero coupon bond with maturity T, thereby generating income $P(t,T)$.

(b) From the generated income, establish a long position of $P(t,T)/P(t,T+\delta)$ units of zero coupon bonds with maturity $T+\delta$. The combined portfolio costs nothing to set up.

(c) At time T, the short position needs to be covered, necessitating a payment of 1 unit of currency (the payoff of the zero coupon bond).

(d) At time $T+\delta$, the long position of $P(t,T)/P(t,T+\delta)$ shares in zero coupon bonds with maturity $T+\delta$ provides a return of 1 unit of currency per share held, i.e. $P(t,T)/P(t,T+\delta)$.

In summary, we have locked in at time t a return of $P(t,T)/P(t,T+\delta)$ at time $T+\delta$ per unit of currency invested at time T. The *simple return* L_t corresponding to this investment is defined as the (net) return of the investment divided by the investing period, i.e.

$$L_t = \frac{1}{\delta}\left(\frac{P(t,T)}{P(t,T+\delta)} - 1\right) = \frac{1}{\delta}\frac{P(t,T) - P(t,T+\delta)}{P(t,T+\delta)}.$$

[29] See [55, §10.3.1], on which the following description is based, for details. We do not consider transaction costs and liquidity issues in this text.

It is called the forward LIBOR valid on the time interval $[T, T+\delta]$, and the duration δ is in this context called the *tenor* of the LIBOR. In the classical LIBOR market model, maturity times $0 \le t_0 < \cdots < t_N \le \overline{T}$ are fixed, and the forward LIBORs

$$L_t^i := \frac{1}{\delta_i} \frac{P(t, t_{i-1}) - P(t, t_i)}{P(t, t_i)}, \qquad t \le t_{i-1},\ i \in \{1, \dots, N\}, \tag{3.47}$$

where $\delta_i := t_i - t_{i-1}$, are described by SDEs ((3.48) below). This is in contrast to the so-called short rate models encountered in the context of local stochastic volatility models (Definition 3.1), where the short-term risk-free interest rate process $r = (r_t)_{t \in \mathbb{R}_+}$ is the fundamental process being modelled under a risk-neutral measure, and other quantities, say the family $(P(t,T))_{0 \le t \le T \le \overline{T}}$ of prices of zero coupon bonds, are derived using the risk-neutral pricing framework, see (3.9). We refer to [55, §6.5] for details on short-rate models.

Assumption 3.13 (Construction of the LMM, [43, §2 and §4] and [17, §2, (1)–(3)])**.** The classical construction of $L^1, \dots, L^N$ is based on a backward induction approach, which employs the following postulates.

(a) The initial term structure $(P(0, t_i))_{i \in \{0, \dots, N\}}$ is deterministic, positive and non-increasing, hence $L_0^i \ge 0$ for all $i \in \{1, \dots, N\}$.

(b) For each $i \in \{1, \dots, N\}$, there is a bounded measurable volatility[30] $\sigma^i : [0, t_{i-1}] \to \mathbb{R}^d$, a probability measure $\mathbb{Q}^i$, the so-called forward measure associated with the maturity date t_i and a d-dimensional standard $\mathbb{Q}^i$-Brownian motion W^i, such that w.r.t. $\mathbb{Q}^i$,

$$\mathrm{d}L_t^i = L_t^i \sigma^i(t)\, \mathrm{d}W_t^i\,, \qquad t \in [0, t_{i-1}]. \tag{3.48}$$

(c) The forward measures $(\mathbb{Q}^i)_{i \in \{1, \dots, N\}}$ are related by their Radon–Nikodym derivatives,

$$\frac{\mathrm{d}\mathbb{Q}^{i-1}}{\mathrm{d}\mathbb{Q}^i} = \mathcal{E}(\gamma^i \cdot W^i)_{t_{i-1}}, \qquad i \in \{2, \dots, N\}, \tag{3.49}$$

where $\mathcal{E}(\gamma^i \cdot W^i)$ denotes the stochastic exponential of $\gamma^i \cdot W^i$ ([53, Definition 6.59]) and for each $i \in \{1, \dots, N\}$,

$$\gamma_s^i := \frac{\delta_i L_s^i}{\delta_i L_s^i + 1} \sigma^i(s), \qquad s \in [0, t_{i-1}]. \tag{3.50}$$

In addition, the Brownian motions $(W^i)_{i \in \{1, \dots, N\}}$ satisfy

$$W^{i-1} = W^i - \int_0^{\cdot} \gamma_s^i \, \mathrm{d}s\,, \tag{3.51}$$

i.e. W^{i-1} is the $\mathbb{R}^d$-valued standard $\mathbb{Q}^{i-1}$-Brownian motion obtained from W^i using (3.49) and Girsanov's theorem and Lévy's characterization of Brownian motion ([53, Theorem 7.41 and Corollary 7.2]).

[30] To remain within the framework of Definition 2.12, we define $\sigma^i(t) = 0$ for all $t > t_{i-1}$ and $i \in \{1, \dots, N\}$.

The complete family $(L^i)_{i\in\{1,\dots,N\}}$ of forward LIBORs is then constructed inductively by starting from a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{Q})$ (Definition C.5(a)) carrying an $\mathbb{R}^d$-valued standard Brownian motion W, and where the filtration $\mathbb{F}$ is the augmentation by $(\mathbb{Q}, \mathcal{F}^W_{\bar{T}})$-null sets (see Remark C.4(a)) of the natural filtration $\mathbb{F}^W = (\mathcal{F}^W_t)_{t\in\bar{\mathbb{R}}_+}$ generated by W. With $\mathbb{Q}^N := \mathbb{Q}$ and $W^N := W$, we have by (3.48) with $i = N$ the forward LIBOR for the interval $[t_{N-1}, t_N]$,

$$L^N_t \overset{\text{a.s.}}{=} L^N_0 \mathcal{E}(\sigma^N \cdot W^N)_t, \qquad t \in [0, t_{N-1}],$$

with equality up to indistinguishability. Since σ^N is deterministic, L^N is the stochastic exponential of the Gaussian martingale $\sigma^N \cdot W^N$ (see [53, Theorem 7.1 and Exercise 7.6]) and thus lognormally distributed at each $t \in [0, t_{N-1}]$. By (3.50), L^N determines the process γ^N which determines through (3.49) the measure[31] $\mathbb{Q}^{N-1}$ and through (3.51) the $\mathbb{R}^d$-valued standard $\mathbb{Q}^{N-1}$-Brownian motion W^{N-1}. Proceeding inductively, we obtain $L^{N-1}, L^{N-2}, \dots, L^1$. By (3.48), each LIBOR L^i has lognormal marginal distributions under its respective forward measure $\mathbb{Q}^i$. This attractive feature of the model is preserved in the mean-field extension described in Subsection 3.2.3 below.

3.2.2. Strong Existence and Pathwise Uniqueness for a McKean SDE with local Lipschitz Continuity in the Measure Argument

A main aim of the mean-field LMM (described in Subsection 3.2.3 below) is to address the problem of explosion of forward rates in the classical LMM. To this end, the deterministic volatilities in (3.48) are replaced by volatilities depending jointly on time and the laws of the respective forward LIBORs, see (3.119) and (3.122) below. For the SDE coefficients arising in this way the Lipschitz property of coefficients of McKean SDEs in the sense of Definition 2.19 is in question,[32] hence our main existence and uniqueness result, Theorem 2.32, does not apply. Nevertheless, a proof of strong existence and pathwise uniqueness for certain SDEs with coefficients not satisfying the Lipschitz property (2.49) is given in [17, Theorem 5.2], in the setting of instantaneous dependence (Definition 2.13(a)), on a finite time horizon, with $\mathbb{K} = \mathbb{R}$ and the Wasserstein order $p = 2$.

In this section, we generalize the SDEs studied in [17]. We consider the infinite time horizon, include the case $\mathbb{K} = \mathbb{C}$, allow for pathwise dependence of the coefficients on both the solution and its law, consider Wasserstein orders $p \in [2, \infty)$ and a more general functional form of the coefficients.[33] The main result is Theorem 3.19, which establishes strong existence and pathwise uniqueness (Definitions 2.26 and 2.29) for the McKean SDE (2.37) with coefficients as in Assumption 3.14 below. The main application is to SDEs

[31] The measure $\mathbb{Q}^{N-1}$ given by (3.49) is well-defined by Novikov's criterion [53, Theorem 7.111].

[32] Of particular interest for the mean-field LMM is Example 3.21, for which we have not been able to (dis)prove Lipschitz continuity in the sense of Definition 2.19.

[33] In [17, Equation (5.4)], coefficients of the form $b(t, x, \mu, \omega) = b_1(t, x, \mu, \omega) + xg(\mu)h_1(t, \omega)$ and $\sigma(t, x, \mu, \omega) = \sigma_1(t, x, \mu, \omega) + xg(\mu)h_2(t, \omega)^\intercal$ for all $(t, x, \mu, \omega) \in [0, \bar{T}] \times \mathbb{R}^n \times \mathcal{P}_2(\mathbb{R}^n) \times \Omega$ are considered, where $\bar{T} \in \mathbb{R}_+$ denotes a time horizon and g, h_1 and h_2 are functions satisfying instantaneous and time-independent versions of the conditions in Assumption 3.14(c) and (d). In comparison, the form of the coefficients in (3.55) is more flexible.

involving variance-dependent terms as in Example 3.21, which feature in the mean-field LMM discussed in the next section.

Assumption 3.14. [34]Fix $p \in [2,\infty)$, a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ (Definition C.5(a)) and let b and σ be as in (2.35).

(a) There are $\mathbb{G}\otimes\mathbb{H}\otimes\mathbb{F}$-progressive coefficients $\tilde{b}$ and $\tilde{\sigma}$ as in (2.35), functions

$$f_1\colon \mathbb{R}_+ \times W_n \to \mathbb{K}^n, \qquad f_2\colon \mathbb{R}_+ \times W_n \to \mathbb{K}^n, \tag{3.52}$$
$$g_1\colon \mathbb{R}_+ \times \widehat{\mathcal{P}}_p(W_n) \to \mathbb{K}, \qquad g_2\colon \mathbb{R}_+ \times \widehat{\mathcal{P}}_p(W_n) \to \mathbb{K}, \tag{3.53}$$
$$h_1\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}, \qquad h_2\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^d, \tag{3.54}$$

such that $f_1(\,\cdot\,,w)$ and $f_2(\,\cdot\,,w)$ are $\mathcal{B}_{\mathbb{R}_+}$-measurable for each $w \in W_n$, $g_1(\,\cdot\,,\mathbb{Q})$ and $g_2(\,\cdot\,,\mathbb{Q})$ are $\mathcal{B}_{\mathbb{R}_+}$-measurable for each $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$ and h_1 and h_2 are $\mathbb{F}$-progressive. Moreover, for all $(t,w,\mathbb{Q},\omega) \in \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \times \Omega$,

$$\begin{aligned} b(t,w,\mathbb{Q},\omega) &= \tilde{b}(t,w,\mathbb{Q},\omega) + f_1(t,w)g_1(t,\mathbb{Q})h_1(t,\omega) \quad \text{and} \\ \sigma(t,w,\mathbb{Q},\omega) &= \tilde{\sigma}(t,w,\mathbb{Q},\omega) + f_2(t,w)g_2(t,\mathbb{Q})h_2(t,\omega)^{\mathsf{T}}. \end{aligned} \tag{3.55}$$

(b) The coefficients $\tilde{b}$ and $\tilde{\sigma}$ have the Lipschitz property[35] and satisfy the linearly bounded growth condition (Definitions 2.19 and 2.20).

(c) There is an increasing function $\tilde{D}\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$|f_i(t,w)| \le \tilde{D}_t(1+p_t(w)), \qquad |g_i(t,\mathbb{Q})| \le \tilde{D}_t, \qquad \text{and} \qquad |h_i(t,\omega)| \le \tilde{D}_t, \tag{3.56}$$

for all $i \in \{1,2\}$ and all $(t,w,\mathbb{Q},\omega) \in \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \times \Omega$.

(d) There is an increasing function $\tilde{c}\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that for each $i \in \{1,2\}$, all $t \in \mathbb{R}_+$ and $w, \tilde{w} \in W_n$,

$$|f_i(t,w) - f_i(t,\tilde{w})| \le \tilde{c}_t p_t(w-\tilde{w}), \tag{3.57}$$

and moreover for all $\mathbb{Q}, \tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}_p(W_n)$,

$$|g_i(t,\mathbb{Q}) - g_i(t,\tilde{\mathbb{Q}})| \le \tilde{c}_t\Big(1 + \widehat{M}_{t,p}(\mathbb{Q}) + \widehat{M}_{t,p}(\tilde{\mathbb{Q}})\Big)\widehat{\mathcal{W}}_{t,p}(\mathbb{Q},\tilde{\mathbb{Q}}), \tag{3.58}$$

where

$$\widehat{M}_{t,p}(\mathbb{Q}) := \int_{W_n} p_t^p(w)\,\mathrm{d}\mathbb{Q}(w)\,, \qquad t \in \mathbb{R}_+. \tag{3.59}$$

Convention 3.15 (Instantaneous dependence)**.** If the right-hand sides in (3.55) depend on the path argument $w \in W_n$ and the measure argument $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$ only through w_t and $\mathbb{Q}_t$ (see (2.17)), i.e. there are functions $\hat{b}\colon \mathbb{R}_+ \times \mathbb{K}^n \times \mathcal{P}_p(\mathbb{K}^n) \times \Omega \to \mathbb{K}^n$, $\hat{\sigma}\colon \mathbb{R}_+ \times \mathbb{K}^n \times$

[34] Assumption 3.14 generalizes [17, §5, Assumptions $\mathbf{A}_b^1$ to $\mathbf{A}_{b\sigma}^2$].

[35] This assumption allows to relax the required $\mathbb{G}\otimes\mathbb{H}\otimes\mathbb{F}$-progressive measurability of $\tilde{b}$ and $\tilde{\sigma}$ using Lemma 2.21.

$\mathcal{P}_p(\mathbb{K}^n) \times \Omega \to \mathbb{K}^{n\times d}$, $\hat{f}_i\colon \mathbb{R}_+ \times \mathbb{K}^n \to \mathbb{K}^n$ and $\hat{g}_i\colon \mathbb{R}_+ \times \mathcal{P}_p(\mathbb{K}^n) \to \mathbb{K}$, $i \in \{1, 2\}$, such that (cf. (2.38))

$$\begin{aligned} \tilde{b}(t, w, \mathbb{Q}, \omega) &= \hat{b}(t, w_t, \mathbb{Q}_t, \omega), & \tilde{\sigma}(t,w, \mathbb{Q}, \omega) &= \hat{\sigma}(t, w_t, \mathbb{Q}_t, \omega),\\ f_i(t, w) &= \hat{f}_i(t, w_t) \quad \text{and} & g_i(t, \mathbb{Q}) &= \hat{g}_i(t, \mathbb{Q}_t), \end{aligned}$$

for all $i \in \{1, 2\}$ and all $(t, w, \mathbb{Q}, \omega) \in \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \times \Omega$, then we speak, as in Definition 2.13, of instantaneous dependence of the SDE coefficients on the solution. In this case, we make no notational distinction between $\tilde{b}$ and $\hat{b}$, $\tilde{\sigma}$ and $\hat{\sigma}$ and, for $i \in \{1, 2\}$, f_i and $\hat{f}_i$, g_i and $\hat{g}_i$, respectively. Furthermore, as in Definition 2.13, we use the notation from (2.39) instead of (2.37).

Remark 3.16. Fix $p \in [2, \infty)$ and $t \in \mathbb{R}_+$.

(a) For every $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$, the function $(0, p] \ni \alpha \mapsto \widehat{M}^{1/\alpha}_{t,\alpha}(\mathbb{Q})$ is increasing by Jensen's inequality. From the elementary inequality $x^{1/p} \le 1 + x$, which holds for all $x \in \mathbb{R}_+$, it follows that

$$\sup_{\alpha\in(0,p]} \widehat{M}^{1/\alpha}_{t,\alpha}(\mathbb{Q}) = \widehat{M}^{1/p}_{t,p}(\mathbb{Q}) \le 1 + \widehat{M}_{t,p}(\mathbb{Q}), \qquad \mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n). \tag{3.60}$$

Moreover, for every $\alpha \in [0, p]$ and all $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$,

$$\begin{aligned} \widehat{M}_{t,\alpha}(\mathbb{Q}) &= \int_{W_n} p^\alpha_t(w)\, \mathrm{d}\mathbb{Q}(w)\\ &= \int_{W_n} p^\alpha_t(w)\, \mathbb{1}_{\{p_t\le 1\}}(w)\, \mathrm{d}\mathbb{Q}(w) + \int_{W_n} p^\alpha_t(w)\, \mathbb{1}_{\{p_t>1\}}(w)\, \mathrm{d}\mathbb{Q}(w)\\ &\le 1 + \int_{W_n} p^p_t(w)\, \mathrm{d}\mathbb{Q}(w) = 1 + \widehat{M}_{t,p}(\mathbb{Q}). \end{aligned}$$

Hence

$$\sup_{\alpha\in[0,p]} \widehat{M}_{t,\alpha}(\mathbb{Q}) \le 1 + \widehat{M}_{t,p}(\mathbb{Q}), \qquad \mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n).$$

Consequently, the condition (3.58) from Assumption 3.14(d) is satisfied whenever for each $i \in \{1, 2\}$, all $t \in \mathbb{R}_+$ and all $\mathbb{Q}, \tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}_p(W_n)$,

$$|g_i(t, \mathbb{Q}) - g_i(t, \tilde{\mathbb{Q}})| \le \hat{c}_t(1 + \Phi_t(\mathbb{Q}))\widehat{\mathcal{W}}_{t,p}(\mathbb{Q}, \tilde{\mathbb{Q}}),$$

for some increasing function $\hat{c}\colon \mathbb{R}_+ \to \mathbb{R}_+$, where $\Phi\colon \mathbb{R}_+ \times \widehat{\mathcal{P}}_p(W_n) \to \mathbb{R}$ is an arbitrary linear combination of the functions in $\{\mathbb{R}_+ \times \widehat{\mathcal{P}}_p(W_n) \ni (t, \mathbb{Q}) \mapsto \widehat{M}_{t,\alpha}(\mathbb{Q}) : \alpha \in (0, p]\} \cup \{\mathbb{R}_+ \times \widehat{\mathcal{P}}_p(W_n) \ni (t, \mathbb{Q}) \mapsto \widehat{M}^{1/\alpha}_{t,\alpha}(\mathbb{Q}) : \alpha \in (0, p]\}$.

(b) Recall the Wasserstein distance $\mathcal{T}_{t,p}$ on the space $\mathcal{P}_p(W^t_n)$ from Definition 2.5. Let $Q \in \mathcal{P}_p(W^t_n)$ and denote by $\tilde{\delta}_0$ the Dirac measure at $0 \in W^t_n$. Observe that for every coupling $R \in \mathcal{P}(\tilde{\delta}_0, Q)$,

$$R(A \times B) = \mathbb{1}_A(0)Q(B), \qquad A, B \in \mathcal{B}_{W^t_n},$$

since if $0 \notin A$, then using that $A \times B \subseteq A \times W_n^t$, we have $R(A \times B) \leq R(A \times W_n^t) = \tilde{\delta}_0(A) = 0$, and if $0 \in A$, then $R(A^{\mathsf{c}} \times B) = 0$ by what we have just shown, hence $R(A \times B) = R(A \times B) + R(A^{\mathsf{c}} \times B) = R(W_n^t \times B) = Q(B)$. Consequently, $\mathcal{P}(\tilde{\delta}_0, Q) = \{\tilde{\delta}_0 \otimes Q\}$, which the definition (2.19) of $\mathcal{T}_{t,p}$ implies

$$\mathcal{T}_{t,p}(\tilde{\delta}_0, Q) = \int_{W_n^t \times W_n^t} p_t^p(w - \tilde{w})\, \mathrm{d}\tilde{\delta}_0 \otimes Q\,(w, \tilde{w}) = \int_{W_n^t} p_t^p(\tilde{w})\, \mathrm{d}Q\,(\tilde{w}).$$

With (2.21), the law of the unconscious statistician and (3.59), it follows that

$$\widehat{\mathcal{T}}_{t,p}(\mathbb{Q}, \delta_0) = \widehat{M}_{t,p}(\mathbb{Q}), \qquad \mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n), \tag{3.61}$$

where δ_0 denotes the Dirac measure at $0 \in W_n$.

Remark 3.17 (Properties of coefficients and certain classical SDEs obtained under Assumption 3.14)**.**

(a) By Lemma C.29 (b), (3.57) together with the assumed Borel-measurability of $f_1(\,\cdot\,, w)$ and $f_2(\,\cdot\,, w)$ for each $w \in W_n$ (see Assumption 3.14 (a)) implies that f_1 and f_2 are $\mathbb{G}$-progressive. Similarly, (3.58) implies, for each $t \in \mathbb{R}_+$, continuity[36] of $\widehat{\mathcal{P}}_p(W_n) \ni \mathbb{Q} \mapsto g(t, \mathbb{Q})$ w.r.t. the topology $\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}$ induced on $\widehat{\mathcal{P}}_p(W_n)$ by the pseudometric $\widehat{\mathcal{W}}_{t,p}$ defined in (2.21). Hence arguments analogous to those in the proof of Lemma 2.21 (b) show that the assumed $\mathcal{B}_{\mathbb{R}_+}$-measurability of $g_1(\,\cdot\,, \mathbb{Q})$ and $g_2(\,\cdot\,, \mathbb{Q})$ for each $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$ together with (3.58) implies $\mathbb{H}$-progressive measurability (see Definition 2.8) of g_1 and g_2. Together with the assumed $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$-progressive measurability of $\tilde{b}$ and $\tilde{\sigma}$, it follows that the coefficients b and σ given by (3.55) are $\mathbb{G} \otimes \mathbb{H} \otimes \mathbb{F}$-progressively measurable, ensuring that the stochastic integrals implicitly given by the associated McKean SDE are well-defined.

(b) Let $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$ and consider the $\mathbb{G} \otimes \mathbb{F}$-progressive coefficients

$$\begin{aligned} b^{\mathbb{Q}} &\colon \mathbb{R}_+ \times W_n \times \Omega \ni (t, w, \omega) \mapsto b(t, w, \mathbb{Q}, \omega) \in \mathbb{K}^n \quad \text{and} \\ \sigma^{\mathbb{Q}} &\colon \mathbb{R}_+ \times W_n \times \Omega \ni (t, w, \omega) \mapsto \sigma(t, w, \mathbb{Q}, \omega) \in \mathbb{K}^{n \times d}, \end{aligned}$$

defined by fixing $\mathbb{Q}$ in (3.55). For all $w, \tilde{w} \in W_n$, using that $\tilde{\sigma}$ has the Lipschitz property (see (2.49)) together with (3.56) and (3.57),[37]

$$\begin{aligned} |\sigma^{\mathbb{Q}}(t, w, \omega) - \sigma^{\mathbb{Q}}(t, \tilde{w}, \omega)| &\leq |\tilde{\sigma}(t, w, \mathbb{Q}, \omega) - \tilde{\sigma}(t, \tilde{w}, \mathbb{Q}, \omega)| \\ &\qquad + |(f_2(t, w) - f_2(t, \tilde{w}))g_2(t, \mathbb{Q})h_2(t, \omega)^{\mathsf{T}}| \\ &\leq c_t\, p_t(w - \tilde{w}) + \tilde{D}_t |f_2(t, w) - f_2(t, \tilde{w})||h_2(t, \omega)| \\ &\leq c_t\, p_t(w - \tilde{w}) + \tilde{c}_t \tilde{D}_t^2 p_t(w - \tilde{w}) \\ &= (c_t + \tilde{c}_t \tilde{D}_t^2)\, p_t(w - \tilde{w}). \end{aligned} \tag{3.62}$$

[36] The terms involving $\widehat{M}_{t,p}$ in (3.58) pose no problem, since by Theorem D.6 (a) $\Rightarrow$ (c), convergence of a sequence $(\mathbb{Q}^{(n)})_{n \in \mathbb{N}}$ in $\widehat{\mathcal{P}}_p(W_n)$ to a limit $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$ implies (see also (3.61)) that $\widehat{M}_{t,p}(\mathbb{Q}^{(n)}) \to \widehat{M}_{t,p}(\mathbb{Q})$.

[37] For the second inequality, we use also that the Frobenius norm of an outer product of two vectors is equal to the product of Euclidean norms of the vectors.

Exactly the same bound holds for $b^{\mathbb{Q}}$ instead of $\sigma^{\mathbb{Q}}$, hence $b^{\mathbb{Q}}$ and $\sigma^{\mathbb{Q}}$ have the Lipschitz property (Definition C.25) for classical SDEs, using instead of the function c in (C.15) the function

$$c'\colon \mathbb{R}_+ \to \mathbb{R}_+, \quad c'_t := 2(c_t + \tilde{c}_t \tilde{D}_t^2). \tag{3.63}$$

(c) Using that $\tilde{\sigma}$ satisfies the bounded growth condition (Definition 2.20) together with (3.56),

$$\begin{aligned} |\sigma^{\mathbb{Q}}(t,w,\omega)| &\le |\tilde{\sigma}(t,w,\mathbb{Q},\omega)| + |f_2(t,w)g_2(t,\mathbb{Q})h_2(t,\omega)^{\mathsf{T}}| \\ &\le D_t(1+p_t(w)) + \tilde{D}_t^3(1+p_t(w)) = (D_t+\tilde{D}_t^3)(1+p_t(w)). \end{aligned} \tag{3.64}$$

The same bound holds for $b^{\mathbb{Q}}$ instead of $\sigma^{\mathbb{Q}}$, hence the coefficients $b^{\mathbb{Q}}$ and $\sigma^{\mathbb{Q}}$ satisfy the bounded growth condition (Definition C.27) for classical SDEs, using instead of the function D in (C.17) the function

$$D'\colon \mathbb{R}_+ \to \mathbb{R}_+, \quad D'_t := 2(D_t + \tilde{D}_t^3). \tag{3.65}$$

(d) Fix $p \in [2,\infty)$, an initial distribution $\mu \in \mathcal{P}_p(\mathbb{K}^n)$ and a set-up $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P},B,\xi)$ (see Definition C.5 (b)) with $\mathcal{L}_{\mathbb{P}}[\xi] = \mu$. By parts (a) to (c) and Theorem C.34, there is, for every $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$, strong existence and pathwise uniqueness of solutions of the SDE

$$\mathrm{d}X_t = b^{\mathbb{Q}}(t,X,\,\cdot\,)\,\mathrm{d}t + \sigma^{\mathbb{Q}}(t,X,\,\cdot\,)\,\mathrm{d}W_t \tag{3.66}$$

for the initial distribution μ. In addition, defining the function $K'\colon \mathbb{R}_+ \to \mathbb{R}_+$ by (cf. (C.42))

$$K'_t := \beta(t)\max\Big\{(c'_t)^p, (D'_t)^p\,(1+M_p(\mu))\Big\}, \qquad t \in \mathbb{R}_+, \tag{3.67}$$

with the function M from (D.12), β from (C.22), each (pathwise unique) strong solution $X^{\mathbb{Q}}$ of (3.66) on the given set-up satisfies by Remark C.35, for every $t \in \mathbb{R}_+$,

$$\begin{aligned} \|p_t(X^{\mathbb{Q}})\|_{L^p(\mathbb{P})} &\le M_p^{1/p}(\mu) + \|p_t(X^{\mathbb{Q}} - \xi)\|_{L^p(\mathbb{P})} \\ &\le M_p^{1/p}(\mu) + \sum_{k=1}^{\infty}\Big(\frac{(K'_t t)^k}{k!}\Big)^{1/p} < \infty. \end{aligned} \tag{3.68}$$

In particular, $\mathcal{L}_{\mathbb{P}}[X^{\mathbb{Q}}] \in \widehat{\mathcal{P}}_p(W_n)$. Consider the function $M\colon \mathbb{R}_+ \to \mathbb{R}_+$ defined by

$$M_t = 2^{p-1}\Bigg(M_p(\mu) + \bigg(\sum_{k=1}^{\infty}\Big(\frac{(K'_t t)^k}{k!}\Big)^{1/p}\bigg)^p\Bigg), \qquad t \in \mathbb{R}_+. \tag{3.69}$$

Then using (3.59) and the law of the unconscious statistician for the equality and (3.68) together with Jensen's inequality (see (C.18)) for the bound, we have

$$\widehat{M}_{t,p}(\mathcal{L}[X^{\mathbb{Q}}]) = \mathbb{E}[p_t^p(X^{\mathbb{Q}})] \le M_t, \qquad (t,\mathbb{Q}) \in \mathbb{R}_+ \times \widehat{\mathcal{P}}_p(W_n). \tag{3.70}$$

Lemma 3.18. *Fix $p \in [2,\infty)$, a distribution $\mu \in \mathcal{P}_p(\mathbb{K}^n)$ and let*

$$\widehat{\mathcal{P}}'_p(W_n) := \{\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n) : \widehat{M}_{t,p}(\mathbb{Q}) \leq M_t \text{ for all } t \in \mathbb{R}_+\}, \tag{3.71}$$

with the function $\widehat{M}_{\cdot,p}$ from (3.59) and the function M from (3.69). Denote by $\mathcal{O}_{\widehat{\mathcal{P}}_p(W_n)}$ the Polish topology on $\widehat{\mathcal{P}}_p(W_n)$ from Proposition 2.7, which is induced by the metric $\widehat{\mathcal{W}}_p$ from (2.26). Then $\widehat{\mathcal{P}}'_p(W_n)$ is closed in $\widehat{\mathcal{P}}_p(W_n)$ w.r.t. $\mathcal{O}_{\widehat{\mathcal{P}}_p(W_n)}$ and in fact a Polish space when equipped with the restriction of $\widehat{\mathcal{W}}_p$.

Proof. Since the topology $\mathcal{O}_{\widehat{\mathcal{P}}_p(W_n)}$ is induced by the metric $\widehat{\mathcal{W}}_p$ from (2.26), it is first countable. Hence it suffices to check that for every sequence $(\mathbb{Q}^{(n)})_{n\in\mathbb{N}}$ in $\widehat{\mathcal{P}}'_p(W_n)$ converging to some limit $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$, we have in fact $\mathbb{Q} \in \widehat{\mathcal{P}}'_p(W_n)$. Fix $t \in \mathbb{R}_+$ and note that $\mathbb{Q}^{(n)} \xrightarrow[n\to\infty]{} \mathbb{Q}$ w.r.t. the metric $\widehat{\mathcal{W}}_p$ implies by (2.26) and Remark 2.6(e) that $\mathbb{Q}^{(n)} \xrightarrow[n\to\infty]{} \mathbb{Q}$ w.r.t. the pseudometric $\widehat{\mathcal{W}}_{t,p}$. This implies[38] that

$$\widehat{\mathcal{W}}_{t,p}(\mathbb{Q}^{(n)}, \delta_0) \xrightarrow[n\to\infty]{} \widehat{\mathcal{W}}_{t,p}(\mathbb{Q}, \delta_0). \tag{3.72}$$

From (2.20), it follows that $\widehat{\mathcal{T}}_{t,p}(\mathbb{Q}^{(n)}, \delta_0) \xrightarrow[n\to\infty]{} \widehat{\mathcal{T}}_{t,p}(\mathbb{Q}, \delta_0)$. With (3.61), using also that for each $n \in \mathbb{N}$, $\widehat{M}_{t,p}(\mathbb{Q}^{(n)}) \leq M_t$ since $\mathbb{Q}^{(n)} \in \widehat{\mathcal{P}}'_p(W_n)$, this implies

$$\widehat{M}_{t,p}(\mathbb{Q}) = \lim_{n\to\infty} \widehat{M}_{t,p}(\mathbb{Q}^{(n)}) \leq M_t.$$

Since the preceding display holds for every $t \in \mathbb{R}_+$, we have $\mathbb{Q} \in \widehat{\mathcal{P}}'_p(W_n)$. Therefore, $\widehat{\mathcal{P}}'_p(W_n)$ is closed w.r.t. $\mathcal{O}_{\widehat{\mathcal{P}}_p(W_n)}$. It follows that $\widehat{\mathcal{P}}'_p(W_n)$ is a Polish space when equipped with the restriction of the metric $\widehat{\mathcal{W}}_p$ to $\widehat{\mathcal{P}}'_p(W_n)$. □

Theorem 3.19 (Strong Existence and Pathwise Uniqueness for the McKean SDE under Assumption 3.14)**.** *Consider the setting of Assumption 3.14. For every initial distribution $\mu \in \mathcal{P}_p(\mathbb{K}^n)$ and every $\mathbb{R}^d$- or $\mathbb{K}^d$-valued standard Brownian motion B, there is strong existence and pathwise uniqueness (Definitions 2.26 and 2.29) for the McKean SDE (2.37) with coefficients b and σ as defined in (3.55). The solution process X satisfies $\mathcal{L}_{\mathbb{P}}[X] \in \widehat{\mathcal{P}}'_p(W_n)$ (see (3.71) and (3.69)), in particular $p_t(X)$ is p-integrable for each $t \in \mathbb{R}_+$.*

Furthermore, in the case of nonrandom dependence (Definition 2.12), there is uniqueness in law (Definition 2.28), and given an $\mathcal{F}_0$-measurable initial random vector ξ with law μ, the solution X with $X_0 \overset{\text{a.s.}}{=} \xi$ is progressive w.r.t. the smallest filtration $\widetilde{\mathbb{F}} = (\widetilde{\mathcal{F}}_t)_{t\in\mathbb{R}_+}$ generated by ξ and B and containing all $\mathbb{P}$-null sets of $\sigma(\xi, (B_t)_{t\in\mathbb{R}_+})$.

Proof. The proof is similar to the proof of Theorem 2.32, and we describe only the necessary modifications. The main difference is that the space $\widehat{\mathcal{P}}_p(W_n)$ in (2.69) is replaced in (3.74)

[38] We use here that the pseudometric $\widehat{\mathcal{W}}_{t,p} \colon \widehat{\mathcal{P}}_p(W_n) \times \widehat{\mathcal{P}}_p(W_n) \to \mathbb{R}_+$ is continuous w.r.t. the product $\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}} \otimes \mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}$ of the topology $\mathcal{O}_{\widehat{\mathcal{W}}_{t,p}}$ which it induces on W_n. This follows from the triangle inequality exactly as for metrics. Alternatively, (3.72) can also be obtained using Theorem D.6.

below by the space $\widehat{\mathcal{P}}'_p(W_n)$ defined in (3.71). This replacement is made possible by the observation from Remark 3.17(d), i.e. that the (laws of the) solutions $X^{\mathbb{Q}}$, $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$, of the classical SDEs from (3.66) satisfy the moment bound (3.70), which means that the image of the map Φ defined in (3.73) below is contained in $\widehat{\mathcal{P}}'_p(W_n)$ (see (3.71)). Consequently, the moments involving $\widehat{M}$ in (3.58) are bounded on the image of Φ, so the coefficients defined by (3.55) can effectively be assumed to be of the Lipschitz type already seen in Theorem 2.32, allowing to reuse the same arguments. The main technical complication is that obtaining the bound (3.75), which is required to apply the fixed point theorem A.16, is more laborious due to the more complicated form of the coefficients.

Step 1 (Reduction to an equivalent fixed point problem). Fix a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B)$ (see Definition C.5(b)) with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ and where the initial random vector ξ has law μ. For every $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$, there exists by Remark 3.17(d) a pathwise unique strong solution $X^{\mathbb{Q}}$ of (3.66). With the map

$$\Phi\colon \widehat{\mathcal{P}}_p(W_n) \to \widehat{\mathcal{P}}_p(W_n), \quad \mathbb{Q} \mapsto \mathcal{L}[X^{\mathbb{Q}}], \tag{3.73}$$

exactly the same arguments as in Step 1 of the proof of Theorem 2.32 show that it suffices to establish that Φ has a unique fixed point $\mathbb{Q}^*$ in $\widehat{\mathcal{P}}_p(W_n)$. For each $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$, $\mathcal{L}[X^{\mathbb{Q}}] \in \widehat{\mathcal{P}}'_p(W_n)$, see (3.71) and (3.70), hence Φ can have no fixed point in $\widehat{\mathcal{P}}_p(W_n) \setminus \widehat{\mathcal{P}}'_p(W_n)$. Therefore it suffices to show existence of a unique fixed point $\mathbb{Q}^*$ of the map

$$\Phi'\colon \widehat{\mathcal{P}}'_p(W_n) \to \widehat{\mathcal{P}}'_p(W_n), \quad \mathbb{Q} \mapsto \mathcal{L}[X^{\mathbb{Q}}]. \tag{3.74}$$

This is achieved using Theorem A.16 by verifying parts (i) and (ii) in Step 2 of the proof of Theorem 2.32, for which it suffices (see the discussion following (2.75)) in turn to show that for all $\mathbb{Q}, \tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}'_p(W_n)$,

$$\widehat{\mathcal{T}}_{t,p}(\Phi'(\mathbb{Q}), \Phi'(\tilde{\mathbb{Q}})) \le R_t \int_0^t \widehat{\mathcal{T}}_{s,p}(\mathbb{Q}, \tilde{\mathbb{Q}})\,\mathrm{d}s\,, \qquad t \in \mathbb{R}_+, \tag{3.75}$$

with $\widehat{\mathcal{T}}_{\cdot,p}$ from (2.21) and the increasing function $R\colon \mathbb{R}_+ \to \mathbb{R}_+$ defined in (3.90) below.

Step 2 (Establishing (3.75)). Fix $\mathbb{Q}, \tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}'_p(W_n)$ for now and define

$$h(t) := \mathbb{E}[p_t^p(X^{\mathbb{Q}} - X^{\tilde{\mathbb{Q}}})], \qquad t \in \mathbb{R}_+, \tag{3.76}$$

which is an increasing, $\mathbb{R}_+$-valued function due to the fact that $\mathcal{L}[X^{\mathbb{Q}}]$ and $\mathcal{L}[X^{\tilde{\mathbb{Q}}}]$ are in $\widehat{\mathcal{P}}'_p(W_n)$. Fix $t \in \mathbb{R}_+$ until (3.91) below. By construction, $X^{\mathbb{Q}}$ is a strong solution of (3.66), hence[39]

$$\begin{aligned} X^{\mathbb{Q}} \overset{\text{uti}}{=} &\overbrace{\xi + \int_0^{\cdot} \tilde{b}(s, X^{\mathbb{Q}}, \mathbb{Q}, \,\cdot\,)\,\mathrm{d}s + \int_0^{\cdot} \tilde{\sigma}(s, X^{\mathbb{Q}}, \mathbb{Q}, \,\cdot\,)\,\mathrm{d}B_s}^{=:\tilde{X}^{\mathbb{Q}}} \\ &+ \underbrace{\int_0^{\cdot} f_1(s, X^{\mathbb{Q}}) g_1(s, \mathbb{Q}) h_1(s, \,\cdot\,)\,\mathrm{d}s + \int_0^{\cdot} f_2(s, X^{\mathbb{Q}}) g_2(s, \mathbb{Q}) h_2(s, \,\cdot\,)^{\mathsf{T}}\,\mathrm{d}B_s}_{=:\hat{X}^{\mathbb{Q}}} \end{aligned} \tag{3.77}$$

[39] Due to Remark 3.17(a) and Lemma 2.15(b), the stochastic integrals in (3.77) and (3.78) are well-defined.

and similarly

$$
\begin{aligned}
X^{\tilde{\mathbb{Q}}} \overset{\mathrm{uti}}{=} & \overbrace{\xi + \int_0^\cdot \tilde{b}(s, X^{\tilde{\mathbb{Q}}}, \tilde{\mathbb{Q}}, \,\cdot\,)\,\mathrm{d}s + \int_0^\cdot \tilde{\sigma}(s, X^{\tilde{\mathbb{Q}}}, \tilde{\mathbb{Q}}, \,\cdot\,)\,\mathrm{d}B_s}^{=:\tilde{X}^{\tilde{\mathbb{Q}}}} \\
& + \underbrace{\int_0^\cdot f_1(s, X^{\tilde{\mathbb{Q}}}) g_1(s, \tilde{\mathbb{Q}}) h_1(s, \,\cdot\,)\,\mathrm{d}s + \int_0^\cdot f_2(s, X^{\tilde{\mathbb{Q}}}) g_2(s, \tilde{\mathbb{Q}}) h_2(s, \,\cdot\,)^{\mathsf{T}}\,\mathrm{d}B_s}_{=:\hat{X}^{\tilde{\mathbb{Q}}}}\,.
\end{aligned}
\tag{3.78}
$$

With the decompositions $X^{\mathbb{Q}} = \tilde{X}^{\mathbb{Q}} + \hat{X}^{\mathbb{Q}}$ and $X^{\tilde{\mathbb{Q}}} = \tilde{X}^{\tilde{\mathbb{Q}}} + \hat{X}^{\tilde{\mathbb{Q}}}$ defined by the previous displays, the triangle inequality for the seminorm p_t and Jensen's inequality for $x \mapsto x^p$ (see (C.18)) yield

$$
p_t^p(X^{\mathbb{Q}} - X^{\tilde{\mathbb{Q}}}) \le 2^{p-1}\big(p_t^p(\tilde{X}^{\mathbb{Q}} - \tilde{X}^{\tilde{\mathbb{Q}}}) + p_t^p(\hat{X}^{\mathbb{Q}} - \hat{X}^{\tilde{\mathbb{Q}}})\big),
$$

hence

$$
h(t) \le 2^{p-1}\big(\mathbb{E}\big[p_t^p(\tilde{X}^{\mathbb{Q}} - \tilde{X}^{\tilde{\mathbb{Q}}})\big] + \mathbb{E}\big[p_t^p(\hat{X}^{\mathbb{Q}} - \hat{X}^{\tilde{\mathbb{Q}}})\big]\big). \tag{3.79}
$$

Using (2.58) from Lemma 2.31 (a) with $\tilde{X} := \tilde{X}^{\mathbb{Q}}$ and $\tilde{Y} := \tilde{X}^{\tilde{\mathbb{Q}}}$,

$$
\mathbb{E}\big[p_t^p(\tilde{X}^{\mathbb{Q}} - \tilde{X}^{\tilde{\mathbb{Q}}})\big] \le \beta(t) \int_0^t c_s^p\, h(s)\,\mathrm{d}s + \beta(t) \int_0^t c_s^p \widehat{\mathcal{W}}_{s,p}^p(\mathbb{Q}, \tilde{\mathbb{Q}})\,\mathrm{d}s\,, \tag{3.80}
$$

where the function c originates from the Lipschitz property (Definition 2.19) and β was defined in (C.22). We next bound the second term on the right-hand side in (3.79). From the definition of $\hat{X}^{\mathbb{Q}}$ and $\hat{X}^{\tilde{\mathbb{Q}}}$, see (3.77) and (3.78), using also Jensen's inequality (see (C.18)),

$$
\begin{aligned}
&\mathbb{E}\big[p_t^p(\hat{X}^{\mathbb{Q}} - \hat{X}^{\tilde{\mathbb{Q}}})\big] \\
&\quad \le 2^{p-1}\mathbb{E}\bigg[p_t^p\bigg(\int_0^\cdot \big(f_1(s, X^{\mathbb{Q}}) g_1(s, \mathbb{Q}) - f_1(s, X^{\tilde{\mathbb{Q}}}) g_1(s, \tilde{\mathbb{Q}})\big) h_1(s, \,\cdot\,)\,\mathrm{d}s\bigg]\bigg) \\
&\qquad + 2^{p-1}\mathbb{E}\bigg[p_t^p\bigg(\int_0^\cdot \big(f_2(s, X^{\mathbb{Q}}) g_2(s, \mathbb{Q}) - f_2(s, X^{\tilde{\mathbb{Q}}}) g_2(s, \tilde{\mathbb{Q}})\big) h_2(s, \,\cdot\,)^{\mathsf{T}}\,\mathrm{d}B_s\bigg)\bigg] \\
&\quad = 2^{p-1}\bigg(\mathbb{E}\bigg[p_t^p\bigg(\int_0^\cdot U_s\,\mathrm{d}s\bigg)\bigg] + \mathbb{E}\big[p_t^p(V \cdot B)\big]\bigg),
\end{aligned}
\tag{3.81}
$$

where for all $(s, \omega) \in \mathbb{R}_+ \times \Omega$,

$$
\begin{aligned}
U_s(\omega) &:= \big(f_1(s, X^{\mathbb{Q}}) g_1(s, \mathbb{Q}) - f_1(s, X^{\tilde{\mathbb{Q}}}) g_1(s, \tilde{\mathbb{Q}})\big) h_1(s, \,\cdot\,) \quad \text{and} \\
V_s(\omega) &:= \big(f_2(s, X^{\mathbb{Q}}) g_2(s, \mathbb{Q}) - f_2(s, X^{\tilde{\mathbb{Q}}}) g_2(s, \tilde{\mathbb{Q}})\big) h_2(s, \,\cdot\,)^{\mathsf{T}},
\end{aligned}
\tag{3.82}
$$

which are $\mathbb{F}$-progressive by Remark 3.17 (a) and Lemma 2.15 (b). Then the same arguments as in the proof of Lemma 2.31 (a), see (2.64) and (2.65), show that

$$
\begin{aligned}
p_t^p\bigg(\int_0^\cdot U_s\,\mathrm{d}s\bigg) &\le t^{p-1} \int_0^t |U_s|^p\,\mathrm{d}s \quad \text{and} \\
\mathbb{E}\big[p_t^p(V \cdot B)\big] &\le C_p c_{\mathbb{K}}^{p/2} t^{p/2-1} \mathbb{E}\bigg[\int_0^t |V_s|^p\,\mathrm{d}s\bigg],
\end{aligned}
\tag{3.83}
$$

where $c_{\mathbb{K}} := \dim_{\mathbb{R}}(\mathbb{K})$ and C_p originates from the Burkholder–Davis–Gundy inequality [53, Theorem 6.32]. Combining (3.81) to (3.83), we obtain

$$\mathbb{E}\big[p_t^p\big(\hat{X}^{\mathbb{Q}} - \hat{X}^{\tilde{\mathbb{Q}}}\big)\big] \leq \underbrace{2^{p-1}(t^{p-1} + C_p c_{\mathbb{K}}^{p/2} t^{p/2-1})}_{=\beta(t)/3^{p-1}}\bigg(\int_0^t \mathbb{E}[|U_s|^p] + \mathbb{E}\,[|V_s|^p]\,\mathrm{d}s\bigg), \tag{3.84}$$

with the function β from (C.22), and where we have used Fubini's theorem to exchange expectation and integral, which is justified by $\mathbb{F}$-progressive measurability of U and V.

For all $s \in [0, t]$, we have using (3.82) and (3.56) for the first, introducing the term $f_1(s, X^{\tilde{\mathbb{Q}}})g_1(s, \mathbb{Q})$, using the triangle inequality and Jensen's inequality (see (C.18)) for the second, again (3.56) for the third, (3.57) together with (3.58) for the fourth and twice more Jensen's inequality together with (3.71) and the fact that $\mathbb{Q}, \tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}'_p(W_n)$ (see (3.71)) for the last inequality,

$$\begin{aligned}
|U_s|^p &\leq \tilde{D}_s^p \big|f_1(s, X^{\mathbb{Q}})g_1(s, \mathbb{Q}) - f_1(s, X^{\tilde{\mathbb{Q}}})g_1(s, \tilde{\mathbb{Q}})\big|^p \\
&\leq 2^{p-1}\tilde{D}_s^p\Big(\big|\big(f_1(s, X^{\mathbb{Q}}) - f_1(s, X^{\tilde{\mathbb{Q}}})\big)g_1(s, \mathbb{Q})\big|^p \\
&\qquad\qquad + \big|f_1(s, X^{\tilde{\mathbb{Q}}})\big(g_1(s, \mathbb{Q}) - g_1(s, \tilde{\mathbb{Q}})\big)\big|^p\Big) \\
&\leq 2^{p-1}\tilde{D}_s^{2p}\big|f_1(s, X^{\mathbb{Q}}) - f_1(s, X^{\tilde{\mathbb{Q}}})\big|^p \\
&\qquad\qquad + 2^{p-1}\tilde{D}_s^{2p}\big(1 + p_s(X^{\tilde{\mathbb{Q}}})\big)^p |g_1(s, \mathbb{Q}) - g_1(s, \tilde{\mathbb{Q}})|^p \\
&\leq 2^{p-1}\tilde{c}_s^p\tilde{D}_s^{2p}p_s^p(X^{\mathbb{Q}} - X^{\tilde{\mathbb{Q}}}) \\
&\qquad\qquad + 2^{p-1}\tilde{c}_s^p\tilde{D}_s^{2p}\big(1 + p_s(X^{\tilde{\mathbb{Q}}})\big)^p\Big(1 + \widehat{M}_{s,p}(\mathbb{Q}) + \widehat{M}_{s,p}(\tilde{\mathbb{Q}})\Big)^p \widehat{\mathcal{W}}_{s,p}^p(\mathbb{Q}, \tilde{\mathbb{Q}}) \\
&\leq 12^{p-1}\Big(\tilde{c}_s^p\tilde{D}_s^{2p}p_s^p(X^{\mathbb{Q}} - X^{\tilde{\mathbb{Q}}}) + \tilde{c}_s^p\tilde{D}_s^{2p}\big(1 + p_s^p(X^{\tilde{\mathbb{Q}}})\big)\big(1 + 2M_s^p\big)\widehat{\mathcal{W}}_{s,p}^p(\mathbb{Q}, \tilde{\mathbb{Q}})\Big) \\
&= 12^{p-1}\tilde{c}_s^p\tilde{D}_s^{2p}\Big(p_s^p(X^{\mathbb{Q}} - X^{\tilde{\mathbb{Q}}}) + \big(1 + p_s^p(X^{\tilde{\mathbb{Q}}})\big)\big(1 + 2M_s^p\big)\widehat{\mathcal{W}}_{s,p}^p(\mathbb{Q}, \tilde{\mathbb{Q}})\Big).
\end{aligned} \tag{3.85}$$

Due to the similar structure of the second terms on the right-hand sides in (3.55) and the fact that the Frobenius norm of an outer product of two vectors is the product of Euclidean norms of those vectors, exactly the same bound as in the previous display applies to V, i.e.

$$|V_s|^p \leq 12^{p-1}\tilde{c}_s^p\tilde{D}_s^{2p}\Big(p_s^p(X^{\mathbb{Q}} - X^{\tilde{\mathbb{Q}}}) + \big(1 + p_s^p(X^{\tilde{\mathbb{Q}}})\big)\big(1 + 2M_s^p\big)\widehat{\mathcal{W}}_{s,p}^p(\mathbb{Q}, \tilde{\mathbb{Q}})\Big). \tag{3.86}$$

for all $(s, \omega) \in \mathbb{R}_+ \times \Omega$. Combining (3.84), (3.85) and (3.86) for the first and using that $\mathbb{E}\big[p_s^p(X^{\tilde{\mathbb{Q}}})\big] \leq M_s$ by (3.70) for the second inequality,

$$\begin{aligned}
&\mathbb{E}\big[p_t^p\big(\hat{X}^{\mathbb{Q}} - \hat{X}^{\tilde{\mathbb{Q}}}\big)\big] \\
&\leq 2 \cdot 4^{p-1}\beta(t)\int_0^t \tilde{c}_s^p\tilde{D}_s^{2p}\Big(\overbrace{\mathbb{E}\big[p_s^p(X^{\mathbb{Q}} - X^{\tilde{\mathbb{Q}}})\big]}^{=h(s)} + \big(1 + \overbrace{\mathbb{E}\big[p_s^p(X^{\tilde{\mathbb{Q}}})\big]}^{\leq M_s}\big)(1 + 2M_s^p)\widehat{\mathcal{W}}_{s,p}^p(\mathbb{Q}, \tilde{\mathbb{Q}})\Big)\,\mathrm{d}s \\
&\leq 2^{2p-1}\beta(t)\int_0^t \tilde{c}_s^p\tilde{D}_s^{2p}\Big(h(s) + (1 + M_s)(1 + 2M_s^p)\widehat{\mathcal{W}}_{s,p}^p(\mathbb{Q}, \tilde{\mathbb{Q}})\Big)\,\mathrm{d}s\,.
\end{aligned}$$

From (3.79), (3.80) and the previous display,

$$
\begin{aligned}
h(t) &\leq 2^{3p-2}\beta(t)\int_0^t (c_s^p + \tilde{c}_s^p\tilde{D}_s^{2p})h(s)\,\mathrm{d}s \\
&\quad + 2^{3p-2}\beta(t)\int_0^t (c_s^p + \tilde{c}_s^p\tilde{D}_s^{2p}(1+M_s)(1+2M_s^p))\widehat{\mathcal{W}}_{s,p}^p(\mathbb{Q},\tilde{\mathbb{Q}})\,\mathrm{d}s \\
&\leq 2^{3p-2}\beta(t)\int_0^t (c_s^p + \tilde{c}_s^p\tilde{D}_s^{2p})h(s)\,\mathrm{d}s \\
&\quad + \beta(t)\,\underbrace{2^{3p-2}(c_t^p + \tilde{c}_t^p\tilde{D}_t^{2p}(1+M_s)(1+2M_s^p))}_{=:\hat{c}_t}\int_0^t \widehat{\mathcal{W}}_{s,p}^p(\mathbb{Q},\tilde{\mathbb{Q}})\,\mathrm{d}s \qquad (3.87)\\
&\leq \beta(t)\int_0^t \hat{c}_s h(s)\,\mathrm{d}s + \underbrace{\beta(t)\hat{c}_t\int_0^t \widehat{\mathcal{W}}_{s,p}^p(\mathbb{Q},\tilde{\mathbb{Q}})\,\mathrm{d}s}_{=:\alpha(t)} \\
&= \alpha(t) + \beta(t)\int_0^t \hat{c}_s h(s)\,\mathrm{d}s\,,
\end{aligned}
$$

where for each $s \in \mathbb{R}_+$,

$$
\alpha(s) := \beta(s)\hat{c}_s\int_0^s \widehat{\mathcal{W}}_{u,p}^p(\mathbb{Q},\tilde{\mathbb{Q}})\,\mathrm{d}u \quad \text{and} \quad \hat{c}_s := 2^{3p-2}(c_s^p + \tilde{c}_s^p\tilde{D}_s^{2p}(1+M_s)(1+2M_s^p)). \qquad (3.88)
$$

By Grönwall's inequality ([53, Theorem 8.21 and Remark 8.24 (c)]), using also that by (2.20) and (2.21), $\widehat{\mathcal{W}}_{s,p}^p(\mathbb{Q},\tilde{\mathbb{Q}}) = \widehat{\mathcal{T}}_{s,p}(\mathbb{Q},\tilde{\mathbb{Q}})$, (3.87) implies that

$$
h(t) \leq \alpha(t)\exp\bigl(t(\beta(t)\vee\hat{c}_t)^2\bigr) = R_t\int_0^t \widehat{\mathcal{T}}_{s,p}(\mathbb{Q},\tilde{\mathbb{Q}})\,\mathrm{d}s\,, \qquad (3.89)
$$

where[40]

$$
R_s := \beta(s)\hat{c}_s\exp\bigl(s(\beta(s)\vee\hat{c}_s)^2\bigr), \qquad s\in\mathbb{R}_+, \qquad (3.90)
$$

which defines an $\mathbb{R}_+$-valued increasing function. Recall the notation $\mathcal{P}(\mu,\nu)$ for the set of couplings of two probability measures μ and ν (Definition D.2). Since $\mathcal{L}_{\mathbb{P}}[(X^{\mathbb{Q}}, X^{\tilde{\mathbb{Q}}})] \in \mathcal{P}(\Phi'(\mathbb{Q}),\Phi'(\tilde{\mathbb{Q}}))$, we have $\mathcal{L}_{\mathbb{P}}[((X_s^{\mathbb{Q}})_{s\in[0,t]}, (X^{\tilde{\mathbb{Q}}_s})_{s\in[0,t]})] \in \mathcal{P}(\Phi'(\mathbb{Q})_{[0,t]},\Phi'(\tilde{\mathbb{Q}})_{[0,t]})$. Consequently, by (2.21) from Definition 2.5,

$$
\widehat{\mathcal{T}}_{t,p}(\Phi'(\mathbb{Q}),\Phi'(\tilde{\mathbb{Q}})) \leq \mathbb{E}\Bigl[p_t^p(X^{\mathbb{Q}} - X^{\tilde{\mathbb{Q}}})\Bigr] = h(t). \qquad (3.91)
$$

Combining (3.89) and (3.91), which both hold for all $t\in\mathbb{R}_+$, (3.75) follows. □

Remark 3.20 (Propagation of chaos and the convergence of the particle approximation under Assumption 3.14)**.** The fact that the proof of Theorem 2.32 can be adapted with relative ease to the setting of Assumption 3.14 makes it plausible that both the particle

[40] From the definitions (C.22), (3.69), (3.67) and (3.88) of the functions β, M, K' and $\hat{c}$, we see that R_t depends apart from t only on the order p, the absolute p-th moment of the initial distribution and the constants c_t (from (2.49)), D_t (from (2.50)), $\tilde{c}_t$ and $\tilde{D}_t$ (from (3.56) and (3.57))

approximation and propagation of chaos (see Theorems 2.56 and 2.60) should hold also in this setting. Unfortunately, the continuity property (3.58) of the functions g_1 and g_2 in (3.55) preclude obtaining bounds identical to the ones in the proof of Theorem 2.56. For example, defining the process $\tilde{X}^1$ as in (2.111) but with the coefficients b and σ as in (3.55), (2.113) no longer follows, since the term involving the Wasserstein distance of the empirical measures μ^N and $\tilde{\mu}^N$ is multiplied with $\big(1 + \widehat{M}_{t,p}(\mu^N) + \widehat{M}_{t,p}(\tilde{\mu}^N)\big)^p$. We conjecture that convergence of the particle approximation and the propagation of chaos hold also under Assumption 3.14, but do not pursue a proof in this thesis.

The next example, which generalizes ideas[41] from [17, §2.1], motivates the form of the coefficients in (3.55). SDEs involving terms as in (3.92) are of central importance in the mean-field LMM, which is described in Subsection 3.2.3 below.

Example 3.21 (Variance-based volatility dampening)**.** Let $\alpha\colon \mathbb{R}_+ \to \mathbb{R}_+$ and $\beta\colon \mathbb{R}_+ \to \mathbb{R}_+$ be Borel-measurable with α locally bounded, and define $g\colon \mathbb{R}_+ \times \widehat{\mathcal{P}}_2(W_n) \to \mathbb{R}_+$ by

$$g(t,\mathbb{Q}) := \exp(-\alpha_t(\mathbb{V}[\mathbb{Q}_t] - \beta_t)^+), \qquad (t,\mathbb{Q}) \in \mathbb{R}_+ \times \widehat{\mathcal{P}}_2(W_n), \tag{3.92}$$

where we used the notation (2.17) and

$$\mathbb{V}[\mu] := \int_{\mathbb{K}^n} |z|^2 \, \mathrm{d}\mu(z) - \left| \int_{\mathbb{K}^n} z \, \mathrm{d}\mu(z) \right|^2, \qquad \mu \in \mathcal{P}_1(\mathbb{K}^n), \tag{3.93}$$

denotes the variance of measures on $\mathbb{K}^n$.[42] If at time t, $\mathbb{V}[\mathbb{Q}_t]$ exceeds the threshold value β_t, then $g(t,\mathbb{Q})$ will be, provided $\alpha_t > 0$, less than one, and the more $\mathbb{V}[\mathbb{Q}_t]$ overshoots β_t, the closer $g(t,\mathbb{Q})$ will be to zero. The function α governs the strength of this dampening. In Figure 3.3, $\mathbb{R}_+ \ni v \mapsto \hat{g}_{\alpha,\beta}(v) := \exp(-\alpha(v-\beta)^+)$ is plotted against v with constant values of α and β. By Lemma 3.22 below, g has the continuity property (3.58) required in Assumption 3.14(d).[43]

The following lemma generalizes an observation made in [17, Remark 5.3].

Lemma 3.22. *Let $\alpha\colon \mathbb{R}_+ \to \mathbb{R}_+$ and $\beta\colon \mathbb{R}_+ \to \mathbb{R}_+$ be Borel-measurable and define $g\colon \mathbb{R}_+ \times \widehat{\mathcal{P}}_2(W_n) \to \mathbb{R}_+$ as in (3.92) (see also Convention 3.15). Then for all $\mathbb{Q}, \tilde{\mathbb{Q}} \in \widehat{\mathcal{P}}_2(W_n)$,*

$$|g(t,\mathbb{Q}) - g(t,\tilde{\mathbb{Q}})| \le 3\alpha_t(M_2^{1/2}(\mathbb{Q}_t) + M_2^{1/2}(\tilde{\mathbb{Q}}_t))d_2(\mathbb{Q}_t,\tilde{\mathbb{Q}}_t). \tag{3.94}$$

where d_2 denotes the second-order Wasserstein metric on $\mathbb{K}^n$ (Definitions 2.4 and D.2) and $M_2(\,\cdot\,)$ denotes the second absolute moment of a measure, see (D.12). In particular, if α is locally bounded, then there exists an increasing function $\tilde{\alpha}\colon \mathbb{R}_+ \to \mathbb{R}_+$ with $\alpha \le \tilde{\alpha}$, and for every such function $\tilde{\alpha}$, g satisfies the continuity condition (3.58) with $\tilde{c} = 6\tilde{\alpha}$.

[41] In [17, §2.1], the function $g(t,\mathbb{Q})$ from (3.92) is considered for $\mathbb{K} = \mathbb{R}$, $n = 1$ and with the functions α and β replaced by constants.

[42] Let $Z\colon \mathbb{R}^n \to \mathbb{R}^n$ denote the identity on $\mathbb{K}^n$. By the componentwise definition of the expectation of $\mathbb{K}^n$-valued random vectors, $\mathbb{V}[\mu] = \sum_{k=1}^n \mathbb{V}[\mathcal{L}_\mu[\mathrm{Re}(Z_k)]] + \mathbb{V}[\mathcal{L}_\mu[\mathrm{Im}(Z_k)]]$.

[43] That g also satisfies the measurability property required in Assumption 3.14(a) is the content of Footnote 45.

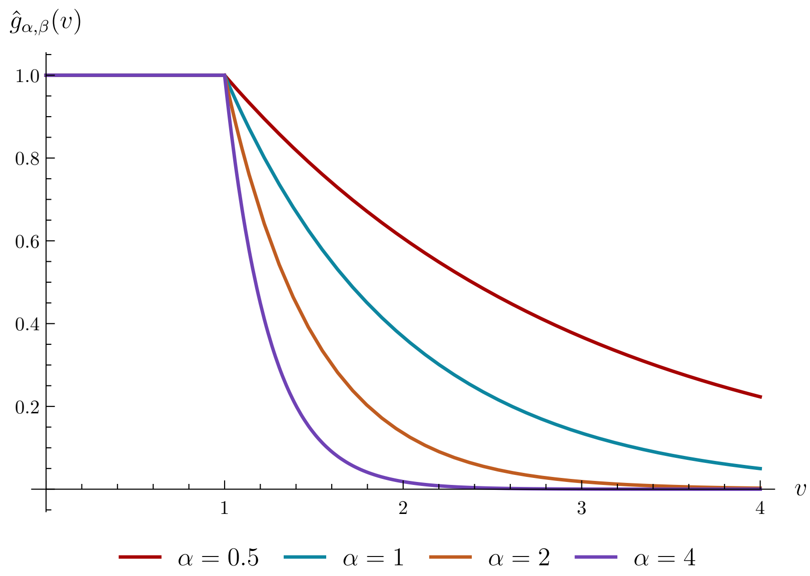

Figure 3.3. The function $\mathbb{R}_+ \ni v \mapsto \hat{g}_{\alpha,\beta}(v) := \exp(-\alpha(v-\beta)^+)$ governing the variance-dampening effect of the function g in (3.92), plotted vs. v for dampening strengths $\alpha \in \{0.5, 1, 2, 4\}$ dampening threshold $\beta = 1$.

For the proof of Lemma 3.22, the following lemma, which is based on ideas from [17, Remark 5.3], is useful.[44]

Lemma 3.23 (Continuity of the variance w.r.t. the Wasserstein metric)**.** *Let μ and ν be probability measures in $\mathcal{P}_2(\mathbb{K}^n)$. Then*

$$|\mathbb{V}[\mu] - \mathbb{V}[\nu]| \le 3(M_2^{1/2}(\mu) + M_2^{1/2}(\nu))d_2(\mu, \nu), \tag{3.95}$$

where d_2 denotes the second-order Wasserstein metric on $\mathbb{K}^n$ (Definitions 2.4 and D.2) and $M_2(\,\cdot\,)$ denotes the second absolute moment of a measure, see (D.12).[45]

Proof. Using the notation (D.12), we will show first[46] that for all $p \in [1, \infty)$,

$$\left|M_p^{1/p}(\mu) - M_p^{1/p}(\nu)\right| \le d_p(\mu, \nu), \tag{3.96}$$

where d_p denotes the p-th order Wasserstein metric from Definition 2.4. By the reverse triangle inequality for d_p,

$$|d_p(\mu, \delta_0) - d_p(\delta_0, \nu)| \le d_p(\mu, \nu),$$

[44] We give a proof different from the one in [17, Remark 5.3], which contains as a byproduct the moment inequality (3.96).

[45] Continuity of $\mathcal{P}_2(\mathbb{K}^n) \ni \mu \mapsto \mathbb{V}[\mu] \in \mathbb{R}_+$ follows with (3.95) from Theorem D.6 (a) $\Rightarrow$ (c), or alternatively directly from Theorem D.6 (a) $\Rightarrow$ (d). Together with Lemma 2.10, it follows that the process $\mathbb{R}_+ \times \widehat{\mathcal{P}}_p(W_n) \ni (t, \mathbb{Q}) \mapsto \mathbb{V}[\mathbb{Q}_t]$ is continuous and $\mathbb{H}$-progressive, in particular each section $\mathbb{R}_+ \ni t \mapsto \mathbb{V}[\mathbb{Q}_t]$ is Borel-measurable. This, together with Borel-measurability of α and β ensures that the function g defined in (3.92) is such that $g(\,\cdot\,, \mathbb{Q})$ is Borel-measurable for each $\mathbb{Q} \in \widehat{\mathcal{P}}_p(W_n)$, as demanded in Assumption 3.14 (a). In fact, if α and β are continuous, then $g(\,\cdot\,, \mathbb{Q})$ is continuous as a composition of continuous functions.

[46] The bound (3.96) holds also when $\mathbb{K}^n$ is replaced by a Banach space $(X, \|\cdot\|)$ if the absolute value in (3.96) is replaced by the norm, and d_p is replaced by the Wasserstein metric $\mathcal{W}_p$ on $\mathcal{P}_p(X)$, see Definition D.2.

where δ_0 denotes the Dirac measure at $0 \in \mathbb{K}^n$. Using arguments entirely analogous to those in Remark 3.16(b), $\mathcal{P}(\mu, \delta_0) = \{\mu \otimes \delta_0\}$ and similarly $\mathcal{P}(\delta_0, \nu) = \{\delta_0 \otimes \nu\}$. By the definition (D.2) of the Wasserstein metric, $d_p(\delta_0, \mu) = \big(\int_{\mathbb{K}^n} |x|^p \,\mathrm{d}\mu(x)\big)^{1/p} = M_p^{1/p}(\mu)$ and similarly for ν. This shows (3.96). Let $a := M_p^{1/p}(\mu)$ and $b := M_p^{1/p}(\nu)$ such that $M_p(\mu) = a^p$ and $M_p(\nu) = b^p$. By the mean value theorem of differential calculus,

$$\big|M_p(\mu) - M_p(\nu)\big| = |a^p - b^p| = |a - b|\, p\,\xi^{p-1},$$

for some number ξ between $a \wedge b$ and $a \vee b$. Since $\mathbb{R}_+ \ni x \mapsto x^{p-1}$ is increasing, $\xi^{p-1} \le a^{p-1} \vee b^{p-1} = M_p^{(p-1)/p}(\mu) \vee M_p^{(p-1)/p}(\nu)$. By (3.96), $|a - b| \le d_p(\mu, \nu)$. Consequently,

$$\big|M_p(\mu) - M_p(\nu)\big| \le p\, d_p(\mu, \nu)\,\big(M_p^{(p-1)/p}(\mu) \vee M_p^{(p-1)/p}(\nu)\big). \tag{3.97}$$

For $p = 2$, we have $|M_2(\mu) - M_2(\nu)| \le 2\, d_2(\mu, \nu)\big(M_2^{1/2}(\mu) \vee M_2^{1/2}(\nu)\big)$. Let Z denote the identity map on $\mathbb{K}^n$. From (3.93) and (3.97),

$$\begin{aligned} |\mathbb{V}[\mu] - \mathbb{V}[\nu]| &= |M_2(\mu) - |\mathbb{E}_\mu[Z]|^2 - (M_2(\nu) - |\mathbb{E}_\nu[Z]|^2)| \\ &\le |M_2(\mu) - M_2(\nu)| + ||\mathbb{E}_\mu[Z]|^2 - |\mathbb{E}_\nu[Z]|^2| \\ &\le 2d_2(\mu, \nu)\big(M_2^{1/2}(\mu) \vee M_2^{1/2}(\nu)\big) + ||\mathbb{E}_\mu[Z]|^2 - |\mathbb{E}_\nu[Z]|^2|. \end{aligned} \tag{3.98}$$

To bound the second term in the previous display, observe that for every coupling $\pi \in \mathcal{P}(\mu, \nu)$,

$$|\mathbb{E}_\mu[Z] - \mathbb{E}_\nu[Z]| = \left|\int_{\mathbb{K}^n} x - y \,\mathrm{d}\pi(x, y)\right| \le \int_{\mathbb{K}^n} |x - y| \,\mathrm{d}\pi(x, y)\,. \tag{3.99}$$

Taking the infimum over all couplings, $|\mathbb{E}_\mu[Z] - \mathbb{E}_\nu[Z]| \le d_1(\mu, \nu)$ follows from the definition (D.2) of the Wasserstein metric. Using the reverse triangle inequality for the Frobenius norm for the first, (3.99) for the second inequality and Remark D.4 with the fact that $M_1(\rho) \le M_2^{1/2}(\rho)$ for each $\rho \in \mathcal{P}_1(\mathbb{K}^n)$,

$$\begin{aligned} \big||\mathbb{E}_\mu[Z]|^2 - |\mathbb{E}_\nu[Z]|^2\big| &= |(|\mathbb{E}_\mu[Z]| - |\mathbb{E}_\nu[Z]|)(|\mathbb{E}_\mu[Z]| + |\mathbb{E}_\nu[Z]|)| \\ &\le |\mathbb{E}_\mu[Z] - \mathbb{E}_\nu[Z]|(\mathbb{E}_\mu[|Z|] + \mathbb{E}_\nu[|Z|]) \\ &\le d_1(\mu, \nu)(M_1(\mu) + M_1(\nu)) \\ &\le d_2(\mu, \nu)(M_2^{1/2}(\mu) + M_2^{1/2}(\nu)). \end{aligned}$$

Combining the previous display with (3.98), the bound (3.95) follows. □

Proof of Lemma 3.22, [17, Remark 5.3]. Since the function $\mathbb{R}_+ \ni x \mapsto \mathrm{e}^{-x}$ is Lipschitz continuous[47] with Lipschitz constant 1, we have

$$|\exp(-x^+) - \exp(-y^+)| \le |x^+ - y^+| \le |x - y|, \qquad x, y \in \mathbb{R}. \tag{3.100}$$

With the functions α and β from Example 3.21, let $q\colon \mathbb{R}_+ \times \widehat{\mathcal{P}}_2(W_n) \to \mathbb{R}_+$ be defined by

$$q(t, \mathbb{Q}) = \alpha_t(\mathbb{V}[\mathbb{Q}_t] - \beta_t), \qquad (t, \mathbb{Q}) \in \mathbb{R}_+ \times \widehat{\mathcal{P}}_2(W_n), \tag{3.101}$$

[47] This follows from the mean value theorem of differential calculus and the fact that $\mathbb{R}_+ \ni x \mapsto \mathrm{e}^{-x}$ has a derivative not exceeding one in absolute value on $\mathbb{R}_+$.

such that the map g from (3.92) satisfies

$$g(t,\mathbb{Q}) = \exp(-q(t,\mathbb{Q})^+), \qquad (t,\mathbb{Q}) \in \mathbb{R}_+ \times \widehat{\mathcal{P}}_2(W_n).$$

From the previous display and (3.100) for the first and (3.95) for the second inequality,

$$\begin{aligned}|g(t,\mathbb{Q}) - g(t,\tilde{\mathbb{Q}})| &\le |q(t,\mathbb{Q}) - q(t,\tilde{\mathbb{Q}})|\\ &\le 3\alpha_t(M_2^{1/2}(\mathbb{Q}_t) + M_2^{1/2}(\tilde{\mathbb{Q}}_t))d_2(\mathbb{Q}_t,\tilde{\mathbb{Q}}_t).\end{aligned}$$

This establishes (3.94). Combining the previous display with the inequalities $d_2(\mathbb{Q}_t,\tilde{\mathbb{Q}}_t) \le \widehat{\mathcal{W}}_{t,2}(\mathbb{Q},\tilde{\mathbb{Q}})$ (see (2.25)), $M_2(\mathbb{Q}_t) \le \widehat{M}_{t,2}(\mathbb{Q})$ and $M_2(\tilde{\mathbb{Q}}_t) \le \widehat{M}_{t,2}(\tilde{\mathbb{Q}})$ (see (D.12) and (3.59)) for the first and (3.60) for the second inequality,

$$\begin{aligned}|g(t,\mathbb{Q}) - g(t,\tilde{\mathbb{Q}})| &\le 3\alpha_t\big(\widehat{M}_{t,2}^{1/2}(\mathbb{Q}) + \widehat{M}_{t,2}^{1/2}(\tilde{\mathbb{Q}})\big)\widehat{\mathcal{W}}_{t,2}(\mathbb{Q},\tilde{\mathbb{Q}})\\ &\le 3\alpha_t\big(2 + \widehat{M}_{t,2}(\mathbb{Q}) + \widehat{M}_{t,2}(\tilde{\mathbb{Q}})\big)\widehat{\mathcal{W}}_{t,2}(\mathbb{Q},\tilde{\mathbb{Q}}),\\ &\le 6\alpha_t\big(1 + \widehat{M}_{t,2}(\mathbb{Q}) + \widehat{M}_{t,2}(\tilde{\mathbb{Q}})\big)\widehat{\mathcal{W}}_{t,2}(\mathbb{Q},\tilde{\mathbb{Q}}),\end{aligned}$$

i.e. g satisfies (3.58) with $\tilde{c} = 6\tilde{\alpha}$ for every increasing function[48] $\tilde{\alpha}\colon \mathbb{R}_+ \to \mathbb{R}_+$ with $\alpha \le \tilde{\alpha}$. □

We conclude this section with concrete examples of SDEs involving variance-based volatility dampening, see Example 3.21.

Example 3.24 (Variance-based volatility dampening in arithmetic[49] Brownian motion)**.** Let $\mu\colon \mathbb{R}_+ \to \mathbb{R}^n$ and $\sigma\colon \mathbb{R}_+ \to \mathbb{R}^{n\times d}$ be continuous functions. Fix an $\mathbb{R}^d$-valued standard Brownian motion $B = (B_t)_{t\in\mathbb{R}_+}$ defined on a stochastic base (Definition C.5(a)) $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$. We consider the SDE

$$\mathrm{d}X_t = \mu_t\,\mathrm{d}t + g(\mathcal{L}[X_t])\sigma_t\,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+, \tag{3.102}$$

for an $\mathbb{R}^n$-valued solution with initial random vector $X_0 \overset{\text{a.s.}}{=} 0$, where g is as in (3.92) with α and β assumed to be nonnegative constants, and we follow Convention 3.15 to write g as a function of $\mathcal{L}[X_t]$ instead of $(t,\mathcal{L}[X])$. By the assumed continuity of μ and σ and Lemma 3.22, the coefficients in (3.102) satisfy Assumption 3.14. Therefore, by Theorem 3.19, there is uniqueness, pathwise and in law, as well as strong existence of solutions of (3.102), and every solution is a square integrable process. Let $X = (X_t)_{t\in\mathbb{R}_+}$ be a strong solution. Since $\mathbb{R}_+ \ni t \mapsto \tilde{g}_t := g(\mathcal{L}[X_t]) \in \mathbb{R}_+$ is continuous (see Footnote 45), the coefficients in (3.102) can be viewed as deterministic and continuous functions of time only. Our next aim is to determine the law of X, for which it suffices to show that X has independent increments, and to determine the distribution of those increments. By (3.102), X can be decomposed as

$$X \overset{\text{uti}}{=} \int_0^{\cdot} \mu_s\,\mathrm{d}s + Y, \tag{3.103}$$

[48] One such function is given by the running supremum, $\tilde{\alpha}_t := \sup_{s\in[0,t]} \alpha_s$, $t \in \mathbb{R}_+$, which is real-valued by local boundedness of α and Borel-measurable as it is increasing.

[49] We use the term arithmetic Brownian motion for any strong solution of the SDE $\mathrm{d}X_t = \mu_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t$, where μ and σ are deterministic and locally integrable respectively square integrable.

with the continuous local martingale $Y := \int_0^\cdot \tilde{g}_s\sigma_s \,\mathrm{d}B_s$. By [53, Exercise 5.116 (c)], $[X] \overset{\text{uti}}{=} [Y] \overset{\text{uti}}{=} \int_0^\cdot \tilde{g}_s^2\sigma_s\sigma_s^\mathsf{T}\,\mathrm{d}s$, which is deterministic. From [53, Theorem 7.1], Y has $\mathbb{F}$-independent increments, and

$$\mathcal{L}[Y_t - Y_s] = \mathcal{N}\big(0, [Y]_t - [Y]_s\big) = \mathcal{N}\bigg(0, \int_s^t \tilde{g}_u^2\sigma_u\sigma_u^\mathsf{T}\,\mathrm{d}u\bigg), \qquad s \le t \text{ in } \mathbb{R}_+. \tag{3.104}$$

Since μ is deterministic, (3.103) implies that X has $\mathbb{F}$-independent increments and

$$\mathcal{L}[X_t - X_s] = \mathcal{N}\bigg(\int_s^t \mu_u\,\mathrm{d}u\,, \int_s^t \tilde{g}_u^2\sigma_u\sigma_u^\mathsf{T}\,\mathrm{d}u\bigg), \qquad s \le t \text{ in } \mathbb{R}_+.$$

By definition, $\tilde{g}_u = g(\mathcal{L}[X_u]) = \exp(-\alpha(\mathbb{V}[X_u] - \beta)^+)$, $u \in \mathbb{R}_+$, (see (3.92)) hence the distribution of X is determined through the previous display by μ, σ and the curve $\mathbb{R}_+ \ni u \mapsto \mathbb{V}[X_u]$. By (3.103), (3.104) and the definition (3.93) of the variance of $\mathbb{R}^n$-valued random variables (see also Footnote 42),

$$\varphi_t := \mathbb{V}[X_t] = \mathbb{V}[Y_t] = \sum_{i=1}^n \mathbb{V}[Y_t^i] = \int_0^t \tilde{g}_u^2\,\mathrm{tr}(\sigma_u\sigma_u^\mathsf{T})\,\mathrm{d}u = \int_0^t \tilde{g}_u^2|\sigma_u|^2\,\mathrm{d}u\,, \qquad t \in \mathbb{R}_+,$$

where we used the definition of the Frobenius norm for the last equality. Since both $\tilde{g}$ and σ are continuous, the fundamental theorem of calculus implies that

$$\frac{\mathrm{d}\varphi_t}{\mathrm{d}t} = \tilde{g}_t^2|\sigma_t|^2 = \exp(-2\alpha(\varphi_t - \beta)^+)|\sigma_t|^2, \qquad t \in \mathbb{R}_+.$$

Multiplying by the reciprocal of the exponential factor and integrating w.r.t. t and using the substitution rule for integration, we obtain

$$\int_0^t \exp(2\alpha(\varphi_u - \beta)^+)\frac{\mathrm{d}\varphi_u}{\mathrm{d}u}\,\mathrm{d}u = \int_0^{\varphi(t)} \exp(2\alpha(x - \beta)^+)\,\mathrm{d}x = \int_0^t |\sigma_u|^2\,\mathrm{d}u\,, \tag{3.105}$$

for every $t \in \mathbb{R}_+$. We have

$$F(y) := \int_0^y \exp(2\alpha(x-\beta)^+)\,\mathrm{d}x = y \wedge \beta + \frac{1}{2\alpha}\big(\mathrm{e}^{2\alpha((y\vee\beta)-\beta)} - 1\big), \qquad y \in \mathbb{R}_+.$$

The inverse of the function F defined in the previous display is given by

$$F^{-1}(s) = \begin{cases} s & \text{if } s \le \beta, \\ \beta + \frac{1}{2\alpha}\log(2\alpha(s-\beta)+1) & \text{if } s > \beta, \end{cases} \qquad s \in \mathbb{R}_+. \tag{3.106}$$

From (3.105), $F \circ \varphi = \int_0^\cdot |\sigma_u|^2\,\mathrm{d}u$. Consequently,

$$\mathbb{V}[X_t] = \varphi_t = F^{-1}\bigg(\int_0^t |\sigma_u|^2\,\mathrm{d}u\bigg), \qquad t \in \mathbb{R}_+. \tag{3.107}$$

The function F^{-1} defined in (3.106) is plotted in Figure 3.4 below for specific values of the variance dampening parameters, $\alpha = 4$ and $\beta = 1.5$. Its form reflects the fact that the process $X = (X_t)_{t\in\mathbb{R}_+}$ governed by (3.102) is indistinguishable from an ordinary arithmetic Brownian motion until $\mathbb{V}[X_t]$ exceeds the threshold level β. From this point on, the presence of the function g in (3.102) causes a significant reduction in variance, and in fact, $\mathbb{V}[X_t] \sim \frac{1}{2\alpha}\log(\int_0^t |\sigma_u|^2\,\mathrm{d}u)$ as $t\to\infty$.

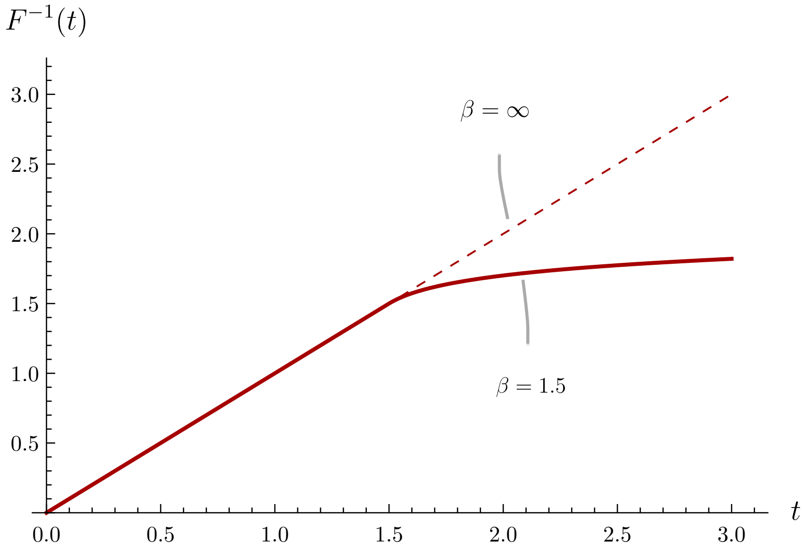

Figure 3.4. The function F^{-1} defined in (3.106), for the dampening strength $\alpha = 4$ and dampening threshold $\beta = 1.5$. The dashed line corresponds to the undampened case, i.e. $\beta = \infty$. By (3.107), $F^{-1}(t)$ is the variance at time t of the variance-dampened arithmetic Brownian motion $X = (X_t)_{t\in\mathbb{R}_+}$ when the volatility coefficient σ in (3.102) satisfies $|\sigma_u|^2 = 1$ for all $u \in \mathbb{R}_+$. This is the case in particular if $n = d$ and σ has the constant value $\frac{1}{\sqrt{n}} I_n$, where I_n denotes the $n \times n$ identity matrix.

Remark 3.25 (Deterministic vs. measure-dependence induced variance dampening)**.**

(a) Fix a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$ (Definition C.5(b)) and consider the McKean SDE (2.37) with coefficients as in (3.55) from Assumption 3.14, with initial condition $X_0 \overset{\text{a.s.}}{=} \xi$. As seen in Example 3.24 and observed in a more general context in Step 1 of the proof of Theorem C.34 (see also (2.70)), once strong existence and pathwise uniqueness of a solution $X = (X_t)_{t\in\mathbb{R}_+}$ of (2.37) is established, the functions $\mathbb{R}_+ \ni t \mapsto g_i(t, \mathcal{L}[X])$, $i \in \{1, 2\}$, are deterministic, measurable and locally bounded functions of time.[50] Therefore, the influence of measure dependence on the solution of the McKean SDE can be accounted for entirely through equivalent deterministic coefficients, obtained by replacing the measure dependent terms $f_i(t, w)g_i(t, \mathbb{Q})$ in (3.55) by $\tilde{f}_i(t, w) := f_i(t, w)g_i(t, \mathcal{L}[X])$ for all $(t, w) \in \mathbb{R}_+ \times W_n$ and $i \in \{1, 2\}$.

In the concrete setting of Example 3.24, this means that the SDE (3.102) with initial condition $X_0 \overset{\text{a.s.}}{=} 0$ is equivalent to the SDE

$$\mathrm{d}X_t = \mu_t\,\mathrm{d}t + \tilde{\sigma}_t\,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+,$$

with initial condition $X_0 \overset{\text{a.s.}}{=} 0$ and

$$\tilde{\sigma}_t := g(\mathcal{L}[X]_t)\sigma_t, \qquad t \in \mathbb{R}_+. \tag{3.108}$$

[50] If there is also weak uniqueness, these functions are in addition independent of the underlying set-up.

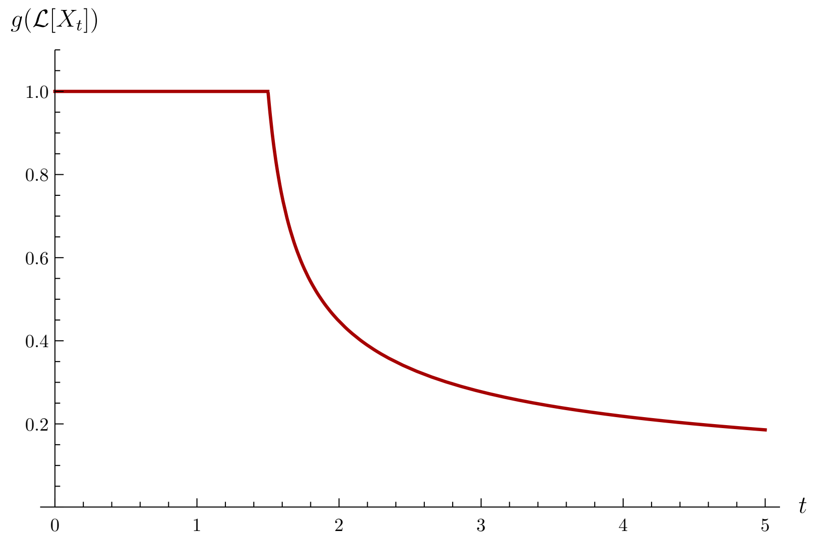

Figure 3.5. The function $t \mapsto g(\mathcal{L}[X_t])$ from (3.109), see also (3.110), where $X = (X_t)_{t\in\mathbb{R}_+}$ is the pathwise unique strong solution of (3.102) from Example 3.24, for the the dampening strength $\alpha = 4$ and dampening threshold $\beta = 1.5$, assuming also that $|\sigma_u|^2 = 1$ for all $u \in \mathbb{R}_+$.

By the definition[51] (3.92) of g and (3.107), we have, for all $t \in \mathbb{R}_+$,

$$g(\mathcal{L}[X]_t) = \exp(-\alpha(\mathbb{V}[X_t] - \beta)^+) = \exp\biggl(-\alpha\biggl(F^{-1}\biggl(\int_0^t |\sigma_u|^2 \,\mathrm{d}u\biggr)\biggr)^+\biggr), \tag{3.109}$$

with F^{-1} defined in (3.106), which simplifies to

$$g(\mathcal{L}[X_t]) = \begin{cases} 1, & \text{if } \int_0^t |\sigma_u|^2 \,\mathrm{d}u \le \beta, \\ \Bigl(1 - 2\alpha\beta + 2\alpha \int_0^t |\sigma_u|^2 \,\mathrm{d}u\Bigr)^{-1/2} & \text{otherwise.} \end{cases} \tag{3.110}$$

This function is plotted in Figure 3.5 below for $\alpha = 4$ and $\beta = 1.5$, assuming in addition that $|\sigma_u|^2 = 1$ for all $u \in \mathbb{R}_+$. Therefore, instead of the SDE (3.102) in Example 3.24, we could have started with the classical SDE

$$\mathrm{d}X_t = \mu_t \,\mathrm{d}t + \tilde{\sigma}_t \,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+,$$

with $\tilde{\sigma}$ defined in (3.108), and obtained exactly the same solutions.

(b) The observation made in part (a) raises the question in which sense variance dampening by means of McKean SDEs is advantageous from a modelling perspective over using classical SDEs with appropriate time-dependent volatility coefficients.

The answer to this question lies in the following observation: The form of variance dampening in Example 3.21 depends on the process variance directly – volatility is suppressed once it exceeds a certain threshold level, β. To achieve the same effect

[51] We use again Convention 3.15 to write g as a function of $\mathcal{L}[X_t]$ instead of time and $\mathcal{L}[X]$.

using classical SDEs, it is necessary to understand how the curve $\mathbb{R}_+ \ni t \mapsto \mathbb{V}[X_t]$ of variances of the solution depends on the volatility coefficient σ, in order to then modify σ appropriately to produce the desired variance profile. In the simple case of the arithmetic Brownian motion considered in Example 3.24, this is straightforward, using the arguments that lead to (3.104), and similar methods work in the case of geometric Brownian motion considered in Example 3.26 below (see (3.113)). However, as soon as the coefficients become more complicated (e.g. if σ is no longer deterministic or depends nonlinearly on the solution, as e.g. in the local stochastic volatility model (Definition 3.1)), this approach becomes prohibitively complicated. In comparison, the variance dampening approach from Example 3.21 using McKean SDEs can be applied without modification to complicated volatilities.

Example 3.26 (Variance-based volatility dampening in geometric Brownian motion with zero drift)**.** Let $\sigma\colon \mathbb{R}_+ \to \mathbb{R}^{n\times d}$ be a continuous function. Fix an $\mathbb{R}^d$-valued standard Brownian motion $B = (B_t)_{t\in\mathbb{R}_+}$ defined on a stochastic base (Definition C.5 (a)) $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. We consider the SDE

$$\mathrm{d}X_t = g(\mathcal{L}[X_t])\,\mathrm{diag}(X_t)\,\sigma_t\,\mathrm{d}B_t\,, \qquad t\in\mathbb{R}_+, \tag{3.111}$$

for an $\mathbb{R}^n$-valued[52] solution with deterministic initial condition $X_0 \in (0,\infty)^n$, where g is as in (3.92) with $\alpha\in\mathbb{R}_+$ and $\beta\in\mathbb{R}_+$ assumed to be constant, and where we follow Convention 3.15 to write g as a function of $\mathcal{L}[X_t]$ instead of $(t, \mathcal{L}[X])$. By the assumed continuity of σ and Lemma 3.22, the coefficient in (3.111) satisfies Assumption 3.14. Therefore, by Theorem 3.19, there is uniqueness, pathwise and in law, as well as strong existence of solutions of (3.111), and every solution is a square integrable process. Let $X = (X_t)_{t\in\mathbb{R}_+}$ be a strong solution. Since $\mathbb{R}_+ \ni t \mapsto \tilde g_t := g(\mathcal{L}[X_t]) \in \mathbb{R}_+$ is continuous (see Footnote 45), the product of the terms $g(\mathcal{L}[X_t])$ and σ_t in (3.111) is a deterministic and continuous function of time only. By a characterization of the stochastic exponential (see [53, Theorem 6.61]) applied to each of the n SDEs defined by (3.111) for the components $X^1, \dots, X^n$ of X, it follows for all $i\in\{1,\dots,n\}$ that,

$$X^i \overset{\text{uti}}{=} X_0^i\mathcal{E}((\tilde g\sigma^i)\cdot B) = X_0^i \exp\biggl(\int_0^\cdot \tilde g_s\sigma_s^i\,\mathrm{d}B_s - \frac{1}{2}\int_0^\cdot \tilde g_s^2|\sigma_s^i|^2\,\mathrm{d}s\biggr), \tag{3.112}$$

where $\sigma_s^i := (\sigma_s^{ij})_{j\in\{1,\dots,d\}}$, $s\in\mathbb{R}_+$, denotes the i-th row of σ. It follows that

$$X_t^i/X_0^i \sim \mathrm{LogN}\biggl(-\frac{1}{2}\int_0^t \tilde g_s^2|\sigma_s^i|^2\,\mathrm{d}s\,,\int_0^t \tilde g_s^2|\sigma_s^i|^2\,\mathrm{d}s\biggr), \qquad t\in\mathbb{R}_+,$$

where LogN denotes the lognormal distribution, see [53, Exercise 7.6] for details. It is well known (see e.g. [53, Exercise 2.57 (c)]) that for all $(\tilde\mu,\tilde\sigma)\in\mathbb{R}^2$, the variance of $\mathrm{LogN}(\tilde\mu,\tilde\sigma^2)$ is $(\exp(\tilde\sigma^2)-1)\exp(2\tilde\mu+\tilde\sigma^2)$, hence

$$\mathbb{V}[X_t^i/X_0^i] = \exp\biggl(\int_0^t \tilde g_s^2|\sigma_s^i|^2\,\mathrm{d}s\biggr) - 1. \tag{3.113}$$

[52] We will specialize after (3.114) below to the case $n=1$ in order to obtain more explicit results.

By the definition (3.93) of the variance of $\mathbb{R}^n$-valued random variables (see also Footnote 42),

$$\varphi_t := \mathbb{V}[X_t] = \sum_{i=1}^n \mathbb{V}[X_t^i] = \sum_{i=1}^n (X_0^i)^2 \left[\exp\left(\int_0^t \tilde{g}_s^2 |\sigma_s^i|^2 \,\mathrm{d}s\right) - 1\right],$$

for each $t \in \mathbb{R}_+$. The function $\tilde{g}$ can be written as $\tilde{g}_s = \hat{g}(\varphi_s)$ with $\hat{g}(v) := \exp(-\alpha(v-\beta)^+)$, $v \in \mathbb{R}_+$, hence φ satisfies

$$\varphi_t = \sum_{i=1}^n (X_0^i)^2 \left[\exp\left(\int_0^t \hat{g}^2(\varphi_s) |\sigma_s^i|^2 \,\mathrm{d}s\right) - 1\right], \qquad t \in \mathbb{R}_+. \tag{3.114}$$

In the special case $n = 1$, which we consider from now on, this integral equation[53] can be reduced to an ordinary differential equation. To simplify the notation, we will assume that $X_0^1 = 1$ and write σ instead of σ^1. Then from the previous display,

$$1 + \varphi_t = \exp\left(\int_0^t \hat{g}^2(\varphi_s) |\sigma_s|^2 \,\mathrm{d}s\right),$$

for each $t \in \mathbb{R}_+$. Taking logarithms, $\log(1+\varphi_t) = \int_0^t \hat{g}^2(\varphi_s)|\sigma_s|^2 \,\mathrm{d}s$, and differentiating w.r.t. t yields

$$\frac{1}{1+\varphi_t} \frac{\mathrm{d}\varphi_t}{\mathrm{d}t} = \hat{g}^2(\varphi_t)|\sigma_t|^2,$$

and thus, using the definition of $\hat{g}$,

$$\frac{\exp(2\alpha(\varphi_t - \beta)^+)}{1+\varphi_t} \frac{\mathrm{d}\varphi_t}{\mathrm{d}t} = |\sigma_t|^2,$$

for all $t \in \mathbb{R}_+$. Integrating w.r.t. t, it follows that

$$F(\varphi_t) = \int_0^t |\sigma_s|^2 \,\mathrm{d}s\,, \qquad t \in \mathbb{R}_+, \tag{3.115}$$

where

$$F(y) := \int_0^y \frac{\exp(2\alpha(x-\beta)^+)}{1+x} \,\mathrm{d}x\,, \qquad y \in \mathbb{R}_+. \tag{3.116}$$

This integral can, for $y > \beta$, not be evaluated in terms of elementary functions. However, it can be expressed using elementary substitutions through the exponential integral Ei, defined for $y \in (0, \infty)$ as $\mathrm{Ei}(y) = \int_{-\infty}^y \frac{\mathrm{e}^t}{t} \,\mathrm{d}t$ in the sense of the Cauchy principal value, as

$$F(y) = \begin{cases} \log(1+y), & \text{if } 0 \le y \le \beta, \\ \log(1+\beta) + \mathrm{e}^{-2\alpha(1+\beta)}(\mathrm{Ei}(2\alpha(1+y)) - \mathrm{Ei}(2\alpha(1+\beta))), & \text{if } y > \beta. \end{cases} \tag{3.117}$$

[53] Without the exponential, the integral equation (3.114) would be a nonlinear Volterra equation of the second kind.

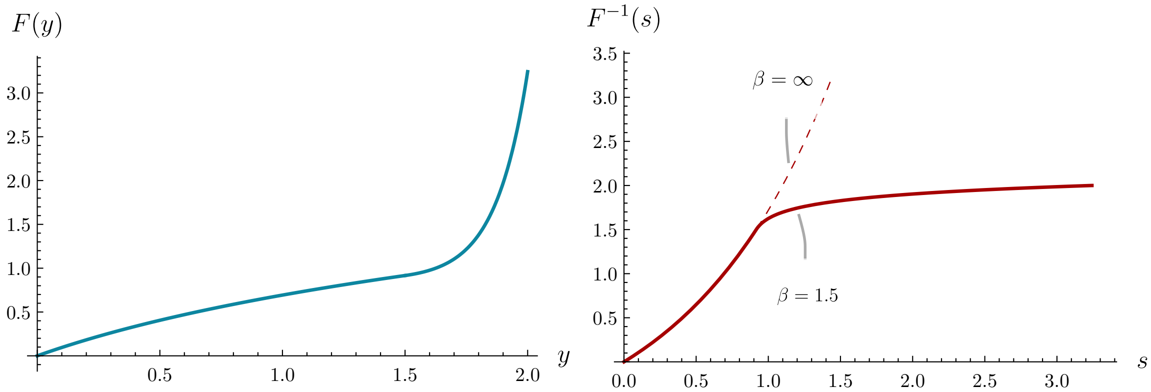

Figure 3.6. Left: The function F defined in (3.117) for the dampening strength $\alpha = 4$ and dampening threshold $\beta = 1.5$. Right: the inverse of F, for the same parameter values (solid line) and for the undampened case, i.e. $\beta = \infty$ (dashed line). By (3.118), $F^{-1}(t)$ is the variance at time t of the variance-dampened one-dimensional geometric Brownian motion $X = (X_t)_{t\in\mathbb{R}_+}$ defined by (3.111) in the case where the volatility coefficient σ in (3.111) is equal to one.

By (3.116), F is continuous and strictly increasing and thus a homeomorphism of $\mathbb{R}_+$. In particular, it is invertible, and (3.115) implies[54] that

$$\mathbb{V}[X_t] = F^{-1}\left(\int_0^t |\sigma_s|^2 \,\mathrm{d}s\right), \qquad t \in \mathbb{R}_+. \tag{3.118}$$

The function F defined in (3.117) and its inverse are plotted in Figure 3.6 below. Figure 3.7 shows on the left the first eight sample paths of approximate solutions $\hat{X}^1, \ldots, \hat{X}^N$ of (3.111) for the dimension $n = 1$ and with $\sigma = 1$, computed[55] using the particle approximation described in Subsection 2.4.2 with $N = 250$ particles. We note that by Remark 3.20, convergence of the particle approximation has not been formally justified in the setting of Assumption 3.14. Despite this, the empirical variances[56] computed from the vector $\hat{X} = \hat{X}^1, \ldots, \hat{X}^N$, which can be seen in the right plot in Figure 3.7, are in line with the theoretical values, see (3.118) and Figure 3.6.

3.2.3. The Mean-Field LIBOR Market Model

To address the problem of explosion described at the beginning of Section 3.2, it is proposed in [17] to allow, for each $i \in \{1, \ldots, N\}$, for a dependence of the volatility σ^i of the i-th

[54] The fact that X has by assumption the deterministic initial value $X_0 = 1$ implies (e.g. by the Burkholder–Davis–Gundy inequality, [53, Theorem 5.84]), together with (3.118) and local boundedness of σ that the quadratic variation $[X]$ of the solution X of (3.111) in the one-dimensional case is an integrable process. This implies (see [53, Lemma 5.89]) that X is a square-integrable true martingale.

[55] We used Wolfram Mathematica's `ItoProcess[]` and `RandomFunction[]` routines with a time step of 0.005 to simulate solutions of the particle approximation SDE (see Definition 2.51) associated with (3.111).

[56] By the empirical variance $\hat{\mathbb{V}}[\hat{x}]$ of a vector $\hat{x} = (\hat{x}^1, \ldots, \hat{x}^N)$ of N numbers, we understand the variance of the associated empirical distribution, multiplied with $\frac{N}{N-1}$ in order to obtain an unbiased estimator of the variance when $\hat{x}$ is replaced by a random vector whose components are i.i.d. square-integrable random variables.

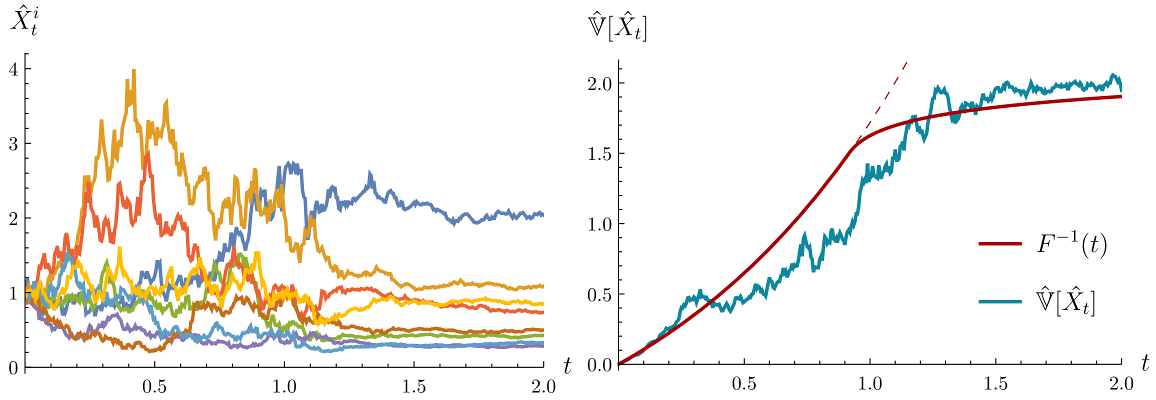

Figure 3.7. Left: The first eight sample paths of approximate solutions $\hat{X}^1, \ldots, \hat{X}^N$ of (3.111) for the dimension $n = 1$, with the volatility σ identically equal to 1, dampening strength $\alpha = 4$ and dampening threshold $\beta = 1.5$, obtained using the particle approximation described in Subsection 2.4.2 with $N = 250$ particles. Right: Comparison between the empirical variance of the vector $\hat{X} = (\hat{X}^1, \ldots, \hat{X}^N)$ of approximate solutions and the theoretical prediction, cf. (3.118) and Figure 3.6. The dashed line corresponds to the theoretical prediction for the case without any dampening.

forward LIBOR L^i (see Assumption 3.13) on both time and the distribution under its respective forward measure $\mathbb{Q}^i$. More precisely, (measurable) volatilities of the form[57]

$$\sigma^i \colon [0, t_{i-1}] \times \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^d, \qquad i \in \{1, \ldots, N\}, \tag{3.119}$$

(cf. (3.48)) are permitted.[58]

The construction of the mean-field LMM then proceeds by backward induction, exactly as in the classical case, but with the deterministic volatilities replaced by their extended forms from (3.119). That is, for each $i \in \{1, \ldots, N\}$, (3.48) is replaced by

$$\mathrm{d}L^i_t = L^i_t \sigma^i(t, \mathcal{L}_{\mathbb{Q}^i}[L^i_t])\, \mathrm{d}W^i_t\,, \qquad t \in [0, t_{i-1}], \qquad \text{w.r.t. } \mathbb{Q}^i, \tag{3.120}$$

and (3.50) by

$$\gamma^i_s := \frac{\delta_i L^i_s}{\delta_i L^i_s + 1} \sigma^i(s, \mathcal{L}_{\mathbb{Q}^i}[L^i_s]), \qquad s \in [0, t_{i-1}], \tag{3.121}$$

and with these modifications, we retain Assumption 3.13.

Example 3.27 (Volatility dampening in the mean-field LMM, [17, §2.1]). Since the objective in the mean-field LMM is to prevent explosion, the example

$$\sigma^i(t, \mu) := \hat{\sigma}^i(t) \exp\bigl(-\max\{\mathbb{V}[\mu] - \beta^i, 0\}\bigr), \qquad (t, \mu) \in [0, t_{i-1}] \times \mathcal{P}_2(\mathbb{R}), \tag{3.122}$$

is of particular importance, where for each $i \in \{1, \ldots, N\}$, $\hat{\sigma}^i \colon [0, t_{i-1}] \to \mathbb{R}^d$ is a bounded measurable function, β^i is a nonnegative constant and $\mathbb{V}[\mu]$ is the variance of the measure

[57] In this section, we consider only the case (2.39) of McKean SDEs with instantaneous dependence, hence the measure argument of σ^i lies in $\mathcal{P}_2(\mathbb{K}^n)$ instead of $\widehat{\mathcal{P}}_2(W_n)$ and we follow Convention 3.15.

[58] Compared to Subsection 3.2.2, we consider in this section SDEs with coefficients defined on finite time intervals. By Remark 2.30, this creates only superficial complications, and we can apply the previously developed results also in this setting.

μ, see also (3.93). Comparing (3.119) and (3.120) for a fixed $i \in \{1, \ldots, N\}$ with (3.55), the dimension n is equal to one, $\mathbb{K} = \mathbb{R}$, the functions $\tilde{b}$, $\tilde{\sigma}$ as well as f_1, g_1 and h_1 are all identically equal to zero, the function $f_2 \colon \mathbb{R} \to \mathbb{R}$ is the identity map, $g_2 \colon \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$ is defined as

$$g_2(\mu) = \exp\bigl(-\max\bigl\{\mathbb{V}[\mu] - \beta^i, 0\bigr\}\bigr), \qquad \mu \in \mathcal{P}_2(\mathbb{R}), \tag{3.123}$$

(cf. Example 3.21) and the function $h_2 \colon [0, t_{i-1}] \to \mathbb{R}^d$ can be identified with the function $\hat{\sigma}^i$. By Lemma 3.22 and the assumed boundedness of $\hat{\sigma}^i$, these coefficients satisfy Assumption 3.14. Thus Theorem 3.19 applies to the SDE in (3.120), for each $i \in \{1, \ldots, N\}$.

To construct the full mean-field LMM (i.e. the forward LIBORs $L^1, \ldots, L^N$ together with their associated forward measures $\mathbb{Q}^1, \ldots, \mathbb{Q}^N$), fix a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{Q})$ (Definition C.5 (a)) carrying an $\mathbb{R}^d$-valued standard Brownian motion W, and where the filtration $\mathbb{F}$ is the augmentation by $(\mathbb{Q}, \mathcal{F}^W_{\overline{T}})$-null sets (see Remark C.4 (a)) of the natural filtration $\mathbb{F}^W = (\mathcal{F}^W_t)_{t \in [0, \overline{T}]}$ generated by W. Assume that the coefficients σ^i from (3.119) are such that there is strong existence and pathwise uniqueness of solutions of (3.120) for each $i \in \{1, \ldots, N\}$. This is the case in particular when the volatilities are as in Example 3.27 above. Let $\mathbb{Q}^N := \mathbb{Q}$ and $W^N := W$. From (3.120) with $i = N$, we obtain the forward LIBOR for the interval $[t_{N-1}, t_N]$,

$$L^N_t \overset{\text{a.s.}}{=} L^N_0 \mathcal{E}\left(\int_0^t \sigma^N(s, \mathcal{L}_{\mathbb{Q}^N}[L^N_s])\, \mathrm{d}W^N_s\right), \qquad t \in [0, t_{N-1}],$$

where we used also that, once existence and uniqueness of solutions of (3.120) is established, (3.120) can be seen as a classical SDE[59] and applied a characterization of the stochastic exponential ([53, Theorem 6.61]). Since $[0, t_{N-1}] \ni s \mapsto \sigma^N(s, \mathcal{L}_{\mathbb{Q}^N}[L^N_s])$ is in the setting of Example 3.27 a bounded deterministic function of time, see Footnote 45, L^N is, as in the classical case, the stochastic exponential of the Gaussian martingale $\int_0^\cdot \sigma^N(s, \mathcal{L}_{\mathbb{Q}^N}[L^N_s])\, \mathrm{d}W^N_s$ (see [53, Theorem 7.1 and Exercise 7.6]) and thus lognormally distributed.[60] In fact, the variance (and thus the distribution) of L^N_t have been obtained analytically in Example 3.26.

Boundedness of σ^N ensures that γ^N defined by (3.121) is bounded, hence $\mathbb{Q}^{N-1}$ from (3.49) is well-defined. Moreover, by Girsanov's theorem ([53, Theorem 7.41]), W^{N-1} defined by (3.51) is an $\mathbb{R}^d$-valued standard $\mathbb{Q}^{N-1}$-Brownian motion. Applying the same arguments inductively, we obtain pathwise unique strong solutions $L^N, \ldots, L^1$ of (3.120).

Remark 3.28 (Dynamics of the forward LIBORs w.r.t. $\mathbb{Q}^N$)**.** The forward LIBORs $L^1, \ldots,$ L^N obtained above are, a priori, solutions to SDEs w.r.t. different measures. For simulation purposes, it is desirable to describe their dynamics by a single SDE w.r.t. the terminal measure $\mathbb{Q}^N$. As shown in [17, Theorem 2.1 and Remark 2.3], to which we refer for details, this is indeed possible.

We refer to [17, §3] for a discussion of several practical aspects of the mean-field LMM, such as the pricing of interest rate caps and calibration. There, it is also reported, based on numerical investigations, that with the concrete form (3.122) of the volatility, a significant

[59] See also the discussion surrounding (2.70).

[60] The initial value $L^N_0 = \frac{1}{\delta_N}\frac{P(0,t_{N-1}) - P(0,t_N)}{P(0,t_N)}$ is assumed to be deterministic, hence it does not interfere with lognormality.

reduction in the probability of explosion is achieved, which aligns with the observations made in Example 3.26.

A. Pseudometric Spaces

This appendix reviews elements of the theory of pseudometric spaces. The main result is Theorem A.16, a variant of Banach's classical fixed point theorem for topologies induced by countable families of pseudometrics, which is used in the proof of Theorem 2.32.

Definition A.1 (Pseudometric space)**.** [9, §1.1] A *pseudometric space* (X, d) is a set X together with a function $d\colon X \times X \to \mathbb{R}_+$, called a *pseudometric*, such that

(a) $d(x, x) = 0$ for all $x \in X$,

(b) $d(x, y) = d(y, x)$ for all $x, y \in X$ (symmetry),

(c) $d(x, z) \le d(x, y) + d(y, z)$ for all $x, y, z \in X$ (triangle inequality).

Example A.2. (a) Let X be a set, (Y, ϱ) a pseudometric space (e.g. a metric or normed space) and $\Phi\colon X \to Y$ a function. Then $d(x, y) = \varrho(\Phi(x), \Phi(y))$ for $x, y \in X$ defines a pseudometric on X, see also Lemma A.6 below.

(b) For each $t \in \mathbb{R}_+$, the seminorm p_t of uniform convergence defined in (2.1) induces a pseudometric ϱ_t on W_n defined by $\varrho_t(f, g) = p_t(f - g)$ for $f, g \in W_n$.

(c) [53, Remark 15.2] Let $\mathcal{F}$ be an algebra of sets on a set Ω, and let $\mu\colon \mathcal{F} \to \overline{\mathbb{R}}_+$ be a finitely additive set function with $\mu(\varnothing) = 0$. Then

$$d_\mu(A, B) = \mu(A \triangle B), \qquad A, B \in \mathcal{F},$$

defines, for every $A \in \mathcal{F}$, a pseudometric on $\mathcal{R}_{A,\mu} := \{B \in \mathcal{F} : \mu(A \triangle B) < \infty\}$. In particular, d_μ is a pseudometric on the σ-algebra $\mathcal{F}$ of every finite measure space $(\Omega, \mathcal{F}, \mu)$.

Definition A.3 (Topology induced by a pseudometric)**.** [9, §2] The topology induced by a pseudometric d on a set X is defined by its basis

$$\mathcal{B} = \{B_\varepsilon(x) : x \in X,\ \varepsilon > 0\},$$

where $B_\varepsilon(x) := \{y \in X : d(x, y) < \varepsilon\}$ denotes the ball of radius $\varepsilon > 0$ centered at $x \in X$.

Remark A.4. (a) Pseudometric spaces can fail to be Hausdorff, and in fact a pseudometric d on a space X is a metric if and only if it generates a Hausdorff topology.

(b) Pseudometric spaces share many of the desirable properties of metric spaces, since much of metric space theory does not rely on the Hausdorff property. For example, the topology induced by a pseudometric is first countable (and thus determined by convergent sequences), with a neighbourhood basis given by the open balls with rational radii. We refer to [9] for a textbook presentation of general topology in which pseudometric spaces are used throughout.

Definition A.5 (Cauchy sequences, completeness in a pseudometric space)**.** [9, §8.1]

(a) A sequence $(x_n)_{n\in\mathbb{N}}$ in a pseudometric space (X,d) is called a *Cauchy sequence* if for every $\varepsilon > 0$, there is an integer N such that $d(x_m, x_n) < \varepsilon$ for all $m, n > N$.

(b) Let (X,d) be a pseudometric space in which every Cauchy sequence converges. Then (X,d), (and often also the pseudometric d) is called *complete.*

Lemma A.6 (Pullback of a pseudometric)**.** *Let X be a set, let (Y,ϱ) be a pseudometric space and $\Phi\colon X \to Y$ a map. Let d denote the pseudometric on X defined by*[1]

$$d(x,\tilde{x}) = \varrho(\Phi(x), \Phi(\tilde{x})), \qquad x, \tilde{x} \in X. \tag{A.1}$$

(a) Let $\mathcal{O}_Y$ be the topology generated by the metric ϱ on Y. Then the topology $\mathcal{O}_X$ induced by the pseudometric d coincides with the initial topology induced on X by the map Φ, i.e.,

$$\mathcal{O}_X = \Phi^{-1}(\mathcal{O}_Y). \tag{A.2}$$

(b) Let (Y,ϱ) be a complete pseudometric space and assume that the image $\Phi(X)$ is closed in Y (this is the case in particular when Φ is surjective). Then (X,d) is a complete pseudometric space.

Proof. (a) By (A.1), Φ is an isometry. Therefore it is continuous, implying that $\Phi^{-1}(\mathcal{O}_Y) \subseteq \mathcal{O}_X$. In addition, every ball $B_r^d(x)$ for $r > 0$ and $x \in X$ w.r.t. d is of the form

$$\begin{aligned} B_r^d(x) &= \{\tilde{x} \in X : d(x,\tilde{x}) < r\} = \{\tilde{x} \in X : \varrho(\Phi(x), \Phi(\tilde{x})) < r\} \\ &= \{\varrho(\Phi(x), \Phi(\,\cdot\,)) < r\} = \Phi^{-1}(B_r^\varrho(\Phi(x))), \end{aligned}$$

where $B_r^\varrho(\Phi(x))$ denotes the ϱ-ball of radius r centered at $\Phi(x)$. Hence the basis $\mathcal{B}$ of $\mathcal{O}_X$ (see Definition A.3) is contained in the initial topology $\Phi^{-1}(\mathcal{O}_Y)$. This proves $\mathcal{O}_X \subseteq \Phi^{-1}(\mathcal{O}_Y)$ and thereby (A.2).

(b) Let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in (X,d). Since Φ is an isometry by (A.1), it follows that $(\Phi(x_n))_{n\in\mathbb{N}}$ is a Cauchy sequence in the complete pseudometric space (Y,ϱ) and thus converges to a limit $y \in Y$. Since $\Phi(X)$ is closed, there exists an $x \in X$ such that $y = \Phi(x)$. It follows that $d(x, x_n) = \varrho(\Phi(x), \Phi(x_n)) \to 0$. Hence $(x_n)_{n\in\mathbb{N}}$ converges to x in (X,d).

□

Definition A.7 (Contraction, fixed point on a pseudometric space)**.** [9, §8.5] Let (X,d) be a pseudometric space.

(a) A map $\Phi\colon X \to Y$ from (X,d) into a pseudometric space (Y,ϱ) is called a *contraction*, if there is a real number $\theta \in [0,1)$, called *contraction constant*, such that

$$\varrho(\Phi(x), \Phi(\tilde{x})) \le \theta\, d(x,\tilde{x}) \qquad x, \tilde{x} \in X.$$

[1] The pseudometric d is a metric if and only if Φ is injective and ϱ is definite on $\Phi(X)$ (that is, $\varrho(y,\tilde{y}) = 0$ implies $y = \tilde{y}$ for all $y, \tilde{y} \in \Phi(X)$).

(b) An element $x \in X$ is called a *pseudo fixed point* of a map $\Phi\colon X \to X$, if $d(\Phi(x), x) = 0$. The set of all pseudo fixed points of Φ is denoted $\mathrm{Fix}(\Phi)$. A point $x \in X$ is called a (true) *fixed point*, if $\Phi(x) = x$.

Let X be a set and $\Phi\colon X \to X$ a map. For every $n \in \mathbb{N}$, let $\Phi^{\circ n}$ denote the n-fold composition of Φ with itself. The sequence $(\Phi^{\circ n}(x_0))_{n\in\mathbb{N}}$ defined by iterated application of Φ to a starting point $x_0 \in X$ is called fixed point iteration. A fixed point iteration does not always converge, but if it does converge to a limit $x \in X$ w.r.t. a pseudometric d on X and Φ is continuous in x w.r.t. d, then the limit is a pseudo fixed point of Φ w.r.t. d.

Lemma A.8. *Let (X, d) be a pseudometric space and $\Phi\colon X \to X$ a map.*

(a) [9, §8.5] If $x_0 \in X$ is such that the sequence $(\Phi^{\circ n}(x_0))_{n\in\mathbb{N}}$ converges to a limit x in X and Φ is continuous in x, then $x \in \mathrm{Fix}(\Phi)$.

(b) If Φ is a contraction and $\mathrm{Fix}(\Phi) \neq \varnothing$, then

$$\mathrm{Fix}(\Phi) = \{y \in X : d(x, y) = 0\},$$

where $x \in \mathrm{Fix}(\Phi)$ is an arbitrary pseudo fixed point.

Proof. (a) For every $n \in \mathbb{N}$, let $x_n := \Phi^{\circ n}(x_0)$. Since $x_n \xrightarrow[n\to\infty]{} x$ by assumption, continuity of Φ in x implies $\Phi(x_n) \xrightarrow[n\to\infty]{} \Phi(x)$. Consequently

$$d(\Phi(x), x) = \lim_{n\to\infty} d(\Phi(x_n), x) = \lim_{n\to\infty} d(x_{n+1}, x) = 0,$$

where we have used in the first equality that pseudometrics, as metrics, are continuous in each argument with respect to their generated topologies. Hence x is indeed a pseudo fixed point of Φ.

(b) Let $x, y \in \mathrm{Fix}(\Phi)$ and let $\theta \in [0, 1)$ a contraction constant for Φ. Then by Definition A.7(b), $d(\Phi(x), x) = d(\Phi(y), y) = 0$. Hence the triangle inequality for d together with the contraction property of Φ implies

$$d(x, y) \le d(x, \Phi(x)) + d(\Phi(x), \Phi(y)) + d(\Phi(y), y) = d(\Phi(x), \Phi(y)) \le \theta\, d(x, y),$$

hence $d(x, y) = 0$. Conversely, assume that $y \in X$ satisfies $d(x, y) = 0$. By the contraction property, $d(\Phi(x), \Phi(y)) \le \theta\, d(x, y) = 0$, and with the triangle inequality, it follows that

$$d(\Phi(y), y) \le d(\Phi(y), \Phi(x)) + d(\Phi(x), x) + d(x, y) = 0. \qquad \square$$

Example A.9 (Continuity is essential in Lemma A.8(a))**.** Consider the metric space $X = [0, 1]$ with the standard Euclidean metric. Define

$$\Phi\colon X \to X, \qquad \Phi(x) = \begin{cases} 1 & \text{if } x = 0, \\ \frac{x}{2} & \text{if } x \in X \setminus \{0\}, \end{cases}$$

and note that this map has no fixed point. Nevertheless, for every $x_0 \in X$, the sequence $(\Phi^{\circ n}(x_0))_{n\in\mathbb{N}}$ obtained from fixed point iteration converges to 0.

If Φ is a contraction on a non-empty complete pseudometric space (X, d), then the fixed point iteration $(\Phi^{\circ n}(x_0))_{n\in\mathbb{N}}$ from Lemma A.8(a) converges for every starting point $x_0 \in X$. In particular, Φ then has a pseudo fixed point. If d is a metric, this implies $\Phi(x) = x$, hence in this case x is a (true) fixed point, which is unique by Lemma A.8(b). This classical result is due to Banach.

Theorem A.10 (Banach's fixed point theorem for pseudometric spaces)**.** *[9, §8.5] Let (X, d) be a non-empty complete pseudometric space, and let $\Phi\colon X \to X$ be a contraction. Then Φ has a pseudo fixed point $x \in X$, and for every $x_0 \in X$, the sequence $(\Phi^{\circ n}(x_0))_{n\in\mathbb{N}}$ converges to x with respect to d. Moreover, the set of pseudo fixed points of Φ is given by* $\mathrm{Fix}(\Phi) = \{y \in X : d(x, y) = 0\}$.

Proof. The last part follows from Lemma A.8(b). The proof of the first part is essentially identical to that of Banach's fixed point theorem for metric spaces, see also [9, Propositions 8.24 and 8.25]. Fix $x_0 \in X$, let $\theta \in [0, 1)$ be a contraction constant for Φ and define the sequence $(x_n)_{n\in\mathbb{N}}$ by $x_n = \Phi^{\circ n}(x_0)$ for each $n \in \mathbb{N}$. We have

$$d(x_{n+1}, x_n) = d(\Phi(x_n), \Phi(x_{n-1})) \le \theta\, d(x_n, x_{n-1})$$

and thus, inductively, $d(x_{n+1}, x_n) \le \theta^n\, d(x_1, x_0)$ for all $n \in \mathbb{N}$. Using the triangle inequality, this implies that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in (X, d), which by completeness converges to a limit $x \in X$. As a contraction, Φ is continuous, hence $x \in \mathrm{Fix}(\Phi)$ by Lemma A.8(a). □

If (X, d) is a non-empty complete metric space and $\Phi\colon X \to X$ a map, then the fixed point iteration converges to a fixed point if Φ is not necessarily a contraction, as long as some power $\Phi^{\circ m}$ (i.e. the m-fold composition of Φ with itself) is a contraction [8]. This extends to the case where (X, d) is a complete pseudometric space, provided Φ has the property (A.3) below, which is the case in particular if Φ is continuous.

Lemma A.11 (Fixed point from contraction property of a power)**.** *Let $\Phi\colon X \to X$ be a map on a non-empty complete pseudometric space (X, d) such that for all $x, \tilde{x} \in X$,*

$$d(x, \tilde{x}) = 0 \quad\Longrightarrow\quad d(\Phi(x), \Phi(\tilde{x})) = 0, \tag{A.3}$$

which is the case in particular if Φ is continuous. Assume that for some $m \in \mathbb{N}$, the m-fold composition $\Phi^{\circ m}$ of Φ with itself is a contraction.

(a) We have $\mathrm{Fix}(\Phi) = \mathrm{Fix}(\Phi^{\circ m})$, see Definition A.7(b). In particular, Φ has a pseudo fixed point $x \in \mathrm{Fix}(\Phi)$, and $\mathrm{Fix}(\Phi) = \{y \in X : d(x, y) = 0\}$.

(b) For every starting point $x_0 \in X$ and every $x \in \mathrm{Fix}(\Phi)$, the sequence $(\Phi^{\circ n}(x_0))_{n\in\mathbb{N}}$ converges to x with respect to d.

Proof. (a) The last part is implied by the equality $\mathrm{Fix}(\Phi) = \mathrm{Fix}(\Phi^{\circ m})$ and Theorem A.10. For the inclusion $\mathrm{Fix}(\Phi) \subseteq \mathrm{Fix}(\Phi^{\circ m})$, assume w.l.o.g. that $\mathrm{Fix}(\Phi) \neq \varnothing$ and let $x \in \mathrm{Fix}(\Phi)$. Then $d(\Phi(x), x) = 0$ by Definition A.7(b). By the assumed property (A.3) of Φ, it follows that $d(\Phi^2(x), \Phi(x)) = 0$. Inductively, we obtain $d(\Phi^{\circ(k+1)}(x), \Phi^{\circ k}(x)) = 0$ for all $k \in \{1, \dots, m-1\}$, and thus $d(\Phi^{\circ m}(x), x) = 0$ by the triangle inequality. Hence

$x \in \mathrm{Fix}(\Phi^{\circ m})$.
For the inclusion $\mathrm{Fix}(\Phi) \supseteq \mathrm{Fix}(\Phi^{\circ m})$, note that $\mathrm{Fix}(\Phi^{\circ m}) \neq \varnothing$ by Theorem A.10 and let $x \in \mathrm{Fix}(\Phi^{\circ m})$. Then from Definition A.7 (b), $d(\Phi^{\circ m}(x), x) = 0$, which by (A.3) implies

$$0 = d(\Phi(\Phi^{\circ m}(x)), \Phi(x)) = d(\Phi^{\circ m}(\Phi(x)), \Phi(x)),$$

hence $\Phi(x) \in \mathrm{Fix}(\Phi^{\circ m})$. Applying Lemma A.8 (b) for the contraction $\Phi^{\circ m}$, it follows that $d(\Phi(x), x) = 0$ and thus $x \in \mathrm{Fix}(\Phi)$.

(b) Let $x \in \mathrm{Fix}(\Phi)$ be a pseudo fixed point and $x_0 \in X$ a starting point. For convergence of $(\Phi^{\circ n}(x_0))_{n\in\mathbb{N}}$ to x, it suffices that each of the sequences $(\Phi^{\circ(\ell m+r)}(x_0))_{\ell\in\mathbb{N}}$ for $r \in \{0, \dots, m-1\}$ converges to x. Since $\Phi^{\circ(\ell m+r)}(x_0) = (\Phi^{\circ m})^{\circ \ell}(\Phi^{\circ r}(x_0))$, the contraction property of $\Phi^{\circ m}$ together with Theorem A.10 implies that each of the sequences $(\Phi^{\circ(\ell m+r)}(x_0))_{\ell\in\mathbb{N}}$ for $r \in \{0, \dots, m-1\}$ converges to some pseudo fixed point $\tilde{x}_r \in \mathrm{Fix}(\Phi^{\circ m})$. From part (a), $x \in \mathrm{Fix}(\Phi)$ implies $x \in \mathrm{Fix}(\Phi^{\circ m})$, hence by Lemma A.8 (b), $d(x, \tilde{x}_r) = 0$. Thus by the triangle inequality, $\Phi^{\circ(\ell m+r)}(x_0) \xrightarrow[\ell\to\infty]{} x$ for all $r \in \{0, \dots, m-1\}$, which proves the claim. □

We conclude this section with a fixed point theorem for topologies induced by countable families of pseudometrics. Results of this type are usually phrased in the language of uniformizable spaces, see e.g. [6, §IX]. For our purposes, pseudometrics suffice.

Definition A.12 (Topology induced by a family of pseudometrics)**.** Let X be a set, I an index set and $\{d_i\}_{i\in I}$ a family of pseudometrics on X. The topology $\mathcal{O}(d_i : i \in I)$ on X induced by $\{d_i\}_{i\in I}$ is defined by its subbasis

$$\mathcal{S} = \{B^i_\varepsilon(x) : \varepsilon > 0,\, i \in I,\, x \in X\}, \tag{A.4}$$

where

$$B^i_\varepsilon(x) \coloneqq \{y \in X : d_i(x, y) < \varepsilon\}, \qquad \varepsilon > 0,\, i \in I,\, x \in X. \tag{A.5}$$

The set of finite intersections of sets in $\mathcal{S}$ forms a basis of this topology, and an equivalent basis is given by[2] sets of the form

$$U_{x,\varepsilon,i_1,\dots,i_n} \coloneqq \bigcap_{k=1}^{n} B^{i_k}_\varepsilon(x), \qquad n \in \mathbb{N},\ x \in X,\ \varepsilon > 0,\ i_1, \dots, i_n \in I. \tag{A.6}$$

Definition A.13 (Separating family of pseudometrics)**.** A family $\{d_i\}_{i\in I}$ of pseudometrics on a set X is called *separating*, if for all distinct $x, y \in X$, there exists an $i \in I$ with $d_i(x, y) > 0$.

Lemma A.14. *Let X be a set and $\{d_i\}_{i\in I}$ a family of pseudometrics on X. Then the induced topology $\mathcal{O}(d_i : i \in I)$ is Hausdorff if and only if $\{d_i\}_{i\in I}$ is separating.*

[2] It is not hard to see that the family $\mathcal{B}$ of sets of the form (A.6) is the basis of a topology $\mathcal{O}_{\mathcal{B}}$. Every set in $\mathcal{B}$ is a finite intersection of elements of $\mathcal{S}$, hence $\mathcal{O}_{\mathcal{B}} \subseteq \mathcal{O}(d_i : i \in I)$. On the other hand, if a finite intersection $O \coloneqq \bigcap_{k=1}^{n} B^{i_k}_{\varepsilon_k}(x_k)$ of elements of $\mathcal{S}$ is non-empty and $y \in O$, then $y \in U_{y,\delta,i_1,\dots,i_n} \subseteq O$ where $\delta \coloneqq \min_{k\in\{1,\dots,n\}}(\varepsilon_k - d_{i_k}(y, x_k))$, which shows that O is open in $\mathcal{O}_{\mathcal{B}}$. Hence $\mathcal{O}(d_i : i \in I) \subseteq \mathcal{O}_{\mathcal{B}}$.

Proof. For the direction "$\Leftarrow$", assume that $\{d_i\}_{i\in I}$ is separating and let $x, y \in X$ be distinct. By assumption, there exists an $i \in I$ with $\varepsilon := d_i(x, y) > 0$. We verify that the balls $B^i_{\varepsilon/2}(x)$ and $B^i_{\varepsilon/2}(y)$, which are open neighbourhoods of x and y respectively w.r.t. $\mathcal{O}(d_i : i \in I)$, have empty intersection. Assuming towards a contradiction that this is not the case, let $z \in B^i_{\varepsilon/2}(x) \cap B^i_{\varepsilon/2}(y)$. By the triangle inequality for d_i,

$$\varepsilon = d_i(x, y) \le d_i(x, z) + d_i(z, y) < \varepsilon,$$

Since $x, y \in X$ were arbitrary distinct points, it follows that $\mathcal{O}(d_i : i \in I)$ is Hausdorff.
For the converse direction, "$\Rightarrow$", assume that $\mathcal{O}(d_i : i \in I)$ is Hausdorff and let $x, y \in X$ be distinct. By assumption, there exist disjoint open neighbourhoods U of x and V of y. By the definition (A.4) of the subbasis $\mathcal{S}$ of $\mathcal{O}(d_i : i \in I)$, see also the discussion surrounding (A.6), there exist an $n \in \mathbb{N}$, $i_1, \dots, i_n \in I$ and $\varepsilon > 0$ such that $\bigcap_{k=1}^n B^{i_k}_\varepsilon(x) \subseteq U$. Since $U \cap V = \varnothing$, $y \notin \bigcap_{k=1}^n B^{i_k}_\varepsilon(x)$, i.e. there exists a $k_0 \in \{1, \dots, n\}$ with $y \notin B^{i_{k_0}}_\varepsilon(x)$. Therefore, $d_{i_{k_0}}(x, y) \ge \varepsilon > 0$. Hence $\{d_i\}_{i\in I}$ is separating. $\square$

Remark A.15. (a) Let X be a set and $\{d_i\}_{i\in I}$ a family of pseudometrics on X. By the form (A.6) of the basic open sets in $\mathcal{O}(d_i : i \in I)$, a net $(x_\alpha)_{\alpha\in A}$ in X converges to $x \in X$ w.r.t. $\mathcal{O}(d_i : i \in I)$ if and only if

$$d_i(x_\alpha, x) \xrightarrow[\alpha]{} 0, \qquad i \in I.$$

(b) Let $(c_k)_{k\in\mathbb{N}}$ be a null sequence in $\mathbb{R}^*_+$ and $(d_k)_{k\in\mathbb{N}}$ a countable family of pseudometrics on a set X. Then $\mathcal{O}(d_k : k \in \mathbb{N})$ is induced by the pseudometric

$$d(x, y) := \max_{k\in\mathbb{N}}(c_k \wedge d_k(x, y)), \qquad x, y \in X. \tag{A.7}$$

Indeed, by part (a), convergence of a net w.r.t. $\mathcal{O}(d_k : k \in \mathbb{N})$ is equivalent to convergence w.r.t. d_k for all $k \in \mathbb{N}$, which in turn is equivalent to convergence w.r.t. d. This suffices for the equality of $\mathcal{O}(d_k : k \in \mathbb{N})$ and the topology induced by d. In addition, by Lemma A.14, d is a metric if and only if $\{d_k\}_{k\in\mathbb{N}}$ is separating.

Theorem A.16 (Fixed point theorem[3] for topologies induced by families of pseudometrics)**.** *Let $\{d_k\}_{k\in\mathbb{N}}$ a separating family of complete pseudometrics on a non-empty set X. Fix a null sequence $(c_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^*_+$ and denote by d the associated metric from (A.7) for the topology induced by $\{d_k\}_{k\in\mathbb{N}}$. Assume that (X, d) is complete and let $\Phi\colon X \to X$ be a map such that*

(a) for each $k \in \mathbb{N}$, Φ is continuous w.r.t. the topology on X induced by d_k, and
(b) for each $k \in \mathbb{N}$, there exists an $m \in \mathbb{N}$ such that $\Phi^{\circ m}$ is a contraction w.r.t. d_k.

Then Φ has a unique (true) fixed point x in X, and for every starting point $x_0 \in X$, the sequence $(\Phi^{\circ n}(x_0))_{n\in\mathbb{N}}$ converges to x with respect to d.

Proof. Fix $x_0 \in X$ and denote by $\mathrm{Fix}_k(\Phi)$ the set of pseudo fixed points of Φ w.r.t. d_k, see Definition A.7(b). From the assumptions (a) and (b), it follows using Lemma A.11(b)

[3] See [57, Theorem 1.1] for an analogous result in the setting of uniformizable spaces.

that $(\Phi^{\circ n}(x_0))_{n\in\mathbb{N}}$ converges w.r.t. d_k to some limit[4] in $\mathrm{Fix}_k(\Phi)$ for every $k \in \mathbb{N}$. Hence $(\Phi^{\circ n}(x_0))_{n\in\mathbb{N}}$ is Cauchy w.r.t. each pseudometric d_k, and thus, by Definition (A.7) of the metric d, Cauchy w.r.t. d. By completeness of (X, d), there exists an $x \in X$ such that

$$\Phi^{\circ n}(x_0) \xrightarrow[n\to\infty]{} x,$$

w.r.t. d. Since continuity of Φ w.r.t. d_k for every $k \in \mathbb{N}$ is equivalent to continuity of Φ w.r.t. d, see Remark A.15, it follows with Lemma A.8(a) that x is a pseudo fixed point of Φ w.r.t. d, i.e., $d(\Phi(x), x) = 0$. By the fact that $\{d_k\}_{k\in\mathbb{N}}$ is separating, d is a metric. Thus $\Phi(x) = x$, i.e., x is a (true) fixed point of Φ. For uniqueness of x, suppose that $\tilde{x} \in X$ is also a fixed point of Φ. Then $\tilde{x} \in \bigcap_{k\in\mathbb{N}} \mathrm{Fix}_k(\Phi)$ and thus $d_k(x, \tilde{x}) = 0$ for every $k \in \mathbb{N}$ by Lemma A.8(b). Consequently $x = \tilde{x}$. □

Example A.17 (The assumptions on the family of pseudometrics $\{d_k\}_{k\in\mathbb{N}}$ in Theorem A.16 are not sufficient for completeness of (X, d), without which fixed point iteration can fail to converge)**.** Consider the set

$$X = \{(x_k)_{k\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}} : x_k \neq 0 \text{ for at most finitely many } k \in \mathbb{N}\ \}.$$

of $\{0, 1\}$-valued sequences with finite support. For each $k \in \mathbb{N}$, define

$$d_k(x, \tilde{x}) := \sum_{l=1}^{k} |x_l - \tilde{x}_l|, \qquad x, \tilde{x} \in X.$$

Then $\{d_k\}_{k\in\mathbb{N}}$ is a separating family[5] of complete pseudometrics, and the associated metric $d := \max_{k\in\mathbb{N}}(2^{-k} \wedge d_k)$ generates the topology of pointwise convergence in X. The metric d is not complete, since the sequence $(x^{(n)})_{n\in\mathbb{N}}$ defined by $x_k^{(n)} := \mathbb{1}_{k\le n}$ for all $k, n \in \mathbb{N}$ is Cauchy but not convergent w.r.t. d (note that $(1)_{k\in\mathbb{N}} \notin X$). For every $x \in X$, let $T(x) := \min\{k \in \mathbb{N} : x_k = 0\}$ denote the first index $k \in \mathbb{N}$ at which $x_k = 0$. Define

$$\Phi\colon X \to X, \quad \Phi(x) = x + \delta_{T(x)},$$

where for each $n \in \mathbb{N}$, $\delta_n := (\mathbb{1}_{k=n})_{k\in\mathbb{N}}$ and addition of sequences in X is defined pointwise. For each $x \in X$ and every $k \in \mathbb{N}$, the first k elements of the sequence $\Phi^{\circ k}(x)$ are equal to one, hence $d_k(\Phi^{\circ k}(x), \Phi^{\circ k}(\tilde{x})) = 0$ for all $x, \tilde{x} \in X$. In particular, $\Phi^{\circ k}$ is a contraction w.r.t. d_k for each $k \in \mathbb{N}$. Moreover, $d_k(x, \tilde{x}) < \frac{1}{2}$ implies $x_l = \tilde{x}_l$ for all $l \le k$ and all $x, \tilde{x} \in X$, implying that $d_k(\Phi(x), \Phi(\tilde{x})) = 0$ in this case. Hence Φ is continuous w.r.t. each pseudometric d_k. Consequently the family $\{d_k\}_{k\in\mathbb{N}}$ and the map $\Phi\colon X \to X$ satisfy all assumptions of Theorem A.16 except for completeness of (X, d). However, Φ has no fixed point, and there is no starting value $x_0 \in X$ for which the fixed point iteration sequence $(\Phi^{\circ n}(x_0))_{n\in\mathbb{N}}$ converges w.r.t. d.

[4] These limits may be distinct, and in fact $\bigcap_{k\in\mathbb{N}} \mathrm{Fix}_k(\Phi)$ can be empty without completeness of (X, d), see Example A.17.

[5] The family $\{d_k\}_{k\in\mathbb{N}}$ has the monotonicity property $d_k \le d_l$ for all $k \le l$ in $\mathbb{N}$, hence Example A.17 also shows that completeness of (X, d) can fail even with this added property.

B. Measure and Probability Theory

This appendix reviews a number of topics from measure and probability theory which are used in this thesis. Appendix B.1 collects some elementary results, e.g. on the measurability of limits and modifications, factorization of measurable maps and the conditional expectation. Appendix B.2 is about the extension of a sequence of consistent probability measures on the spaces W_n^k, $k \in \mathbb{N}$ (see (2.6)) to a probability measure on the space $W_n = C(\mathbb{R}_+, \mathbb{K}^n)$.

B.1. Selected Facts from Measure and Probability Theory

Lemma B.1 (Measurability of pointwise limits, [53, Lemma 15.29])**.** *Let $(\Omega, \mathcal{F})$ be a measurable space and (X, d) a metric space equipped with its Borel σ-algebra $\mathcal{B}_X$. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of X-valued $\mathcal{F}$-$\mathcal{B}_X$-measurable functions on Ω converging pointwise to a function $f\colon \Omega \to X$, i.e.*

$$f_n(\omega) \xrightarrow[n\to\infty]{} f(\omega),$$

for every $\omega \in \Omega$. Then f is also $\mathcal{F}$-$\mathcal{B}_X$-measurable.

Lemma B.2 ([16, §17])**.** *Let $(\Omega, \mathcal{F})$ be a measurable space, (X, d) a separable[1] metric space and $f\colon \Omega \to X$ a function. The following are equivalent:*

(a) The function f is $\mathcal{F}$-$\mathcal{B}_X$-measurable.

(b) For every $\varepsilon > 0$, there exists a measurable function $g\colon \Omega \to X$ taking countably many values such that

$$\sup_{\omega\in\Omega} d(f(\omega), g(\omega)) < \varepsilon.$$

Proof. The implication (b) $\Rightarrow$ (a) follows from Lemma B.1. For the implication (a) $\Rightarrow$ (b), assume that $f\colon \Omega \to X$ is $\mathcal{F}$-$\mathcal{B}_X$-measurable and let $(x_n)_{n\in\mathbb{N}}$ be a dense sequence in X. Fix $\varepsilon > 0$. For each $n \in \mathbb{N}$, let $B_n := B_\varepsilon(x_n)$ be the open ball of radius ε centered at x_n. Then $C_n := B_n \setminus \left(\bigcup_{k<n} B_k\right)$, $n \in \mathbb{N}$, defines (after possibly dropping all empty sets in the sequence $(C_n)_{n\in\mathbb{N}}$ and renumbering) a measurable partition of X, and the function $g\colon \Omega \to X$ defined by $g(\omega) = x_n$ for all $\omega \in f^{-1}(C_n)$ and $n \in \mathbb{N}$ takes countably many values and satisfies $\sup_{\omega\in\Omega} d(f(\omega), g(\omega)) < \varepsilon$.[2] $\square$

[1] For the direction (b) $\Rightarrow$ (a), the proof below shows that the metric space X does not need to be separable. The reverse implication, (a) $\Rightarrow$ (b) also holds if $f(\Omega)$ is separable and in $\mathcal{B}_X$, which can be seen by applying Lemma B.2 as stated to the separable metric space $X' := f(\Omega)$ instead of X, using that $\mathcal{B}_{X'} = X' \cap \mathcal{B}_X$ and that for maps from Ω to X', $\mathcal{F}$-$\mathcal{B}_{X'}$-measurability and $\mathcal{F}$-$\mathcal{B}_X$-measurability (when viewed as maps into X) are equivalent.

[2] If X is a second countable topological space, a pointwise convergent sequence of measurable functions with countable range can be constructed as in the proof of Lemma B.10, where this is done for the space Y.

Definition B.3 (Augmentation of a sub-σ-algebra by null sets [53, Exercise 3.20])**.** Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and $\mathcal{F} \subseteq \mathcal{A}$ a sub-σ-algebra. Then

$$\bar{\mathcal{F}} \coloneqq \{A \in \mathcal{A} : \text{There exists } F \in \mathcal{F} \text{ with } \mu(A \triangle F) = 0\}, \tag{B.1}$$

is called the *augmentation of $\mathcal{F}$ by $(\mathcal{A}, \mu)$-null sets.*

Lemma B.4 ([53, Exercise 3.20])**.** *In the setting of Definition B.3, let $\mathcal{N} \coloneqq \{N \in \mathcal{A} : \mu(N) = 0\}$ denote the class of $(\mathcal{A}, \mu)$-null sets. The augmentation $\bar{\mathcal{F}}$ of the sub-σ-algebra $\mathcal{F}$ by $(\mathcal{A}, \mu)$-null sets satisfies*

$$\bar{\mathcal{F}} = \{(F \cup N) \setminus N' : F \in \mathcal{F} \text{ and } N, N' \in \mathcal{N}\} = \sigma(\mathcal{F} \cup \mathcal{N}). \tag{B.2}$$

In particular, $\bar{\mathcal{F}}$ is the smallest sub-σ-algebra of $\mathcal{A}$ containing both $\mathcal{F}$ and all μ-null sets in $\mathcal{A}$.

Proof. Let $\mathcal{F}' \coloneqq \{(F \cup N) \setminus N' : F \in \mathcal{F} \text{ and } N, N' \in \mathcal{N}\}$. We will establish (B.2) by proving that each inclusion in

$$\bar{\mathcal{F}} \overset{\text{(a)}}{\subseteq} \mathcal{F}' \overset{\text{(b)}}{\subseteq} \sigma(\mathcal{F} \cup \mathcal{N}) \overset{\text{(c)}}{\subseteq} \bar{\mathcal{F}}$$

holds.

(a) Fix $A \in \bar{\mathcal{F}}$ and let $F \in \mathcal{F}$ with $\mu(A \triangle F) = 0$. Then $N \coloneqq A \setminus F$ and $N' \coloneqq F \setminus A$ are in $\mathcal{N}$, and

$$(F \cup N) \setminus N' = (F \cup (A \setminus F)) \cap (F \cap A^{\mathsf{c}})^{\mathsf{c}} = (A \cup F) \cap (F^{\mathsf{c}} \cup A) = A.$$

Hence $A \in \mathcal{F}'$, which shows that $\bar{\mathcal{F}} \subseteq \mathcal{F}'$.

(b) The inclusion $\mathcal{F}' \subseteq \sigma(\mathcal{F} \cup \mathcal{N})$ follows directly from the definition of $\mathcal{F}'$.

(c) By (B.1), $\mathcal{N} \cup \mathcal{F} \subseteq \bar{\mathcal{F}}$. Hence to show that $\sigma(\mathcal{F} \cup \mathcal{N}) \subseteq \bar{\mathcal{F}}$, it suffices to show that $\bar{\mathcal{F}}$ is a σ-algebra. Trivially, $\varnothing \in \bar{\mathcal{F}}$. If $A \in \bar{\mathcal{F}}$, there exists an $F \in \mathcal{F}$ with $\mu(A \triangle F) = 0$. Since $A \triangle F = A \setminus F \cup F \setminus A = F^{\mathsf{c}} \setminus A^{\mathsf{c}} \cup A^{\mathsf{c}} \setminus F^{\mathsf{c}} = A^{\mathsf{c}} \triangle F^{\mathsf{c}}$, we have $\mu(A^{\mathsf{c}} \triangle F^{\mathsf{c}}) = 0$. Since $\mathcal{F}$ is a σ-algebra, $F^{\mathsf{c}} \in \mathcal{F}$, and thus $A^{\mathsf{c}} \in \bar{\mathcal{F}}$. Hence $\bar{\mathcal{F}}$ is complement-stable. Finally, let A_n, $n \in \mathbb{N}$, be a family of sets in $\bar{\mathcal{F}}$ and let $A \coloneqq \bigcup_{n\in\mathbb{N}} A_n$. For each $n \in \mathbb{N}$, there exists a set $F_n \in \mathcal{F}$ such that $\mu(A_n \triangle F_n) = 0$. Then $F \coloneqq \bigcup_{n\in\mathbb{N}} F_n \in \mathcal{F}$, and

$$\mu(A \triangle F) = \mu(A \setminus F) + \mu(F \setminus A) \leq \sum_{n\in\mathbb{N}} \mu(\underbrace{A_n \setminus F}_{\subseteq A_n \setminus F_n}) + \sum_{n\in\mathbb{N}} \mu(\underbrace{F_n \setminus A}_{\subseteq F_n \setminus A_n}) = 0.$$

Hence $A \in \bar{\mathcal{F}}$, implying that $\bar{\mathcal{F}}$ is stable under taking countable unions and thus a σ-algebra. □

Lemma B.5 (Almost sure equality and augmentation)**.** *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $(E, \mathcal{E})$ a measurable space, $\mathcal{F} \subseteq \mathcal{A}$ a sub-σ-algebra and denote by $\bar{\mathcal{F}}$ the augmentation of $\mathcal{F}$ by $(\mathcal{A}, \mu)$-null sets. Let $f \colon \Omega \to E$ be an $\mathcal{F}$-$\mathcal{E}$-measurable map and $g \colon \Omega \to E$ an $\mathcal{A}$-$\mathcal{E}$-measurable map. If $f = g$, μ-a.e., then g is $\bar{\mathcal{F}}$-measurable.*[3]

[3] If $\mathcal{A}$ is complete in the sense that it contains all subsets of μ-null sets, then Lemma B.5 holds without the explicit assumption of $\mathcal{A}$-measurability of g.

Proof. Let $N := \{f \neq g\}$, which is an $(\mathcal{A}, \mu)$-null set. Fix an arbitrary $E \in \mathcal{E}$ and observe that

$$\begin{aligned}\{g \in E\} &= (\{g \in E\} \cap N^{\mathsf{c}}) \cup (\{g \in E\} \cap N)\\ &= (\{f \in E\} \cap N^{\mathsf{c}}) \cup (\{g \in E\} \cap N),\end{aligned}$$

where $\{f \in E\} \cap N^{\mathsf{c}} \in \bar{\mathcal{F}}$ by $\mathcal{F}$-measurability of f and (B.2). By $\mathcal{A}$-measurability of g, $\{g \in E\} \cap N$ is an $(\mathcal{A}, \mu)$-null set and thus also in $\bar{\mathcal{F}}$. Hence $\{g \in E\} \in \bar{\mathcal{F}}$, proving $\bar{\mathcal{F}}$-measurability of g. ☐

Corollary B.6 (Measurability of modifications)**.** *Let $I \neq \varnothing$ be a totally ordered index set, and let $X = (X_t)_{t \in I}$ and $Y = (Y_t)_{t \in I}$ be stochastic processes on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in a measurable space $(E, \mathcal{E})$. Let $\mathbb{F} = (\mathcal{F}_t)_{t \in I}$ be a filtration of $\mathcal{F}$, assume that X is $\mathbb{F}$-adapted and that Y is a modification of X. Let $\overline{\mathbb{F}} := (\bar{\mathcal{F}}_t)_{t \in I}$, where for each $t \in I$, $\bar{\mathcal{F}}_t$ denotes the augmentation of $\mathcal{F}_t$ by $(\mathcal{F}, \mathbb{P})$-null sets.*

(a) The process Y is $\overline{\mathbb{F}}$-adapted.

(b) If Y is measurable w.r.t. $\mathcal{F}_{t^*} := \sigma(\mathcal{F}_t : t \in I)$ [4] and the filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in I}$ is such that every $\mathcal{F}_t$ contains all $(\mathcal{F}_{t^*}, \mathbb{P})$-null sets,[5] then Y is $\mathbb{F}$-adapted.

Proof. (a) For each $t \in I$, Y_t is $\mathcal{F}$-measurable and a.s. equal to the $\mathcal{F}_t$-measurable random variable X_t. By Lemma B.5, it follows that Y_t is $\bar{\mathcal{F}}_t$-measurable.

(b) This follows from part (a) applied to the probability space $(\Omega, \mathcal{F}_{t^*}, \mathbb{P})$, in which case Lemma B.4 and the fact that each $\mathcal{F}_t$ contains all $(\mathcal{F}_{t^*}, \mathbb{P})$-null sets imply that $\bar{\mathcal{F}}_t = \mathcal{F}_t$. ☐

Lemma B.7. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, (S, d) a Polish space and let $\mathcal{F} \subseteq \mathcal{A}$ be a sub-σ-algebra. Let $\bar{\mathcal{F}}$ be the augmentation of $\mathcal{F}$ by $(\mathcal{A}, \mu)$-null sets and assume that $f \colon \Omega \to S$ is $\bar{\mathcal{F}}$-$\mathcal{B}_S$-measurable. Then there exists an $\mathcal{F}$-$\mathcal{B}_S$-measurable $g \colon \Omega \to S$ with $f = g$, μ-a.e.*

Proof.

Step 1. Consider first the case where f is a $\bar{\mathcal{F}}$-$\mathcal{B}_S$-measurable function assuming countably many values. Let $\{x_k : k \in I\}$ with $I \subseteq \mathbb{N}$ be the distinct elements in $f(\Omega)$ and define $A_k := \{f = x_k\}$ for each $k \in I$, which yields a partition of Ω into $\bar{\mathcal{F}}$-measurable sets. By Definition B.3, there exist sets F_k, $k \in I$ in $\mathcal{F}$ with

$$\mu(A_k \triangle F_k) = 0, \qquad k \in I. \tag{B.3}$$

Denote by $\mathcal{N}$ the class of $(\mathcal{A}, \mu)$-null sets. Let $G_k := F_k \setminus \bigcup_{I \ni \ell < k} F_\ell$ for $k \in I$, and observe that $\{G_k : k \in I\}$ is a disjoint family of sets in $\mathcal{F}$. We next verify that

$$\mu(A_k \triangle G_k) = 0, \qquad k \in I. \tag{B.4}$$

[4] This means that Y_t is $\mathcal{F}_{t^*}$-measurable for each $t \in I$.

[5] If $I = \mathbb{R}_+$, this means that $\mathbb{F}$ satisfies the usual completeness condition (Definition C.3).

We have $G_k \setminus A_k \subseteq F_k \setminus A_k \in \mathcal{N}$ by (B.3), and

$$A_k \setminus G_k = A_k \cap \Bigg(F_k \cap \bigcap_{I \ni \ell < k} F_\ell^{\mathsf{c}}\Bigg)^{\mathsf{c}} = (A_k \setminus F_k) \cup \bigcup_{I \ni \ell < k} A_k \cap F_\ell.$$

By (B.3), $A_k \setminus F_k \in \mathcal{N}$. Furthermore, for each $\ell < k$ in I,

$$A_k \cap F_\ell = \overbrace{A_k \cap (F_\ell \cap A_\ell)}^{\subseteq A_k \cap A_\ell = \varnothing} \cup A_k \cap \overbrace{(F_\ell \setminus A_\ell)}^{\in \mathcal{N} \text{ by (B.3)}} \in \mathcal{N}.$$

Hence $A_k \setminus G_k \in \mathcal{N}$, which establishes (B.4). Fix an arbitrary $x_0 \in S$ and define $G_0 := (\bigcup_{k\in I} G_k)^{\mathsf{c}}$ which is an $(\mathcal{F}, \mu)$-null set, since by the partition property of $\{A_k : k \in I\}$,

$$G_0 = \bigcup_{k \in I} A_k \cap G_0 \subseteq \bigcup_{k \in I} \underbrace{A_k \setminus G_k}_{\in \mathcal{N} \text{ by (B.4)}} \in \mathcal{N}.$$

Moreover, $\big\{G_k : k \in I \cup \{0\}\big\}$ is a partition of Ω by sets in $\mathcal{F}$, and we can define $g \colon \Omega \to S$ by

$$g(\omega) = x_k \quad \Longleftrightarrow \quad \omega \in G_k, \qquad k \in I \cup \{0\}.$$

The map g is $\mathcal{F}$-measurable, and it remains to verify that $f = g$, μ-a.e. Indeed, since G_0 is a null set, it suffices to show that $\{f \neq g\} \setminus G_0 \in \mathcal{N}$, which follows from

$$\{f \neq g\} \setminus G_0 = \bigcup_{k \in I} G_k \cap \{f \neq x_k\} = \bigcup_{k \in I} \overbrace{G_k \cap A_k^{\mathsf{c}}}^{\in \mathcal{N} \text{ by (B.4)}} \in \mathcal{N}.$$

Step 2. We now consider the general case of an $\bar{\mathcal{F}}$-$\mathcal{B}_S$-measurable $f \colon \Omega \to S$. By Lemma B.2, there exists a sequence $(f_k)_{k\in\mathbb{N}}$ of $\bar{\mathcal{F}}$-$\mathcal{B}_S$ measurable functions taking countably many values which converges uniformly to f. By Step 1, there exists for each $k \in \mathbb{N}$ an $\mathcal{F}$-$\mathcal{B}_S$-measurable $g_k \colon \Omega \to S$ such that $N_k := \{f_k \neq g_k\}$ is an $(\mathcal{A}, \mu)$-null set. Let C denote the convergence set of $(g_k)_{k\in\mathbb{N}}$ and note that $C \in \mathcal{F}$ by completeness of S and $\mathcal{F}$-measurability of each g_k, see [16, §16]. Since (f_k) converges, we have $C^{\mathsf{c}} \subseteq N := \bigcup_{k\in\mathbb{N}} N_k$, implying that C^{c} is an $(\mathcal{F}, \mu)$-null set. Fix again an arbitrary $x_0 \in S$ and define

$$\tilde{g}_k(\omega) := \begin{cases} g_k(\omega) & \text{if } \omega \in C, \\ x_0 & \text{if } \omega \in C^{\mathsf{c}}, \end{cases} \qquad \omega \in \Omega,\ k \in \mathbb{N},$$

which gives a sequence of $\mathcal{F}$-measurable functions converging pointwise on Ω to a limit g, which is $\mathcal{F}$-$\mathcal{B}_S$-measurable by Lemma B.1. Moreover, for all $\omega \in N^{\mathsf{c}}$,

$$g(\omega) = \lim_{k\to\infty} \tilde{g}_k(\omega) \overset{N^{\mathsf{c}} \subseteq C}{=} \lim_{k\to\infty} g_k(\omega) \overset{\omega \in N^{\mathsf{c}}}{=} \lim_{k\to\infty} f_k(\omega) = f(\omega).$$

Hence $\{f \neq g\}$ is contained in the $(\mathcal{A}, \mu)$-null set N, proving the claim. $\square$

Part (a) of the next result, which is known as the Doob–Dynkin factorization lemma, states that for a map g taking values in a Polish space, measurability of g w.r.t. the σ-algebra generated by some other map f (see Figure B.1) is equivalent to g being a measurable function of f.[6]

Lemma B.8 (Doob–Dynkin factorization lemma, [16, §18] for part (a))**.** *Let Ω be a non-empty set, $(\Omega', \mathcal{A}')$ a measurable space, (S, d) a Polish space and $f\colon \Omega \to \Omega'$ be a map.*

(a) A map $g\colon \Omega \to S$ is $\sigma(f)$-$\mathcal{B}_S$-measurable if and only if there exists a $\mathcal{A}'$-$\mathcal{B}_S$-measurable map $\varphi\colon \Omega' \to S$ with $g = \varphi \circ f$.

$$\begin{tikzcd}
(\Omega, \sigma(f)) \arrow[r, "g"] \arrow[d, "f"'] & (S, \mathcal{B}_S) \\
(\Omega', \mathcal{A}') \arrow[ur, "\varphi"'] &
\end{tikzcd}$$

Figure B.1. The set-up of Lemma B.8

(b) Let $\mathcal{A}$ be a σ-algebra on Ω, μ a measure on $(\Omega, \mathcal{A})$ and assume that $f\colon \Omega \to \Omega'$ is $\mathcal{A}$-$\mathcal{A}'$-measurable. Let $\overline{\sigma(f)}$ denote the augmentation of $\sigma(f)$ by $(\mathcal{A}, \mu)$-null sets (Definition B.3). Then a map $g\colon \Omega \to S$ is $\overline{\sigma(f)}$-$\mathcal{B}_S$-measurable if and only if it is $\mathcal{A}$-$\mathcal{B}_S$-measurable and there exists an $\mathcal{A}'$-$\mathcal{B}_S$-measurable map $\varphi\colon \Omega' \to S$ such that $g = \varphi \circ f$ holds μ-a.e.

Proof. (a) The direction "$\Leftarrow$" follows from measurability being preserved under composition. For the implication "$\Rightarrow$", we begin with the case where g is a measurable function assuming countably many values $\{x_n\}_{n\in I}$, where $I \subseteq \mathbb{N}$ and $x_n \neq x_m$ for all distinct $n, m \in I$. Let

$$A_n := \{g = x_n\}, \qquad n \in I, \tag{B.5}$$

and note that $\{A_n : n \in I\}$ is a partition of Ω into non-empty $\sigma(f)$-measurable sets. For each $n \in I$, the fact that $A_n \in \sigma(f) = f^{-1}(\mathcal{A}')$ implies existence of a $B_n \in \mathcal{A}'$ with $A_n = f^{-1}(B_n)$. Let $C_n := B_n \setminus \bigl(\bigcup_{I \ni k < n} B_k\bigr)$, $n \in I$, which yields a family of disjoint sets in $\mathcal{A}'$ satisfying

$$f^{-1}(C_n) = f^{-1}(B_n) \setminus \bigcup_{I\ni k<n} f^{-1}(B_k) = A_n \setminus \bigcup_{I \ni k<n} A_k = A_n, \qquad n \in I. \tag{B.6}$$

Let $C_0 := \bigl(\bigcup_{n\in I} C_n\bigr)^{\mathsf{c}}$, and note that $\Omega' = C_0 \cup \bigcup_{n \in I} C_n$. Fix $x_0 \in S$ and define

$$\varphi(\omega') := \begin{cases} x_0 & \text{if } \omega' \in C_0, \\ x_n & \text{if } \omega' \in C_n, \end{cases} \qquad \omega' \in \Omega',$$

which is $\mathcal{A}'$-$\mathcal{B}_S$-measurable. By (B.6), $f(A_n) \subseteq C_n$ for each $n \in I$. Hence

$$(\varphi \circ f)(\omega) = x_n = g(\omega), \qquad \omega \in A_n,\ n \in I,$$

[6] Lemma B.8(a), which is usually stated for the case $S = \mathbb{R}$, can be generalized beyond the case of S being a Polish space. See Remark B.9 below and [53, Lemma 15.58].

which by (B.5) shows $g = \varphi \circ f$.
We now consider the general case of a $\sigma(f)$-$\mathcal{B}_S$-measurable g. Since S is Polish, there exists by Lemma B.2 a sequence $(g_n)_{n\in\mathbb{N}}$ of $\sigma(f)$-$\mathcal{B}_S$-measurable simple functions converging pointwise to g. By the above, each g_n is of the form $g_n = \varphi_n \circ f$ for some $\mathcal{A}'$-$\mathcal{B}_S$-measurable φ_n. Let C denote the convergence set of (φ_n) and note that $C \in \mathcal{A}'$ by completeness of S, see [16, §16]. Set

$$\varphi(\omega') := \begin{cases} \lim_{n\to\infty} \varphi_n(\omega') & \text{if } \omega' \in C, \\ x_0 & \text{if } \omega' \in C^{\mathsf{c}}, \end{cases}$$

which is $\mathcal{A}'$-$\mathcal{B}_S$-measurable by[7] Lemma B.1. Since $(g_n)_{n\in\mathbb{N}} = (\varphi_n \circ f)_{n\in\mathbb{N}}$ converges pointwise, $f(\Omega) \subseteq C$ and thus

$$g(\omega) = \lim_{n\to\infty} g_n(\omega) = \lim_{n\to\infty} \varphi_n(f(\omega)) = \varphi(f(\omega)), \qquad \omega \in \Omega,$$

proving the claim.

(b) The direction "⇒" follows from part (a) together with Lemma B.7, using also that $\overline{\sigma(f)}$ is a sub-σ-algebra of $\mathcal{A}$. For the implication "⇐", let $\tilde{g} := \varphi \circ f$, which is $\sigma(f)$-$\mathcal{B}_S$-measurable and satisfies $\tilde{g} = g$, μ-a.e. by assumption. By $\mathcal{A}$-measurability of g and Lemma B.5, $\overline{\sigma(f)}$-measurability of g follows. □

Remark B.9 (Extension of Lemma B.8 to isomorphic measurable spaces, [34, Lemma 1.14])**.** In the setting of Lemma B.8, let $(E, \mathcal{E})$ be a measurable space which is isomorphic to $(S, \mathcal{B}_S)$ in the measure-theoretic sense[8] and let $\Psi\colon S \to E$ be an isomorphism. Then if $\tilde{g}\colon \Omega \to E$ is $\sigma(f)$-$\mathcal{E}$ measurable, application of Lemma B.8(a) to $g := \Psi^{-1} \circ \tilde{g}$ yields an $\mathcal{A}'$-$\mathcal{B}_S$ measurable factorizing map φ with $g = \varphi \circ f$. Hence

$$\tilde{g} = \Psi \circ g = \Psi \circ \varphi \circ f = \tilde{\varphi} \circ f,$$

with $\tilde{\varphi} := \Psi \circ \varphi$ measurable w.r.t. $\mathcal{A}'$-$\mathcal{E}$, see also Figure B.2.

$$\begin{tikzcd}
(\Omega, \sigma(f)) \arrow[r, "g"] \arrow[rr, bend left, "\tilde{g}"] \arrow[d, "f"'] & (S, \mathcal{B}_S) \arrow[r, shift left, "\Psi"] & (E, \mathcal{E}) \arrow[l, shift left, "\Psi^{-1}"] \\
(\Omega', \mathcal{A}') \arrow[ur, "\varphi"'] \arrow[urr, bend right, "\tilde{\varphi}"'] & &
\end{tikzcd}$$

Figure B.2. Extension of Lemma B.8 via isomorphism.

It can be shown (see e.g. [34, Theorem 1.8]) that every Polish space is isomorphic to $(T, \mathcal{B}_T)$ for some $T \in \mathcal{B}_{[0,1]}$, hence Lemma B.8 can also be proved by reducing to the special case $(S, \mathcal{B}_S) = ([0,1], \mathcal{B}_{[0,1]})$, as is done e.g. in the proof of [34, Lemma 1.14].

[7] It suffices to observe that φ is the pointwise limit of the sequence $(\tilde{\varphi}_n)_{n\in\mathbb{N}}$ of metric space-valued measurable functions, where $\tilde{\varphi}_n$ is defined by $\tilde{\varphi}_n{\restriction_C} = \varphi_n{\restriction_C}$ and $\tilde{\varphi}_n{\restriction_{C^{\mathsf{c}}}} = x_0$.

[8] This means that there exists a measurable bijection $\Psi\colon (S, \mathcal{B}_S) \to (E, \mathcal{E})$ with measurable inverse.

Lemma B.10 (Joint measurability from continuity in one and measurability in the other argument [39])**.** *Let $(X, \mathcal{A})$ be measurable space, (Y, τ) a topological space with countable basis, and (Z, d) a metric space. Let further $f\colon X \times Y \to Z$ be a function such that $f(\cdot, y)$ is measurable for every $y \in Y$ and $f(x, \cdot)$ is continuous for every $x \in X$. Then f is $\mathcal{A} \otimes \mathcal{B}_Y$-measurable.*[9]

Proof. [39] Let $\{U_n : n \in \mathbb{N}\}$ be a basis of τ. For each $n \in \mathbb{N}$, let

$$\mathcal{E}_n := \left\{ \bigcap_{k \in I} U_k \cap \bigcap_{k \in \{1,\dots,n\} \setminus I} U_k^{\mathsf{c}} \quad : \quad I \subseteq \{1, \dots, n\} \right\} \setminus \{\varnothing\}$$

denote the partition of Y generated by $\{U_1, \dots, U_n\}$. Let $\{E_{n,1}, \dots, E_{n,k_n}\}$ be an enumeration of $\mathcal{E}_n$ and choose $y_{n,i} \in E_{n,i}$ for each $i \in \{1, \dots, k_n\}$. By the measurability of the $E_{n,i}$ and the measurability of $f(\cdot, y)$ for every $y \in Y$, the function $f_n\colon X \times Y \to Z$, defined by

$$f_n(x, y) := f\left(x, y_{n,i}\right) \quad \text{for} \quad y \in E_{n,i},$$

is measurable. We will now show that $f_n \to f$ pointwise. Let $(x, y) \in X \times Y$ and let V be a neighbourhood of $f(x, y)$. By continuity of $f(x, \cdot)$, there is a neighbourhood U of y with $f(x, U) \subseteq V$. We have $y \in U_{n_0} \subseteq U$ for some $n_0 \in \mathbb{N}$. Hence, for every $n \geq n_0$, there is an $i_n \in \{1, \dots, k_n\}$ with $y \in E_{n,i_n} \subseteq U$, and hence $f_n(x, y) = f\left(x, y_{n,i_n}\right) \in f(x, U) \subseteq V$. Thus we have indeed that $f_n \to f$ pointwise, hence by Lemma B.1, f is measurable as a limit of a sequence of measurable functions taking values in a metric space. □

Lemma B.11 (Conditional expectation of a function of two random variables, one independent, [53, Exercise 17.16])**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ a sub- σ-algebra and $(S_1, \mathcal{S}_1)$ and $(S_2, \mathcal{S}_2)$ measurable spaces. Let $X\colon \Omega \to S_1$ and $Y\colon \Omega \to S_2$ be random variables and assume that Y is independent of $\sigma(\mathcal{G} \cup \sigma(X))$. In addition, let $F\colon S_1 \times S_2 \to \mathbb{K}^d$ be a $\mathcal{S}_1 \otimes \mathcal{S}_2$-measurable function satisfying one of the following two conditions:*

(i) The real and imaginary part of every component of F is bounded below or above,
(ii) $F(X, Y) \in \mathcal{L}^1(\mathbb{P})$.

Then

$$\mathbb{E}[F(X, Y) \mid \mathcal{G}] \overset{\text{a.s.}}{=} \mathbb{E}[H(X) | \mathcal{G}], \tag{B.7}$$

where $H(x) := \mathbb{E}[F(x, Y)]$ for all $x \in S_1$ for which the expectation exists,[10] *and $H(x) := 0$ otherwise. If in addition X is $\mathcal{G}$-measurable, then*

$$\mathbb{E}[F(X, Y) | \mathcal{G}] \overset{\text{a.s.}}{=} \mathbb{E}[F(x, Y)]\big|_{x = X} \tag{B.8}$$

Proof sketch. By the componentwise definition of the conditional expectation of $\mathbb{K}^d$-valued random vectors, it suffices to treat the case $d = 1$. For part (i), the set

$$\mathcal{H} := \{F\colon S_1 \times S_2 \to \mathbb{K} : \ F \text{ bounded and } \mathcal{S}_1 \otimes \mathcal{S}_2\text{-measurable satisfying (B.7)} \}$$

[9] If $Y = \mathbb{R}$ and $f(x, \cdot)$ is left-continuous (or right-continuous) for each $x \in X$, then the same conclusion holds. See e.g. [53, Lemma 3.41].

[10] By this we mean that the expectation of each of the $d \dim_{\mathbb{R}}(\mathbb{K})$ real and, if $\mathbb{K} = \mathbb{C}$, imaginary components of $F(x, Y)$ exists in $\overline{\mathbb{R}}$.

is a vector space containing the constants, containing (by the tower property, taking out what is known and ignoring independent information) the indicators of sets in the π-system $\mathcal{S}_1\times\mathcal{S}_2$ generating $\mathcal{S}_1\otimes\mathcal{S}_2$, and which (using monotone and conditional monotone convergence) is closed under bounded monotone convergence. Using the monotone class theorem ([53, 15.68]), it follows that $\mathcal{H}$ contains all bounded $\mathcal{S}_1\otimes\mathcal{S}_2$-measurable functions. From this, (i) follows using the monotone convergence theorem for the conditional expectation. For (ii), let $\mathbb{P}_X := \mathbb{P}\circ X^{-1}$ and $\mathbb{P}_Y := \mathbb{P}\circ Y^{-1}$ denote the distributions of X and Y, respectively, and let

$$S_1' := \{x\in S_1 : F(x,\,\cdot\,)^+ \in \mathcal{L}^1(\mathbb{P}_Y) \text{ or } F(x,\,\cdot\,)^- \in \mathcal{L}^1(\mathbb{P}_Y)\}$$

be the set of $x\in S_1$ for which $\mathbb{E}[F(x,Y)]$ exists in $\overline{\mathbb{R}}$. Due to $\mathcal{S}_1$-measurability of $x\mapsto \mathbb{E}[F(x,Y)^\pm]$, we have $S_1'\in\mathcal{S}_1$. By definition,

$$H(x) = \mathbb{1}_{S_1'(x)}\,\mathbb{E}[F(x,Y)^+] - \mathbb{1}_{S_1'(x)}\,\mathbb{E}[F(x,Y)^-], \qquad x\in S_1,$$

which is well-defined with values in $\overline{\mathbb{R}}$, and $\mathcal{S}_1$-measurable. By assumption, $F(X,Y)\in\mathcal{L}^1(\mathbb{P})$, which implies using independence of X and Y with Fubini's theorem that the map

$$G\colon S_1\to\overline{\mathbb{R}}_+,\quad G(x) := \mathbb{E}[|F(x,Y)|]$$

is in $\mathcal{L}^1(\mathbb{P}_X)$ and thus finite $\mathbb{P}_X$-a.s. Since $\{x\in S_1 : G(x)<\infty\}\subseteq S_1'$, it follows that

$$\mathbb{P}[X\in S_1'] = 1. \tag{B.9}$$

For all $k\in\mathbb{N}$, define $F_k := F\,\mathbb{1}_{|F|\le n}$ and $H_k := \mathbb{E}[F_k(x,Y)]$, $x\in S_1$. For all $x\in S_1'$, we have by the classical dominated convergence theorem, using $F(x,\,\cdot\,)^\pm$ as the dominating function for $(F_k^\pm)_{n\in\mathbb{N}}$ if $F(x,\,\cdot\,)^\pm\in\mathcal{L}^1(\mathbb{P}_Y)$, and otherwise, if either $F(x,\,\cdot\,)^+$ or $F(x,\,\cdot\,)^-$ fail to be $\mathbb{P}_Y$-integrable, by the monotone convergence theorem,

$$H_k(x) = \mathbb{E}[F_k(x,Y)^+] - \mathbb{E}[F_k(x,Y)^-] \xrightarrow[k\to\infty]{} \mathbb{E}[F(x,Y)^+] - \mathbb{E}[F(x,Y)^-] = H(x).$$

With (B.9), it follows that $H_k(X)\xrightarrow{\text{a.s.}} H(X)$. In addition, $|H_k|\le G$ for all $n\in\mathbb{N}$, and we have seen above that $G\in\mathcal{L}^1(\mathbb{P}_X)$. Using the conditional dominated convergence theorem with $|F(X,Y)|$ as the dominating function in the first equality, the bounded case (i) in the second equality and again the conditional dominated convergence theorem, now with $G(X)$ as the dominating function in the third equality, we obtain

$$\begin{aligned}\mathbb{E}[F(X,Y)|\mathcal{G}] &\overset{\text{a.s.}}{=} \lim_{n\to\infty}\mathbb{E}[F_k(X,Y)|\mathcal{G}]\\ &\overset{\text{a.s.}}{=} \lim_{n\to\infty}\mathbb{E}[H_k(X)|\mathcal{G}] \overset{\text{a.s.}}{=} \mathbb{E}[H(X)|\mathcal{G}].\end{aligned}$$ □

The next lemma is used in Subsection 3.1.4, see Remark 3.11, to understand a problem associated with an SDE obtained by particle approximation in the context of calibration of local stochastic volatility models. It states the intuitively obvious fact that given the empirical probability distribution μ associated with a finite set $\{(x_1,y_1),\dots,(x_N,y_N)\}$ of points in a product space $(X\times Y,\mathcal{A}\otimes\mathcal{B})$ (cf. (B.11)), the conditional law of the second projection π_Y given the information that the first projection π_X takes a value $x\in X$ is the uniform distribution on the set $\{y_i : i \in \{1, \dots, N\}, x_i = x\}$ of points in $\{y_i : i \in \{1, \dots, N\}\}$ which are "compatible" with this information. We state this lemma in the language of regular conditional distributions, see [36, §8.3, in particular Definition 8.24] for details, and admit for generality also conditioning on a function φ of π_X. The notation $\mathcal{L}_\mu[\pi_Y|\varphi\circ\pi_X = z] = \kappa(z, \cdot\,)$ used in equation (B.12) below means (by definition) that $\kappa\colon Z \times \mathcal{B} \to [0, 1]$ is a stochastic kernel such that

$$\mathbb{E}_\mu[\mathbb{1}_B(\pi_Y)|\sigma(\varphi \circ \pi_X)] = \kappa(\varphi \circ \pi_X, B), \qquad B \in \mathcal{B}, \tag{B.10}$$

i.e. for each $B \in \mathcal{B}$, $\kappa(\,\cdot\,, B)$ is a factorizing map of the $\sigma(\varphi \circ \pi_X)$-measurable conditional probability $\mathbb{E}_\mu[\mathbb{1}_{\{\pi_Y \in B\}}\,|\sigma(\varphi(\pi_X))]$ in the sense of Lemma B.8. The fact that such a representation can be obtained in terms of a stochastic kernel is what is referred to as "regularity" of the conditional distribution.

Lemma B.12 (Conditional expectations under empirical measures)**.** *Let $(X, \mathcal{A})$, $(Y, \mathcal{B})$ and $(Z, \mathcal{C})$ be measurable spaces, assume that $\mathcal{C}$ contains all finite subsets of Z and let $x_1, \dots, x_N$ and $y_1, \dots, y_N$ be points in X and Y respectively. Define a probability measure on $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ by*

$$\mu = \frac{1}{N}\sum_{i=1}^{N} \delta_{(x_i,y_i)}, \tag{B.11}$$

and denote by $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ the canonical projections on $X \times Y$. Let $\varphi\colon X \to Z$ be a measurable map. Then π_Y has a regular conditional distribution given $\varphi \circ \pi_X$, which is given by

$$\mathcal{L}_\mu[\pi_Y|\varphi \circ \pi_X = z] = \kappa(z, \cdot\,), \qquad z \in Z, \tag{B.12}$$

where $\kappa\colon Z \times \mathcal{B} \to [0, 1]$ is the stochastic kernel defined by

$$\kappa(z, \cdot\,) := \begin{cases} \dfrac{1}{|I_z|}\displaystyle\sum_{i\in I_z} \delta_{y_i} & \text{if } I_z \neq \varnothing, \\ \delta_{y_0} & \text{if } I_z = \varnothing, \end{cases} \tag{B.13}$$

where $y_0 \in Y$ is arbitrary and $I_z := \{i \in \{1, \dots, N\} : \varphi(\pi_X(x_i, y_i)) = z\}$, for all $z \in Z$.

Proof. To show (B.12), we need to check that κ is a stochastic kernel and then verify (B.10). Clearly, $\kappa(z, \cdot\,)$ is a probability measure on $(Y, \mathcal{B})$ for every $z \in Z$. Let $X_0 := \{x_i : i \in \{1, \dots, N\}\}$ and

$$Z_0 := \varphi(X_0), \tag{B.14}$$

which is a finite set and thus in $\mathcal{C}$ by assumption. Note that $|I_z| > 0$ if and only if $z \in Z_0$. Hence for every $B \in \mathcal{B}$ and $z \in Z$,

$$\kappa(z, B) = \mathbb{1}_{Z_0^c}(z)\delta_{y_0}(B) + \sum_{\tilde{z}\in Z_0} \mathbb{1}_{\{\tilde{z}\}}(z)\frac{1}{|I_{\tilde{z}}|}\sum_{i\in I_{\tilde{z}}} \delta_{y_i}(B),$$

which is a $\mathcal{C}$-measurable function of z. Hence κ is a stochastic kernel. To verify (B.10), we show that for all $B \in \mathcal{B}$ and $C \in \mathcal{C}$,

$$\mathbb{E}_\mu[\mathbb{1}_B(\pi_Y)\,\mathbb{1}_C(\varphi \circ \pi_X)] = \mathbb{E}_\mu[\kappa(\varphi \circ \pi_X, B)\,\mathbb{1}_C(\varphi \circ \pi_X)]. \tag{B.15}$$

Fix $B \in \mathcal{B}$ and $C \in \mathcal{C}$. For the left-hand side of (B.15), we obtain, using (B.14) for the second equality,

$$\begin{aligned}\mathbb{E}_\mu[\mathbb{1}_B(\pi_Y)\,\mathbb{1}_C(\varphi \circ \pi_X)] &= \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_B(y_i)\,\mathbb{1}_C(\varphi(x_i)) \\ &= \frac{1}{N}\sum_{z \in Z_0} \mathbb{1}_C(z) \sum_{i \in I_z} \mathbb{1}_B(y_i).\end{aligned} \tag{B.16}$$

For the right-hand side of (B.15), we have using (B.14) for the second and (B.13) for the third equality

$$\begin{aligned}\mathbb{E}_\mu[\kappa(\varphi \circ \pi_X, B)\,\mathbb{1}_C(\varphi \circ \pi_X)] &= \frac{1}{N}\sum_{i=1}^{N} \kappa(\varphi(x_i), B)\,\mathbb{1}_C(\varphi(x_i)) \\ &= \frac{1}{N}\sum_{z \in Z_0} |I_z|\,\mathbb{1}_C(z)\kappa(z, B) \\ &= \frac{1}{N}\sum_{z \in Z_0} \mathbb{1}_C(z) \sum_{i \in I_z} \mathbb{1}_B(y_i).\end{aligned} \tag{B.17}$$

From comparison of (B.16) and (B.17), (B.10) follows. □

B.2. Extension of Consistent Probability Measures to W_n

As in Section 2.1, $W_n = C(\mathbb{R}_+, \mathbb{K}^n)$ is equipped with the σ-algebra $\mathcal{G}$ induced by the canonical projections (see also (2.4)). The main result of this section is Proposition B.20, which states that a consistent sequence of probability measures on the spaces W_n^k, $k \in \mathbb{N}$ (see (2.6)) has a unique extension to a probability measure on W_n which has the members of the sequence as its marginals. This is an analogue of Kolmogorov's extension theorem, see [53, Theorem 2.94], which in its usual form applies to products of measurable spaces. Since $(W_n, \mathcal{G})$ is not such a product, Kolmogorov's extension theorem does not apply, and a new result (see Proposition B.20) is needed. The proof of this Proposition uses the theory of compact classes, which we now review briefly.

Definition B.13 (Compact class [53, Definition 15.89]). A family $\mathcal{C}$ of subsets of a set Ω is called a compact class, if every sequence $(C_n)_{n\in\mathbb{N}}$ in $\mathcal{C}$ with the finite intersection property[11] has non-empty intersection.

Example B.14 (Every family of compact sets in a Hausdorff space is a compact class [53, Example 15.90]). Let $(X, \mathcal{O})$ be a Hausdorff space and $\mathcal{C}$ a system of compact subsets of X. We verify that $\mathcal{C}$ is a compact class. Let $(C_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{C}$ with

$$\bigcap_{n\in\mathbb{N}} C_n = \varnothing. \tag{B.18}$$

[11] A family $\mathcal{C}'$ of subsets of a set Ω is said to have the *finite intersection property*, if every finite subfamily has non-empty intersection.

The aim is to show that a finite subsequence of $(C_n)_{n\in\mathbb{N}}$ has empty intersection.

Define $C'_n = C_1 \cap C_n$ for each $n \in \mathbb{N}$. Since X is Hausdorff, compactness of the sets C_n implies that they are all closed. Since each C'_n is the intersection of the closed set C_n with C_1, C'_n is closed in the relative topology $\mathcal{O}_1 := \mathcal{O} \cap C_1$ on C_1. Since C_1 is compact w.r.t. $\mathcal{O}$, C_1 is also compact in $\mathcal{O}_1$. Hence, from (B.18), $(C'_n)_{n\in\mathbb{N}}$ is a sequence of closed subsets of the compact topological space $(C_1, \mathcal{O}_1)$ with empty intersection. Consequently, from a characterization of compactness by the finite intersection property, see [53, Remark 13.29], there exist finitely many sets in the sequence $(C'_n)_{n\geq 2}$ with empty intersection. Therefore, there exists an $n_0 \in \mathbb{N}$ with $\varnothing = \bigcap_{n=2}^{n_0} C'_n = \bigcap_{n=1}^{n_0} C_n$.

Lemma B.15 (Surjective preimages of compact classes are compact classes [53, Lemma 15.91])**.** *Let $f\colon \Omega \to S$ be surjective and $\mathcal{C}'$ a compact class of subsets of S. Then $\mathcal{C} := f^{-1}(\mathcal{C}')$ is a compact class of subsets of Ω.*

Proof. Let $(C_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{C}$ with the finite intersection property. For each $n \in \mathbb{N}$, there exists a $C'_n \in \mathcal{C}'$ such that $C_n = f^{-1}(C'_n)$. Since the preimage commutes with set operations, $f^{-1}(\bigcap_{n=1}^{n_0} C'_n) = \bigcap_{n=1}^{n_0} C_n \neq \varnothing$ for each $n_0 \in \mathbb{N}$. Hence $(C'_n)_{n\in\mathbb{N}}$ has the finite intersection property. Since $\mathcal{C}'$ is a compact class by assumption, $C' := \bigcap_{n\in\mathbb{N}} C'_n \neq \varnothing$. Since f is surjective, it follows that

$$\bigcap_{n\in\mathbb{N}} C_n = \bigcap_{n\in\mathbb{N}} f^{-1}(C'_n) = f^{-1}\Big(\bigcap_{n\in\mathbb{N}} C'_n\Big) = f^{-1}(C') \neq \varnothing. \qquad \square$$

Definition B.16 (Projective limit of sets [53, Definition 15.92])**.** Let $(T, \leq)$ be a non-empty partially ordered set. For each $t \in T$, let S_t be a set. For all $s \leq t$ in T, let $\pi_{t,s}\colon S_t \to S_s$ be a surjective map being the identity if $s = t$, and assume that the consistency condition

$$\pi_{u,s} = \pi_{t,s} \circ \pi_{u,t} \tag{B.19}$$

is satisfied for all $s \leq t \leq u$ in T. The inverse limit of the sets $(S_t)_{t\in T}$ w.r.t. the maps $(\pi_{t,s})_{s\leq t}$ is the subset of the Cartesian product $\bigtimes_{t\in T} S_t$ given by

$$\Omega := \Big\{\omega = (\omega_t)_{t\in T} \in \bigtimes_{t\in T} S_t : \pi_{t,s}(\omega_t) = \omega_s \text{ for all } s \leq t \text{ in } T\Big\}. \tag{B.20}$$

For each $t \in T$, the map

$$\tilde{\pi}_t\colon \Omega \ni \omega = (\omega_s)_{s\in T} \mapsto \omega_t \in S_t$$

is called the canonical projection from the projective limit onto the t-th factor.

Recall the projection maps

$$\pi_{[0,t]}\colon W_n \to W_n^t \qquad \text{and} \qquad \pi_{[0,s]}^{[0,t]}\colon W_n^t \to W_n^s, \qquad s \leq t \text{ in } \mathbb{R}_+,$$

defined in (2.8) and (2.9), respectively.

Example B.17 (W_n as a projective limit and a compact class therein [53, Example 15.95]).

(a) Let $T := \mathbb{N}$ with the usual order and consider for each $k \in \mathbb{N}$ the set W_n^k of continuous $\mathbb{K}^n$-valued functions on $[0,k]$. For all $k \leq l$, the map $\pi_{[0,k]}^{[0,l]}$ is surjective and equal to the identity if $k = l$. Moreover, the family $(\pi_{[0,k]}^{[0,l]})_{k \leq l}$ satisfies the consistency condition (B.19), hence the projective limit Ω of the sets $(W_n^k)_{k\in\mathbb{N}}$ w.r.t. the maps $(\pi_{[0,k]}^{[0,l]})_{k\leq l}$ is well-defined. Each $\omega \in \Omega$ is of the form $\omega = (f_l)_{l\in\mathbb{N}}$ with $f_l \in W_n^l$ for each $l \in \mathbb{N}$ and, due to the consistency condition in the definition (B.20) of the projective limit, is such that for each $k \leq l$ in $\mathbb{N}$, the restriction $\pi_{[0,k]}^{[0,l]}(f_l)$ of f_l to $[0,k]$ is equal to f_k. For this reason, the map

$$\Phi\colon W_n \ni f \mapsto (\pi_{[0,l]}(f))_{l\in\mathbb{N}} \in \Omega \tag{B.21}$$

is a bijection. Hence the projective limit Ω can be identified with W_n.

(b) For each $k \in \mathbb{N}$, equip W_n^k with the topology induced by the supremum norm p_k', see (2.7), and let $\mathcal{C}_k$ denote the set of compact subsets of W_n^k. By Example B.14, $\mathcal{C}_k$ is a compact class in W_n^k. Moreover, using that for all $k \leq l$ in $\mathbb{N}$, the map $\pi_{[0,k]}^{[0,l]}$ is continuous, it follows that $\pi_{[0,k]}^{[0,l]}(\mathcal{C}_l) \subseteq \mathcal{C}_k$. Therefore, [53, Lemma 15.94] implies that $\mathcal{C}' := \bigcup_{k\in\mathbb{N}} \tilde{\pi}_k^{-1}(\mathcal{C}_k)$ is a compact class of subsets of Ω. Since the map Φ from (B.21) is bijective, it follows using Lemma B.15 that $\mathcal{C} := \Phi^{-1}(\mathcal{C}')$ is a compact class in W_n. In addition, for each $k \in \mathbb{N}$,

$$(\tilde{\pi}_k \circ \Phi)(f) = \tilde{\pi}_k((\pi_{[0,l]}(f))_{l\in\mathbb{N}}) = \pi_{[0,k]}(f), \qquad f \in W_n, \tag{B.22}$$

hence $\tilde{\pi}_k \circ \Phi = \pi_{[0,k]}$. Therefore

$$\mathcal{C} = \bigcup_{k\in\mathbb{N}} \pi_{[0,k]}^{-1}(\mathcal{C}_k). \tag{B.23}$$

Definition B.18 (Projective contents on semirings, Kolmogorov extension [53, Definition 15.100]). Let $(T, \leq)$ with $T \neq \varnothing$ be a partially ordered and upward directed set (see [53, Definition 15.14]) and let $(\mathcal{S}_t)_{t\in T}$ be a filtration of semirings, i.e. $\mathcal{S}_t$ is a semiring for all $t \in T$, and $\mathcal{S}_s \subseteq \mathcal{S}_t$ for all $s \leq t$ in T.

(a) For each $t \in T$, let μ_t denote a content on $(\Omega, \mathcal{S}_t)$. Then $((\mu_t, \mathcal{S}_t))_{t\in T}$ is called *projective* or *Kolmogorov consistent*, if for all $s \leq t$ in T the restriction $\mu_t\restriction_{\mathcal{S}_s}$ is equal to μ_s.

(b) Define $\mathcal{S} = \bigcup_{t\in T} \mathcal{S}_t$ and let $\mathcal{F} := \sigma(\mathcal{S})$ denote the generated σ-algebra on Ω. A measure μ on $(\Omega, \mathcal{F})$ is called a *Kolmogorov extension* of $((\mu_t, \mathcal{S}_t))_{t\in T}$, if $\mu\restriction_{\mathcal{S}_t} = \mu_t$ for each $t \in T$.

Theorem B.19 (Existence of a uniquely determined Kolmogorov extension [53, Theorem 15.102]). *Let $((\mu_t, \mathcal{S}_t))_{t\in T}$ be projective according to Definition B.18. Suppose there is a $t_0 \in T$ such that $\Omega \in \mathcal{S}_{t_0}$ and $\mu_{t_0}(\Omega) < \infty$. If there exists a compact class $\mathcal{C}$ of subsets of Ω such that*

$$\mu_t(A) = \sup\{\mu_t(C) : C \subseteq A \text{ and } C \in \mathcal{C} \cap \mathcal{S}_t\}, \qquad t \in T,\ A \in \mathcal{S}_t \tag{B.24}$$

with the convention $\sup\varnothing := 0$*, then there exists a unique Kolmogorov extension* μ *to the* σ*-algebra* $\sigma(\bigcup_{t\in T}\mathcal{S}_t)$.

The next result, which is useful to obtain completeness of $\widehat{\mathcal{P}}_p(W_n)$ (Proposition 2.7), states that a family $Q^{(k)}\in\mathcal{P}(W_n^k)$, $k\in\mathbb{N}$, of probability measures possessing the consistency property (B.25) has a unique extension to W_n.[12] The result itself is known, see [47, Theorem 4.2], where the proof is based on the machinery of standard Borel spaces. Below, we give a proof based on the theory of compact classes.

Proposition B.20 (Existence of a unique extension in $\mathcal{P}(W_n)$ for a consistent sequence of probability measures)**.** *For each* $k\in\mathbb{N}$*, let* $Q^{(k)}$ *be in* $\mathcal{P}(W_n^k)$*, and assume that the family* $\{Q^{(k)}:k\in\mathbb{N}\}$ *is consistent in the sense that*[13]

$$Q^{(l)}\circ\big(\pi^{[0,l]}_{[0,k]}\big)^{-1}=Q^{(k)}\qquad\text{for all } k\le l \text{ in } \mathbb{N}. \tag{B.25}$$

Then there exists a unique probability measure $\mathbb{Q}\in\mathcal{P}(W_n)$ *which has the measures* $\{Q^{(k)}:k\in\mathbb{N}\}$ *as its marginal distributions, meaning that* $\mathbb{Q}\circ\pi^{-1}_{[0,k]}=Q^{(k)}$ *for all* $k\in\mathbb{N}$.

Proof. As in Section 2.1, W_n is equipped with the σ-algebra $\mathcal{G}$ and the filtration $(\mathcal{G}_t)_{t\in\mathbb{R}_+}$ generated by the canonical projections, see (2.4) and (2.5) respectively. The proof, which is an application of Theorem B.19, consists of two main steps. In Step 1, we construct for each $k\in\mathbb{N}$ a probability measure $\mathbb{Q}^{(k)}$ on $(W_n,\mathcal{G}_k)$ corresponding to $Q^{(k)}$, and then verify that $((\mathbb{Q}^{(k)},\mathcal{G}_k))_{k\in\mathbb{N}}$ is Kolmogorov consistent (Definition B.18(a)). In Step 2, we verify the analogue of (B.24) for the family $((\mathbb{Q}^{(k)},\mathcal{G}_k))_{k\in\mathbb{N}}$ and conclude.

Step 1. By Lemma 2.1(c),

$$\mathcal{G}_k=\pi^{-1}_{[0,k]}(\mathcal{B}_{W_n^k}),\qquad k\in\mathbb{N}, \tag{B.26}$$

where $\mathcal{B}_{W_n^k}$ denotes the Borel σ-algebra on W_n^k associated with the topology induced by the supremum norm p'_k, see (2.7). For each $k\in\mathbb{N}$, $\pi_{[0,k]}$ is surjective,[14] hence for all $A\in\mathcal{G}_k$, $B:=\pi_{[0,k]}(A)$ is the unique set in $\mathcal{B}_{W_n^k}$ such that $A=\pi^{-1}_{[0,k]}(B)$. Let

$$\mathbb{Q}^{(k)}(A):=Q^{(k)}(\pi_{[0,k]}(A)),\qquad A\in\mathcal{G}_k,\ k\in\mathbb{N}, \tag{B.27}$$

and note that, again by surjectivity of $\pi_{[0,k]}$,

$$Q^{(k)}(B)=\mathbb{Q}^{(k)}(\pi^{-1}_{[0,k]}(B)),\qquad B\in\mathcal{B}_{W_n^k},\ k\in\mathbb{N}. \tag{B.28}$$

To verify that each $\mathbb{Q}^{(k)}$ is a probability measure on $\mathcal{G}_k$, fix $k\in\mathbb{N}$ for a moment. Let $(A_l)_{l\in\mathbb{N}}$ be a sequence of pairwise disjoint sets in $\mathcal{G}_k$ and $A:=\bigcup_{l\in\mathbb{N}}A_l$, which is also in $\mathcal{G}_k$. Furthermore, for each $l\in\mathbb{N}$, let $B_l:=\pi_{[0,k]}(A_l)$, which is in $\mathcal{B}_{W_n^k}$. Then $\pi_{[0,k]}(A)=\bigcup_{l\in\mathbb{N}}\pi_{[0,k]}(A_l)=$

[12] This does not follow from Kolmogorov's extension theorem, since W_n is not a product space.

[13] The projection maps $\pi^{[0,t]}_{[0,s]}\colon W_n^t\to W_n^s$, $s\le t$ in $\mathbb{R}_+$, have been defined in (2.9).

[14] Given an arbitrary element $w=(w_s)_{s\in[0,k]}$ of W_n^k, a suitable preimage $\tilde w=(\tilde w_s)_{s\in\mathbb{R}_+}$ in W_n with $\pi_{[0,k]}(\tilde w)=w$ can be constructed using the map ι from (2.18) with $t=k$.

$\bigcup_{l\in\mathbb{N}} B_l$. In addition, the sets $(B_l)_{l\in\mathbb{N}}$ are pairwise disjoint, since $A_l = \pi_{[0,k]}^{-1}(B_l)$ for all $l \in \mathbb{N}$ and $\pi_{[0,k]}$ is surjective. Consequently

$$\mathbb{Q}^{(k)}(A) = Q^{(k)}\Big(\bigcup_{l\in\mathbb{N}} B_l\Big) = \sum_{l\in\mathbb{N}} Q^{(k)}(B_l) = \sum_{l\in\mathbb{N}} \mathbb{Q}^{(k)}(A_l).$$

We next verify that $((\mathbb{Q}^{(k)}, \mathcal{G}_k))_{k\in\mathbb{N}}$ is Kolmogorov consistent (Definition B.18 (a)). Fix $k \le l$ in $\mathbb{N}$ as well as $A \in \mathcal{G}_k$ and let $B \in \mathcal{B}_{W_n^k}$ be such that $A = \pi_{[0,k]}^{-1}(B)$. Using that $\pi_{[0,k]} = \pi_{[0,k]}^{[0,l]} \circ \pi_{[0,l]}$, it follows that

$$A = \pi_{[0,k]}^{-1}(B) = \pi_{[0,l]}^{-1} \circ \big(\pi_{[0,k]}^{[0,l]}\big)^{-1}(B).$$

By (B.27), the previous display and surjectivity of $\pi_{[0,l]}$,

$$\mathbb{Q}^{(l)}(A) = Q^{(l)}(\pi_{[0,l]}(A)) = Q^{(l)}\big(\big(\pi_{[0,k]}^{[0,l]}\big)^{-1}(B)\big). \tag{B.29}$$

From the consistency property (B.25) and (B.28),

$$Q^{(l)}\big(\big(\pi_{[0,k]}^{[0,l]}\big)^{-1}(B)\big) = Q^{(k)}(B) = \mathbb{Q}^{(k)}(A). \tag{B.30}$$

Combining (B.29) and (B.30), it follows that

$$\mathbb{Q}^{(l)}(A) = \mathbb{Q}^{(k)}(A) \quad \text{for all} \quad A \in \mathcal{G}_k.$$

Therefore $((\mathbb{Q}^{(k)}, \mathcal{G}_k))_{k\in\mathbb{N}}$ is Kolmogorov consistent.

Step 2. Let $\mathcal{C}$ denote the compact class in W_n defined in Example B.17 (b). We verify that the analogue of (B.24) is satisfied for the family $((\mathbb{Q}^{(k)}, \mathcal{G}_k))_{k\in\mathbb{N}}$, i.e. that

$$\mathbb{Q}^{(k)}(A) = \sup\{\mathbb{Q}^{(k)}(C) : C \subseteq A \text{ and } C \in \mathcal{C} \cap \mathcal{G}_k\}, \qquad k \in \mathbb{N},\ A \in \mathcal{G}_k. \tag{B.31}$$

Fix $k \in \mathbb{N}$ for now and recall that W_n^k with the topology induced by the supremum norm p_k' from (2.7) is Polish by Lemma 2.1 (a). From [36, Theorem 13.6], it follows that the probability measure $Q^{(k)}$, which is defined on $\mathcal{B}_{W_n^k}$, is inner regular, i.e.

$$Q^{(k)}(B) = \sup\{Q^{(k)}(K) : K \subseteq B \text{ and } K \in \mathcal{C}_k\}, \qquad B \in \mathcal{B}_{W_n^k}, \tag{B.32}$$

where $\mathcal{C}_k$ denotes the family of compact subsets of W_n^k. Fix $A \in \mathcal{G}_k$, and let $B \in \mathcal{B}_{W_n^k}$ be the unique set with $A = \pi_{[0,k]}^{-1}(B)$ (and, therefore, $B = \pi_{[0,k]}(A)$). Using (B.32) and (B.28),

$$\begin{aligned}\mathbb{Q}^{(k)}(A) = Q^{(k)}(B) &= \sup\{Q^{(k)}(K) : K \subseteq B \text{ and } K \in \mathcal{C}_k\} \\ &= \sup\{\mathbb{Q}^{(k)}(\pi_{[0,k]}^{-1}(K)) : K \subseteq B \text{ and } K \in \mathcal{C}_k\}\end{aligned} \tag{B.33}$$

By (B.23) and (B.26), $\pi_{[0,k]}^{-1}(\mathcal{C}_k) \subseteq \mathcal{C} \cap \mathcal{G}_k$. In combination with (B.33), it follows that

$$\mathbb{Q}^{(k)}(A) \le \sup\{\mathbb{Q}^{(k)}(C) : C \subseteq A \text{ and } C \in \mathcal{C} \cap \mathcal{G}_k\}.$$

The reverse inequality in the previous display is trivial, hence

$$\mathbb{Q}^{(k)}(A) = \sup\{\mathbb{Q}^{(k)}(C) : C \subseteq A \text{ and } C \in \mathcal{C} \cap \mathcal{G}_k\}.$$

Since $k \in \mathbb{N}$ was arbitrary, (B.31) follows. By Step 1, $((\mathbb{Q}^{(k)}, \mathcal{G}_k))_{k\in\mathbb{N}}$ is Kolmogorov consistent (Definition B.18(a)), hence Theorem B.19 yields the existence of a unique Kolmogorov extension $\mathbb{Q}$ to the σ-algebra $\mathcal{G} = \sigma(\bigcup_{l\in\mathbb{N}} \mathcal{G}_l)$. In particular, $\mathbb{Q}{\restriction_{\mathcal{G}_k}} = \mathbb{Q}^{(k)}$ for all $k \in \mathbb{N}$, which using (B.26) and (B.28) implies

$$\mathbb{Q} \circ \pi_{[0,k]}^{-1} = \mathbb{Q}^{(k)} \circ \pi_{[0,k]}^{-1} = Q^{(k)}, \qquad k \in \mathbb{N},$$

thus proving the claim. □

The following, slightly less natural formulation of Proposition B.20 is useful for proving completeness of the space $\widehat{\mathcal{P}}_p(W_n)$ (Proposition 2.7).

Corollary B.21. *For each $k \in \mathbb{N}$, let $\mathbb{Q}^{(k)}$ be a probability measure on $(W_n, \mathcal{G}_k)$, and assume that the family $(\mathbb{Q}^{(k)})_{k\in\mathbb{N}}$ is consistent in the sense that*

$$\mathbb{Q}^{(l)}_{[0,k]} = \mathbb{Q}^{(k)}_{[0,k]} \qquad k \le l \text{ in } \mathbb{N}, \tag{B.34}$$

see also the notations (2.17) and (2.15). Then there exists a unique probability measure $\mathbb{Q} \in \mathcal{P}(W_n)$ with $\mathbb{Q}_{[0,k]} = \mathbb{Q}^{(k)}_{[0,k]}$ for all $k \in \mathbb{N}$.

Proof. For each $k \in \mathbb{N}$, let

$$Q^{(k)}(B) := \mathbb{Q}^{(k)}(\pi_{[0,k]}^{-1}(B)), \qquad B \in \mathcal{B}_{W_n^k}, \tag{B.35}$$

which by Lemma 2.1(c) is a well-defined probability measure on $(W_n^k, \mathcal{B}_{W_n^k})$. For all $k \le l$ in $\mathbb{N}$,

$$Q^{(l)} \circ \bigl(\pi^{[0,l]}_{[0,k]}\bigr)^{-1} = \mathbb{Q}^{(l)} \circ \underbrace{\pi_{[0,l]}^{-1} \circ \bigl(\pi^{[0,l]}_{[0,k]}\bigr)^{-1}}_{=\pi_{[0,k]}^{-1}} \overset{(B.35)}{=} \mathbb{Q}^{(l)} \circ \pi_{[0,k]}^{-1} = \mathbb{Q}^{(k)} \circ \pi_{[0,k]}^{-1} = Q^{(k)}.$$

Therefore, $\{Q^{(k)} : k \in \mathbb{N}\}$ satisfies the consistency condition (B.25) of Proposition B.20, by which there exists a unique $\mathbb{Q} \in \mathcal{P}(W_n)$ with $\mathbb{Q} \circ \pi_{[0,k]}^{-1} = Q^{(k)}$ for all $k \in \mathbb{N}$. By (B.35), the claim follows. □

C. Classical Stochastic Differential Equations

This appendix, which builds primarily on [53, §8], provides a detailed overview of the basic theory of classical stochastic differential equations (SDEs) driven by multidimensional Brownian motion. Following [53], we consider SDEs with coefficients depending on a time argument $t \in \mathbb{R}_+$ and the path of the solution process up to time t (see (C.6)). Furthermore, we study SDEs where the coefficients additionally depend on the randomness of the underlying filtered probability space (see (C.7)). Appendix C.1 reviews the terminology and definitions surrounding classical SDEs. Appendix C.2 establishes the classical result of strong existence and pathwise uniqueness of solutions of (C.7) under Lipschitz continuity and boundedness conditions (Theorem C.34). Appendix C.3 considers SDEs driven by multidimensional Brownian motion with correlated components, which arise frequently in financial applications. Finally, in Appendix C.4, we study flows of classical SDEs, obtaining in Corollary C.57 under the conditions of Theorem C.34 also uniqueness in law (Definition C.21) of solutions of the classical SDE (C.6) with nonrandom dependence.

C.1. Definitions

For later reference, we recall the definitions of the spaces of processes which can be integrated in the stochastic sense against a given finite variation process, continuous local martingale or continuous semimartingale.

Definition C.1 (Stochastic integrability [53, Definition 5.96 and Definition 5.110])**.**

(a) For a $\mathbb{K}$-valued adapted continuous process A of locally finite variation with $A_0 = 0$, let $L(A)$ denote the $\mathbb{K}$-vector space of all progressive processes $H\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}$ such that

$$\int_0^t |H_s| \, \mathrm{d}\mathbb{V}_A(s) < \infty, \qquad t \in \mathbb{R}_+, \tag{C.1}$$

where $\mathbb{V}_A(s)$ denotes the total variation of the process A on the interval $[0, s]$.

(b) For a $\mathbb{K}$-valued continuous local martingale M, let $L(M)$ denote the $\mathbb{K}$-vector space of all progressive processes $H\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}$ which are pathwise locally square integrable w.r.t. $[M]$, meaning that

$$\int_0^t |H_s|^2 \, \mathrm{d}[M]_s < \infty, \qquad t \in \mathbb{R}_+. \tag{C.2}$$

(c) For a $\mathbb{K}$-valued continuous semimartingale X with canonical decomposition $X = A+M$, we define the space $L(X)$ of processes which are stochastically integrable against X by

$$L(X) = L(A) \cap L(M). \tag{C.3}$$

(d) For a $\mathbb{K}^d$-valued continuous semimartingale $X = (X^1, \dots, X^d)^\mathsf{T}$ and a progressive process $V\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^{n\times d}$ with components $V^{i,j}$ for $(i,j) \in \{1,\dots,n\} \times \{1,\dots,d\}$, we write $V \in L_n(X)$ (and just $L(X)$ in the case $n = 1$) if $V^{i,j} \in L(X^j)$ for all $(i,j) \in \{1,\dots,n\} \times \{1,\dots,d\}$.

Remark C.2. If X is a $\mathbb{K}^d$-valued continuous semimartingale and $H \in L_n(X)$, then the stochastic integral $\int_0^\cdot H_s \,\mathrm{d}X_s$ is well defined and a continuous semimartingale with decomposition $\int_0^\cdot H_s \,\mathrm{d}A_s + \int_0^\cdot H_s \,\mathrm{d}M_s$. We refer to [53, Definition 5.110 and Remark 5.111] for details.

Definition C.3 (The usual completeness assumption, [53, Assumption 5.54])**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $I \subseteq \overline{\mathbb{R}}_+$ an interval (possibly containing ∞) and $t^* := \sup I$.

(a) A filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in I\cup\{t^*\}}$ in $\mathcal{F}$ is said to satisfy the usual completeness assumption, if $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets of $\mathcal{F}_{t^*}$.

(b) A filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in I}$ with $t^* \notin I$ is said to satisfy the usual completeness assumption, if this is the case for $(\mathcal{F}_t)_{t\in I\cup\{t^*\}}$ in the sense of part (a), where $\mathcal{F}_{t^*} := \sigma(\bigcup_{t\in I} \mathcal{F}_t)$.

Except for Chapter 3, where a finite time horizon is considered, we will always have $I = \mathbb{R}_+$ or $I = \overline{\mathbb{R}}_+$.

Remark C.4. (a) Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$, we can always construct an enlarged filtration $\overline{\mathbb{F}} = (\bar{\mathcal{F}}_t)_{t\in\mathbb{R}_+}$ satisfying Definition C.3 by replacing $\mathcal{F}_t$ with its augmentation $\bar{\mathcal{F}}_t$ by $(\mathcal{F}_\infty, \mathbb{P})$-null sets (Definition B.3 and Lemma B.4). By [53, Exercise 3.20 and Remark 5.55], $\overline{\mathbb{F}}$ inherits right-continuity of the filtration $\mathbb{F}$ if this is the case, and stochastic processes on $(\Omega, \mathcal{F}, \mathbb{P})$ which are (sub- or super-)martingales or have independent increments w.r.t. $\mathbb{F}$ also have the respective property w.r.t. $\overline{\mathbb{F}}$.

(b) The completeness assumption from Definition C.3 is needed to establish many foundational results in stochastic analysis, e.g. the Banach space property of the space $\mathcal{M}^p$ of $L^p(\mathbb{P})$ bounded ($p \in (1,\infty)$) continuous martingales started from zero [53, Theorem 5.56], the existence of the pathwise stochastic integral for càglàd adapted processes [53, Theorem 5.59], the existence of the covariation process [53, Theorem 5.77], the construction of the abstract stochastic integral [53, Theorem 5.99], Girsanov's theorem [53, Theorem 7.41], the predictable integral representation of Brownian local martingales [53, Theorem 7.102] and the strong existence and pathwise uniqueness theorem C.34 for solutions of SDEs.

(c) In the literature, the precise completeness condition imposed on the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ in the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ under consideration varies. E.g. in [38], Definition C.3 above is replaced with the notion of completeness of $\mathbb{F}$, which is defined [38, p. 42] to hold if $\mathcal{F}_0$ contains all subsets of $(\mathcal{F}_\infty, \mathbb{P})$-null sets. Often, $\mathbb{F}$ is required to satisfy the even stronger "usual conditions", defined in [35, Definition 2.25] and [50, §67] as $\mathbb{F}$ being right-continuous and $\mathcal{F}_0$ containing all subsets of $(\mathcal{F}, \mathbb{P})$-null sets. The completeness condition from Definition C.3 is a weaker requirement than all of the above, but nevertheless sufficient for the development of large parts of the theory.

Recall that $W_n = C(\mathbb{R}_+, \mathbb{K}^n)$ carries the σ-algebra $\mathcal{G}$ and the filtration $\mathbb{G} = (\mathcal{G}_t)_{t\in\mathbb{R}_+}$ generated by the canonical projections on W_n, see Section 2.1 for details. The following definition captures the minimal context needed to define solutions of SDEs.

Definition C.5 (Set-up, stochastic base). (a) A filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ satisfying the usual completeness assumption (Definition C.3) is called a *stochastic base.*

(b) A triple $((\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}), B, \xi)$ of a stochastic base, an $\mathbb{R}^d$ or $\mathbb{K}^d$-valued standard $(\mathbb{F}, \mathbb{P})$-Brownian motion $B = (B_t)_{t\in\mathbb{R}_+}$ and an $\mathcal{F}_0$-measurable $\mathbb{K}^n$-valued random vector ξ is called a *set-up* (for an SDE). We will usually write $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$ instead of $((\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}), B, \xi)$.

Definition C.6 (The filtration $\mathbb{G} \otimes \mathbb{F}$). Let $(\Omega, \mathcal{F}, \mathbb{F})$ with $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be a filtered measurable space and consider the measurable space $(W_n, \mathcal{G})$ equipped with the filtration $\mathbb{G} = (\mathcal{G}_t)_{t\in\mathbb{R}_+}$ (see (2.4) and (2.5)). We denote by $\mathbb{G} \otimes \mathbb{F}$ the filtration $(\mathcal{G}_t \otimes \mathcal{F}_t)_{t\in\mathbb{R}_+}$ of $(W_n \times \Omega, \mathcal{G} \otimes \mathcal{F})$.

Progressive measurability with respect to the filtrations $\mathbb{G}$ resp. $\mathbb{G} \otimes \mathbb{F}$ is the measurability requirement placed on coefficients of SDEs in Definition C.8 below. The next lemma is important to ensure that the stochastic integral equations implicit in these SDEs (see (C.13)) are well-defined.

Lemma C.7 ($\mathbb{G} \otimes \mathbb{F}$-progressive measurability and composition). *Let $(\Omega, \mathcal{F}, \mathbb{F})$ with $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be a filtered measurable space and $X\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^n$ be an $\mathbb{F}$-progressive process.*

(a) Let $f\colon \mathbb{R}_+ \times W_n \times \Omega \to \mathbb{K}^m$ be $\mathbb{G} \otimes \mathbb{F}$-progressive. Then the process

$$\mathbb{R}_+ \times \Omega \ni (t, \omega) \mapsto Y_t(\omega) := f(t, X(\cdot, \omega), \omega) \in \mathbb{K}^m$$

is $\mathbb{F}$-progressive.

(b) [53, Lemma 8.2] Let $f\colon \mathbb{R}_+ \times W_n \to \mathbb{K}^m$ be $\mathbb{G}$-progressive. Then the process

$$\mathbb{R}_+ \times \Omega \ni (t, \omega) \mapsto Y_t(\omega) := f(t, X(\cdot, \omega)) \in \mathbb{K}^m$$

is $\mathbb{F}$-progressive.

Proof. The proof is similar to [53, Lemma 8.2]. Part (b) follows from part (a), since every $\mathbb{G}$-progressive function $f\colon \mathbb{R}_+ \times W_n \to \mathbb{K}^m$ is $\mathbb{G} \otimes \mathbb{F}$-progressive when viewed as a function on $\mathbb{R}_+ \times W_n \times \Omega$. For part (a), fix $t \in \mathbb{R}_+$ and observe that the restriction $Y\restriction_{[0,t]\times\Omega}$ can be written as the composition of $f\restriction_{[0,t]\times W_n\times\Omega}$ with the map

$$[0, t] \times \Omega \ni (s, \omega) \mapsto \varphi(s, \omega) := (s, X(\cdot, \omega), \omega) \in [0, t] \times W_n \times \Omega.$$

We check that φ is $(\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t)$-$(\mathcal{B}_{[0,t]} \otimes \mathcal{G}_t \otimes \mathcal{F}_t)$-measurable by verifying that the first, second and third projections of φ are measurable w.r.t. $\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t$ and $\mathcal{B}_{[0,t]}$, $\mathcal{G}_t$ and $\mathcal{F}_t$, respectively, see also [53, Lemma 15.35]. Clearly, the first and third component maps of φ have the required $\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t$-measurability, and this is also true for the second component map $[0,t] \times \Omega \ni (s,\omega) \mapsto X(\omega) \in W_n$, using that X is $\mathbb{F}$-progressive. By assumption, $f\restriction_{[0,t]\times W_n \times \Omega}$ is $(\mathcal{B}_{[0,t]} \otimes \mathcal{G}_t \otimes \mathcal{F}_t)$-measurable, hence the composition $Y\restriction_{[0,t]\times\Omega} = f\restriction_{[0,t]\times W_n\times\Omega} \circ \varphi$ is $\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t$-measurable. □

We consider classical stochastic differential equations in the sense of Itô with coefficients depending on time, the (past) trajectory of the solution process and, possibly, the underlying randomness. The dependence is assumed to be non-anticipative in the sense that the future evolution of the solution process at every time $t \in \mathbb{R}_+$ is determined only by information from times up to t. That is, we admit as coefficients $\mathbb{G}$-progressive functions

$$b\colon \mathbb{R}_+ \times W_n \to \mathbb{K}^n \quad \text{and} \quad \sigma\colon \mathbb{R}_+ \times W_n \to \mathbb{K}^{n\times d}, \tag{C.4}$$

or, given a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ (Definition C.5 (a)), $\mathbb{G} \otimes \mathbb{F}$-progressive functions

$$b\colon \mathbb{R}_+ \times W_n \times \Omega \to \mathbb{K}^n \quad \text{and} \quad \sigma\colon \mathbb{R}_+ \times W_n \times \Omega \to \mathbb{K}^{n\times d}. \tag{C.5}$$

Definition C.8 (Classical stochastic differential equation)**.** A classical *stochastic differential equation* (SDE) with $\mathbb{G}$-progressive coefficients b and σ as in (C.4) has the form

$$\mathrm{d}X_t = b(t, X)\,\mathrm{d}t + \sigma(t, X)\,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+, \tag{C.6}$$

or, given a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and $\mathbb{G} \otimes \mathbb{F}$-progressive coefficients b and σ as in (C.5),

$$\mathrm{d}X_t = b(t, X, \,\cdot\,)\,\mathrm{d}t + \sigma(t, X, \,\cdot\,)\,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+, \tag{C.7}$$

where in both cases $B = (B_t)_{t\in\mathbb{R}_+}$ is an $\mathbb{R}^d$- or $\mathbb{C}^d$-valued Brownian motion. In the first case, we speak of *nonrandom dependence* of the SDE coefficients on the solution. Otherwise, we speak of *random dependence*.

Precise notions of what constitutes a solution of a stochastic differential equation are given in Definition C.15 and Definition C.20 below. First, we introduce some relevant terminology.

Definition C.9 (Instantaneous and time-homogeneous dependence)**.** Consider the setting of Definition C.8. If $\hat{b}\colon \mathbb{R}_+ \times \mathbb{K}^n \times \Omega \to \mathbb{K}^n$ and $\hat{\sigma}\colon \mathbb{R}_+ \times \mathbb{K}^n \times \Omega \to \mathbb{K}^{n\times d}$ are such that for all $(t, w, \omega) \in \mathbb{R}_+ \times W_n \times \Omega$,

$$b(t, w, \omega) = \hat{b}(t, w_t, \omega), \quad \text{and} \quad \sigma(t, w, \omega) = \hat{\sigma}(t, w_t, \omega),$$

then we speak of *instantaneous dependence* of the SDE coefficients on the solution. In this case, it is customary to make no notational distinction between b and $\hat{b}$ and σ and $\hat{\sigma}$, respectively, and write the SDE (C.7) as

$$\mathrm{d}X_t = b(t, X_t, \,\cdot\,)\,\mathrm{d}t + \sigma(t, X_t, \,\cdot\,)\,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+. \tag{C.8}$$

Similarly, if $\hat{b}$ and $\hat{\sigma}$ do not depend on ω, we write

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+, \tag{C.9}$$

and if $\hat{b}$ and $\hat{\sigma}$ in addition do not depend on t, then we write

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t\,, \qquad t \in \mathbb{R}_+ \tag{C.10}$$

and speak of *instantaneous and time-homogeneous dependence* of the coefficients on the solution.[1]

Example C.10 (SDEs with random dependence arising in stochastic control theory)**.** [48, §3.2] In continuous time stochastic control theory, the state of a system is given by a stochastic process $X\colon \mathbb{R}_+ \times \Omega \to \mathbb{R}^n$ on a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ governed by an SDE

$$\mathrm{d}X_t = \tilde{b}(X_t, \alpha_t)\,\mathrm{d}t + \tilde{\sigma}(X_t, \alpha_t)\,\mathrm{d}B_t\,, \tag{C.11}$$

where the coefficients $\tilde{b}\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $\tilde{\sigma}\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n\times d}$ are measurable functions, $B = (B_t)_{t\in\mathbb{R}_+}$ is an $\mathbb{R}^d$-valued standard $(\mathbb{F}, \mathbb{P})$-Brownian motion and $\alpha\colon \mathbb{R}_+ \times \Omega \to \mathbb{R}^m$ is an $\mathbb{F}$-progressive process, the so called *control process*. The aim in stochastic control theory is to find an optimal control process α maximizing a given objective function depending on X. For fixed α, the SDE (C.11) is of the form (C.8) with coefficients

$$b(t, w, \omega) = \tilde{b}(w_t, \alpha_t(\omega)), \qquad \sigma(t, w, \omega) = \tilde{\sigma}(w_t, \alpha_t(\omega))$$

for all $(t, w, \omega) \in \mathbb{R}_+ \times W_n \times \Omega$.

Example C.11 (Local stochastic volatility model)**.** Another example of an SDE with random dependence is given by the local stochastic volatility model, which extends the Black–Scholes model by allowing the risk-free rate and stock volatility to be stochastic processes. See (3.5) in Definition 3.1 for details.

Remark C.12 (The role of the stochastic base in the setting of nonrandom vs. random dependence)**.**

(a) In the case of random dependence of coefficients, a stochastic base is given, while with nonrandom dependence, this need not be the case. This distinction leads to slight differences in the definitions of strong existence (Definition C.19), weak solutions and weak existence (Definition C.20) and weak uniqueness (Definition C.21) below.

(b) In the setting of nonrandom dependence, the question arises to which degree the choice of the stochastic base influences the behaviour of possible solutions. This topic is revisited in Appendix C.4, with the result that under continuity and bounded growth conditions (Definitions C.25 and C.26) on the coefficients, the stochastic base has only very little influence, see also Theorem C.43.

[1] Solutions of SDEs of the form (C.10) with Lipschitz-continuous coefficients have the special property of being Markov processes, see Theorem C.62.

Remark C.13 (Simplified measurability criterion for coefficients with nonrandom and instantaneous dependence)**.** Consider the SDE (C.9), i.e. the case when the dependence of coefficients on the solution is nonrandom and instantaneous. We claim that if the coefficients $b\colon \mathbb{R}_+ \times \mathbb{K}^n \to \mathbb{K}^n$ and $\sigma\colon \mathbb{R}_+ \times \mathbb{K}^n \to \mathbb{K}^{n\times d}$ are $\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{K}^n}$-measurable, then they are $\mathbb{G}$-progressively measurable when viewed as maps on $\mathbb{R}_+ \times W_n$ and thus admissible coefficients in the sense of Definition C.8. To make this explicit, let $\hat{b}\colon \mathbb{R}_+ \times \mathbb{K}^n \to \mathbb{K}^n$ and $\hat{\sigma}\colon \mathbb{R}_+ \times \mathbb{K}^n \to \mathbb{K}^{n\times d}$ be $\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{K}^n}$-measurable functions. Since the canonical process $\mathbb{R}_+ \times W_n \ni (t, w) \mapsto \pi_t(w) = w_t$ on $(W_n, \mathcal{G}, \mathbb{G})$ is continuous and adapted, it is $\mathbb{G}$-progressive. Therefore,

$$b(t, w) := \hat{b}(t, w_t), \quad \text{and} \quad \sigma(t, w) := \hat{\sigma}(t, w_t), \qquad (t, w) \in \mathbb{R}_+ \times W_n$$

define $\mathbb{G}$-progressive maps and thus admissible coefficients for the SDE (C.9).

Convention C.14 (Uniform treatment of SDEs with random and nonrandom dependence)**.** Given a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, every SDE of the form (C.6) (i.e. with nonrandom dependence) can be naturally viewed as an SDE of the form (C.7). Hence we adopt the following convention: in every context in which the stochastic base is given, definitions made for the coefficients (C.5) or the associated SDE (C.7) also include coefficients of the form (C.4) and the associated SDE (C.6), respectively.

Definition C.15 (Strong solution of a stochastic differential equation [53, Definition 8.3])**.** Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$ be a set-up (Definition C.5(b)), and consider $\mathbb{G} \otimes \mathbb{F}$-progressive coefficients b and σ as in (C.5). A continuous $\mathbb{K}^n$-valued $\mathbb{F}$-adapted stochastic process X is called a *strong solution* of the SDE (C.7) with initial condition ξ, if:

(a) The initial value satisfies $X_0 \overset{\text{a.s.}}{=} \xi$.

(b) The stochastic integrals implicitly given by (C.6) are well defined,[2] i.e.,

$$\int_0^t |b(s, X, \cdot\,)| + |\sigma(s, X, \cdot\,)|^2 \, \mathrm{d}s < \infty, \qquad t \in \mathbb{R}_+. \tag{C.12}$$

(c) The process X satisfies the integral version of (C.6), i.e.

$$X \overset{\text{uti}}{=} X_0 + \int_0^{\cdot} b(s, X, \cdot\,)\, \mathrm{d}s + \int_0^{\cdot} \sigma(s, X, \cdot\,)\, \mathrm{d}B_s\,. \tag{C.13}$$

Remark C.16 (Tolerated abuse of notation in (C.13))**.** In (C.13), the dots in the upper bounds of the integrals refer to the time argument, while the dots in the last argument of b and σ refer to ω. We tolerate this abuse of notation, since highlighting the presence of random dependence of the coefficients will improve clarity in the sequel.

Sometimes, if is no ambiguity regarding the stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, the driving Brownian motion B and the initial random vector ξ, we will refer to a strong solution of (C.7) w.r.t. the set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$ simply as a solution.

[2] Note that we require the inequalities in (C.12) to hold surely for all $t \in \mathbb{R}_+$ rather than almost surely, which is also common in the literature.

Remark C.17 (Strong solutions are continuous semimartingales)**.** Consider the setting of Definition C.15. Lemma C.7 ensures that the integrands on the right-hand side in (C.13) are $\mathbb{F}$-progressively measurable. Furthermore, the integrability condition (C.12) ensures that the stochastic integrals are well-defined in the sense of Definition C.1. Hence from Remark C.2, the right-hand side of (C.13) is a continuous semimartingale. Since X is adapted, continuous and by (C.13) equal up to indistinguishability to a continuous semimartingale, X is itself a continuous semimartingale.[3]

Remark C.18 (Strong solutions and indistinguishability)**.** Fix a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$ and consider $\mathbb{G}$-progressive functions b and σ as in (C.5). Let X be a strong solution of (C.7).

(a) If $\tilde{B}$ is a standard $(\mathbb{F}, \mathbb{P})$-Brownian motion with $\tilde{B} \overset{\text{uti}}{=} B$ and $\tilde{\xi}$ is a $\mathbb{K}^n$-valued $\mathcal{F}_0$-measurable random vector with $\tilde{\xi} \overset{\text{a.s.}}{=} \xi$, then X is also a strong solution of the SDE (C.7) on the set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \tilde{B}, \tilde{\xi})$. To see this, note that since X is a strong solution, part (a) of Definition C.15 also holds for $\tilde{\xi}$ instead of ξ and part (b) does not depend on the pair of initial condition and Brownian motion at all. For part (c), since (C.13) holds by assumption, it also holds for $\tilde{B}$ instead of B by invariance up to indistinguishability of the stochastic integral under passage to an indistinguishable modification of the integrator, see also [53, Exercise 5.113].

(b) Let $\tilde{X} = (\tilde{X}_t)_{t\in\mathbb{R}_+}$ be an $\mathcal{F}_\infty$-measurable continuous stochastic process (meaning that $\tilde{X}_t$ is $\mathcal{F}_\infty$-measurable for all $t \in \mathbb{R}_+$) indistinguishable from X. If the integrability condition (C.12) holds for $\tilde{X}$ in place of X, then $\tilde{X}$ is also a strong solution of (C.7). To see this, observe first that since $\mathbb{F}$ satisfies the usual completeness condition (Definition C.3), adaptedness of X implies that of $\tilde{X}$ by Corollary B.6(b). Part (a) of Definition C.15 holds for $\tilde{X}$ by indistinguishability, and stability up to indistinguishability of stochastic integrals under passage to an indistinguishable integrand (see e.g. [53, Exercise 5.113]) implies that $\tilde{X}$ satisfies also part (c) of Definition C.15.

(c) If b and σ satisfy additionally

$$\int_0^t |b(s, w, \omega)| + |\sigma(s, w, \omega)|^2 \, \mathrm{d}s < \infty, \qquad (t, w, \omega) \in \mathbb{R}_+ \times W_n \times \Omega, \tag{C.14}$$

which is guaranteed e.g. by the linearly bounded growth condition (Definition C.27 below), then by part (b), every continuous $\mathcal{F}_\infty$-measurable stochastic process $\tilde{X}$ which is indistinguishable from a strong solution X of (C.7) is also a strong solution of (C.7).

[3] To verify that X is indeed a continuous semimartingale, let $\tilde{X}$ denote the right-hand side of (C.13), and let $\tilde{X} = \tilde{M} + \tilde{A}$ be a decomposition into a continuous adapted finite variation process and a continuous local martingale. Let $N := \{\omega \in \Omega : X(\omega) \neq \tilde{X}(\omega)\}$, which is an $(\mathcal{F}_\infty, \mathbb{P})$-null set. Since $\mathbb{F}$ satisfies by definition C.5(a) of a set-up the usual completeness condition (Definition C.3), we have $N \in \mathcal{F}_0$. We can write

$$X = X\,\mathbb{1}_N + X\,\mathbb{1}_{N^c} = X\,\mathbb{1}_N + \tilde{M}\,\mathbb{1}_{N^c} + \tilde{A}\,\mathbb{1}_{N^c} = M + A$$

with $M := X\,\mathbb{1}_N + \tilde{M}\,\mathbb{1}_{N^c}$ and $A := \tilde{A}\,\mathbb{1}_{N^c}$, which is a continuous adapted finite variation process started from zero. Using that M is adapted, continuous and equal up to indistinguishability to the continuous local martingale $\tilde{M}$, it is straightforward to check that M is also a continuous local martingale with the same localizing sequences as $\tilde{M}$. Hence X is a continuous semimartingale.

Definition C.19 (Strong existence [53, Definition 8.7])**.** Let $\mu \in \mathcal{P}(\mathbb{K}^n)$ be an initial distribution.

(a) For $\mathbb{G}$-progressive functions b and σ as in (C.4), we say that there is strong existence of a solution of the SDE (C.6) with initial distribution μ for $\mathbb{R}^d$ - or $\mathbb{K}^d$-valued Brownian motion, if for every set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B)$ (Definition C.5(b)) with $\mathcal{L}_{\mathbb{P}}[\xi] = \mu$ and where the Brownian motion B is $\mathbb{R}^d$- or $\mathbb{K}^d$-valued, respectively, there exists a strong solution in the sense of Definition C.15.

(b) Fix a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ Definition C.5(a). For $\mathbb{G} \otimes \mathbb{F}$-progressive functions as in (C.5), we say that there is strong existence of a solution of the SDE (C.7) with initial distribution μ for $\mathbb{R}^d$- or $\mathbb{K}^d$-valued Brownian motion, if for every $\mathbb{R}^d$ respectively $\mathbb{K}^d$-valued Brownian motion $B = (B_t)_{t \in \mathbb{R}_+}$ and every $\mathcal{F}_0$-measurable initial random vector ξ with $\mathcal{L}_{\mathbb{P}}[\xi] = \mu$ there exists a strong solution in the sense of Definition C.15.

Definition C.20 (Weak solution and weak existence [53, Definition 8.9])**.** Let $\mu \in \mathcal{P}(\mathbb{K}^n)$ be an initial distribution.

(a) Let b and σ be $\mathbb{G}$-progressive functions as in (C.4). A *weak solution* of the SDE (C.6) with initial law μ for $\mathbb{R}^d$- or $\mathbb{K}^d$-valued Brownian motion is a pair $(\mathfrak{S}, X)$ of a set-up $\mathfrak{S} := (\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B)$ (Definition C.5(b)) where $\mathcal{L}_{\mathbb{P}}[\xi] = \mu$ and the Brownian motion B is $\mathbb{R}^d$-valued or $\mathbb{K}^d$-valued, respectively, and a continuous $\mathbb{K}^n$-valued $\mathbb{F}$-adapted stochastic process $X = (X_t)_{t \in \mathbb{R}_+}$ which is a strong solution (Definition C.15) of (C.6) on the set-up $\mathfrak{S}$.[4]

(b) Fix a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ (Definition C.5(a)) and let b and σ be $\mathbb{G} \otimes \mathbb{F}$-progressive coefficients as in (C.5). A *weak solution* of the SDE (C.7) with initial law μ for $\mathbb{R}^d$- or $\mathbb{K}^d$-valued Brownian motion is a triple (B, ξ, X) of an $\mathbb{R}^d$-valued respectively $\mathbb{K}^d$-valued Brownian motion B, an $\mathcal{F}_0$-measurable initial random vector ξ with $\mathcal{L}_{\mathbb{P}}[\xi] = \mu$ and a continuous $\mathbb{K}^n$-valued $\mathbb{F}$-adapted stochastic process $X = (X_t)_{t \in \mathbb{R}_+}$ which is a strong solution (Definition C.15) of (C.7) on the set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$.

(c) When a weak solution of (C.6) respectively (C.7) exists for the initial distribution μ, we speak of *weak existence* for this initial distribution.

Definition C.21 (Weak uniqueness, uniqueness in law for SDEs with nonrandom dependence, [53, Definition 8.12])**.** Let $\mu \in \mathcal{P}(\mathbb{K}^n)$ be an initial distribution and consider $\mathbb{G}$-progressive coefficients b and σ as in (C.4). We say that there is *weak uniqueness* or *uniqueness in law* of solutions of the SDE (C.6) for the initial distribution μ and $\mathbb{R}^d$- or $\mathbb{K}^d$-valued Brownian motion, respectively, if every weak solution X in the sense of Definition C.20(a) with initial distribution μ has the same law on W_n.

Definition C.22 (Pathwise uniqueness [53, Definition 8.14])**.** Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic base (Definition C.5(a)) and fix $\mathbb{G} \otimes \mathbb{F}$-progressive coefficients b and σ as in (C.5) and an initial distribution $\mu \in \mathcal{P}(\mathbb{K}^n)$. We say that there is *pathwise uniqueness* for the SDE (C.7)

[4] To simplify the notation, we will sometimes write (B, X) or even more simply X to denote a weak solution $((\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B), X)$, leaving the other elements of the set-up implicit. The same convention applies in the setting of random dependence (Definition C.20(b)).

with initial distribution μ for $\mathbb{R}^d$- or $\mathbb{K}^d$-valued Brownian motion, respectively, if whenever (B, X) and $(\tilde{B}, \tilde{X})$ are two weak solutions of (C.7) with initial distribution μ defined on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with $B \overset{\text{uti}}{=} \tilde{B}$ (u.t.i. identical Brownian motions) and $X_0 \overset{\text{a.s.}}{=} \tilde{X}_0$ (a.s. equal initial conditions), then $X \overset{\text{uti}}{=} \tilde{X}$.

Remark C.23 (Pathwise and weak uniqueness)**.** (a) For SDEs with nonrandom dependence of coefficients (Definition C.9), pathwise uniqueness (Definition C.22) implies weak uniqueness by the Yamada–Watanabe theorem. We will prove this under some additional conditions, see Corollary C.57 below and the remark preceding it.

(b) In the case of random dependence, i.e. when considering the SDE (C.7) with $\mathbb{G} \otimes \mathbb{F}$-progressive coefficients b and σ on a given stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ as in (C.5), a natural definition of weak uniqueness of solutions of (C.7) for the initial distribution μ would be requiring that every weak solution X in the sense of Definition C.20 (b) with initial distribution μ has the same law on W_n. But this is too strong, as Example C.24 below shows that this form of weak uniqueness can fail already for very simple coefficients, even in the presence of pathwise uniqueness.

Example C.24 (Failure of weak uniqueness despite pathwise uniqueness in the setting of random dependence)**.** Fix $\mathbb{K} = \mathbb{R}$ and $n = d = 1$ and let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic base (Definition C.5 (a)). Let ξ be an $\mathcal{F}_0$-measurable initial random variable with the uniform distribution[5] $\mathcal{U}_{[-1,1]}$ on $[-1, 1]$ and $B = (B_t)_{t\in\mathbb{R}_+}$ an $\mathbb{R}$-valued standard Brownian motion B. Consider coefficients

$$b(t, x, \omega) := \xi(\omega), \qquad \sigma(t, x, \omega) := 0, \qquad (t, x, \omega) \in \mathbb{R}_+ \times \mathbb{R} \times \Omega,$$

which have the Lipschitz property (Definition C.25) and are bounded (Definition C.26). Since ξ is $\mathcal{F}_0$-measurable, the coefficients are $\mathbb{G} \otimes \mathbb{F}$-progressive, hence the SDE (C.8) is well defined, and we have strong existence and pathwise uniqueness by Theorem C.34 below. Despite this, we do not have weak uniqueness in the sense of Remark C.23 (b). To see this, note that if $X = (X_t)_{t\in\mathbb{R}_+}$ is a strong solution, then

$$X_t \overset{\text{a.s.}}{=} \xi + \int_0^t b(s, X_s, \cdot\,)\, \mathrm{d}s = (1 + t)\xi, \qquad t \in \mathbb{R}_+.$$

Now consider the same SDE for the set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \tilde{\xi}, B)$, where $\tilde{\xi} := -\xi$. The up to indistinguishability unique strong solution $\tilde{X}$ of (C.8) on this set-up is given by

$$\tilde{X}_t \overset{\text{a.s.}}{=} \tilde{\xi} + \int_0^t b(s, \tilde{X}_s, \cdot\,)\, \mathrm{d}s = (t - 1)\xi, \qquad t \in \mathbb{R}_+.$$

In particular, $\tilde{X}_1 \overset{\text{a.s.}}{=} 0$, while X_1 is uniformly distributed on $[-2, 2]$. Hence (B, ξ, X) and $(B, \tilde{\xi}, \tilde{X})$ are weak solutions of the SDE (C.8) with the same initial distribution $\mathcal{U}_{[-1,1]}$ (Definition C.20 (b)), but $\mathcal{L}[X] \neq \mathcal{L}[\tilde{X}]$.

[5] Every bounded nondegenerate symmetric probability distribution on $\mathbb{R}$ works for this example.

C.2. Strong Existence and Pathwise Uniqueness for Classical SDEs

In this section, we develop the main result for strong existence and pathwise uniqueness of solutions of (C.7), Theorem C.34, under combinations of the following conditions.

Definition C.25 (Lipschitz property, [53, Assumption 8.24])**.** Fix a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. Two coefficients b and σ as in (C.5) are said to have the Lipschitz property, if there exists an increasing function $c\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$|b(t, w, \omega) - b(t, \tilde{w}, \omega)| + |\sigma(t, w, \omega) - \sigma(t, \tilde{w}, \omega)| \le c_t p_t(w - \tilde{w}) \tag{C.15}$$

for all $t \in \mathbb{R}_+$, all $w, \tilde{w} \in W_n$ and all $\omega \in \Omega$, where p_t denotes the seminorm given by (2.1).

Definition C.26 (Boundedness in the second argument [53, Assumption 8.25])**.** Fix a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. Two coefficients b and σ as in (C.5) are said to be bounded in the second argument, if there exists an increasing function $D\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$|b(t, w, \omega)| + |\sigma(t, w, \omega)| \le D_t, \qquad (t, w, \omega) \in \mathbb{R}_+ \times W_n \times \Omega. \tag{C.16}$$

Definition C.27 (Linearly bounded growth in the second argument, [53, Assumption 8.26])**.** Fix a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. Two coefficients b and σ as in (C.5) are said to be of linearly bounded growth in the second argument, if there exists an increasing function $D\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$|b(t, w, \omega)| + |\sigma(t, w, \omega)| \le D_t \left(1 + p_t(w)\right), \qquad (t, w, \omega) \in \mathbb{R}_+ \times W_n \times \Omega \tag{C.17}$$

Remark C.28. By the triangle inequality for the seminorm p_t, the Lipschitz property (Definition C.25) implies the bounded growth condition (Definition C.27) as soon as there exists an increasing function $\tilde{D}\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$|b(t, 0, \omega)| + |\sigma(t, 0, \omega)| \le \tilde{D}_t, \qquad (t, \omega) \in \mathbb{R}_+ \times \Omega.$$

This is the case in particular in the case of nonrandom dependence, if in addition the functions $b(\,\cdot\,, 0)$ and $\sigma(\,\cdot\,, 0)$ are continuous, càdlàg, càglàd or, more generally, regulated.

For coefficients with the Lipschitz property, weaker measurability properties ensure the $\mathbb{G}$- or $\mathbb{G} \otimes \mathbb{F}$-progressive measurability required of coefficients of (C.6) or (C.7), respectively.

Lemma C.29 (Weaker measurability requirements for coefficients with the Lipschitz property)**.**

(a) *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$ be a filtered probability space. If $f\colon \mathbb{R}_+ \times W_n \times \Omega \to \mathbb{K}^m$ has the Lipschitz property (i.e. (C.15) holds for f instead of b with $\sigma = 0$) and is such that $(t, \omega) \mapsto f(t, w, \omega)$ is $\mathbb{F}$-progressive for every $w \in W_n$, then f is $\mathbb{G} \otimes \mathbb{F}$-progressive.*

(b) *If $f\colon \mathbb{R}_+ \times W_n \to \mathbb{K}^m$ has the Lipschitz property (i.e. (C.15) holds for f instead of b with $\sigma = 0$) and is such that $\mathbb{R}_+ \ni t \mapsto f(t, w) \in \mathbb{K}^m$ is Borel-measurable for every $w \in W_n$, then f is $\mathbb{G}$-progressive.*

Proof. (a) Fix $t \in \mathbb{R}_+$. By assumption,

$$[0,t] \times \Omega \ni (s,\omega) \mapsto f\restriction_{[0,t]\times W_n \times \Omega}(s,w,\omega) \in \mathbb{K}^m$$

is $\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t$-measurable for each $w \in W_n$. Since f has the Lipschitz property, $w \mapsto f\restriction_{[0,t]\times W_n\times\Omega}(s,w,\omega)$ is continuous w.r.t. the seminorm p_t from (2.1) for all $(s,\omega) \in [0,t] \times \Omega$. Using Lemma B.10 and Lemma 2.1(c), it follows that $f\restriction_{[0,t]\times W_n\times\Omega}$ is $\mathcal{B}_{[0,t]} \otimes \mathcal{G}_t \otimes \mathcal{F}_t$-measurable.

(b) This follows directly from part (a). □

Remark C.30 (Estimates for absolute p-th powers of sums and integrals arising from Jensen's inequality [53, Remark 8.30])**.** For later reference, note that for every $n \in \mathbb{N}$ and $p \in [1,\infty)$ and all $a_1, \dots, a_n \in \mathbb{C}$, the triangle inequality for the absolute value on $\mathbb{C}$ and Jensen's inequality for the uniform probability distribution on $\{1,\dots,n\}$ imply that

$$|a_1 + \dots + a_n|^p = n^p \left|\frac{a_1 + \dots + a_n}{n}\right|^p \le n^{p-1}\left(|a_1|^p + \dots + |a_n|^p\right). \tag{C.18}$$

Similarly, for every bounded interval $I \subset \mathbb{R}$ of length $l > 0$ and every measurable function $f\colon I \to \mathbb{C}$, the triangle inequality for the absolute value of a $\mathbb{C}$-valued integral and Jensen's inequality applied to the uniform probability distribution on I yield

$$\left|\int_I f(s)\mathrm{d}s\right|^p = l^p \left|\int_I \frac{f(s)}{l}\,\mathrm{d}s\right|^p \le l^p \int_I \frac{|f(s)|^p}{l}\,\mathrm{d}s = l^{p-1}\int_I |f(s)|^p \mathrm{d}s. \tag{C.19}$$

Under Lipschitz continuity and boundedness assumptions, Theorem C.34 gives strong existence and pathwise uniqueness[6] of solutions of (C.7). For the proof of this result, we require the following lemma, which establishes proximity in an L^p-sense of solutions whose initial values are close w.r.t. the L^p-norm.

Lemma C.31 (L^p-continuous dependence of certain stochastic integrals on the input data)**.** *Fix $p \in [2,\infty)$, a stochastic base $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$, an $\mathbb{R}^d$- or $\mathbb{K}^d$-valued standard $(\mathbb{F},\mathbb{P})$-Brownian motion $B = (B_t)_{t\in\mathbb{R}_+}$, and consider $\mathbb{G}\otimes\mathbb{F}$-progressive coefficients b and σ as in (C.5) which are Lipschitz continuous in the second argument (Definition C.25).*

(a) Let X and Y be two $\mathbb{K}^n$-valued continuous semimartingales such that

$$\begin{aligned}\tilde{X} &\coloneqq X_0 + \int_0^{\cdot} b(s,X,\,\cdot\,)\,\mathrm{d}s + \int_0^{\cdot} \sigma(s,X,\,\cdot\,)\,\mathrm{d}B_s \quad \textit{and}\\ \tilde{Y} &\coloneqq Y_0 + \int_0^{\cdot} b(s,Y,\,\cdot\,)\,\mathrm{d}s + \int_0^{\cdot} \sigma(s,Y,\,\cdot\,)\,\mathrm{d}B_s\end{aligned} \tag{C.20}$$

are well-defined.[7] Then for all $t \in \mathbb{R}_+$,

$$\mathbb{E}\big[p_t^p(\tilde{X}-\tilde{Y})\big] \le 3^{p-1}\mathbb{E}[|X_0 - Y_0|^p] + \beta(t)\int_0^t c_s^p\,\mathbb{E}[p_s^p(X-Y)]\,\mathrm{d}s\,, \tag{C.21}$$

[6] The question of weak uniqueness is addressed in Appendix C.4.

[7] By Lemma C.7, the integrands in (C.20) are $\mathbb{F}$-progressive. Hence for well-definedness of the stochastic integrals in (C.20), it suffices that the stochastic integrability condition (C.12) holds for both X and Y.

where the function c originates from Definition C.25,[8]

$$\beta(t) \coloneqq 6^{p-1}(t^{p-1} + C_p c_{\mathbb{K}}^{p/2} t^{p/2-1}), \qquad t \in \mathbb{R}_+, \tag{C.22}$$

where C_p originates from the Burkholder–Davis–Gundy inequality [53, Theorem 6.32] and $c_{\mathbb{K}} \coloneqq \dim_{\mathbb{R}}(\mathbb{K})$.

(b) Let $X = (X_t)_{t\in\mathbb{R}_+}$ and $Y = (Y_t)_{t\in\mathbb{R}_+}$ be two strong solutions of (C.7). Then for all $t \in \mathbb{R}_+$,

$$\mathbb{E}\left[p_t^p(X-Y)\right] \le 6^{p-1}\mathbb{E}[|X_0 - Y_0|^p] \exp\left(t(\beta(t) \vee c_t^p)^2\right). \tag{C.23}$$

Proof. The proof is based on Step 1 of the proof of [53, Theorem 8.27], with modifications to account for the distinct starting values X_0 and Y_0 and for the explicit dependence of the coefficients on ω.

(a) For each $N \in \mathbb{N}$ define $\tau_N = \inf\{t \in \mathbb{R}_+ : |X_t - Y_t| \ge N\}$, which is a stopping time by [53, Lemma 3.53(b)].[9] Since X and Y have continuous paths, it follows for every $\omega \in \Omega$ that the increasing sequence $(\tau_N(\omega))_{N\in\mathbb{N}}$ cannot have an accumulation point in any compact subset of $\mathbb{R}_+$, hence $\tau_N(\omega) \to \infty$ as $N \to \infty$.

Fix $N \in \mathbb{N}$ for the remaining part of this step. We will omit this index where appropriate. Define the processes $U\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^n$ and $V\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^{n\times d}$ by

$$\begin{aligned} U_s(\omega) &= \mathbb{1}_{[0,\tau_N(\omega)]}(s)(b(s, X(\cdot,\omega),\omega) - b(s, Y(\cdot,\omega),\omega)), \\ V_s(\omega) &= \mathbb{1}_{[0,\tau_N(\omega)]}(s)(\sigma(s, X(\cdot,\omega),\omega) - \sigma(s, Y(\cdot,\omega),\omega)) \end{aligned} \tag{C.24}$$

for all $s \in \mathbb{R}_+$ and $\omega \in \Omega$. To see that the processes U and V are progressive, note that $\mathbb{1}_{[0,\tau_N]}$ is adapted and left-continuous, hence progressive, and that the other terms in (C.24) are progressive by Lemma C.7. By the Lipschitz property of coefficients (Definition C.25), these processes satisfy, for every $s \in \mathbb{R}_+$ and $\omega \in \Omega$,

$$\begin{aligned} |U_s(\omega)| + |V_s(\omega)| &\le c_s\, \mathbb{1}_{[0,\tau_N(\omega)]}(s)\; p_s(X(\cdot,\omega) - Y(\cdot,\omega)) \\ &\le c_s\, p_s\left(X^{\tau_N}(\cdot,\omega) - Y^{\tau_N}(\cdot,\omega)\right) \end{aligned} \tag{C.25}$$

We will omit writing the argument ω in the following. Define

$$\tilde{u}_N(t) = \mathbb{E}[p_t^p(\tilde{X}^{\tau_N} - \tilde{Y}^{\tau_N})], \qquad t \in \mathbb{R}_+. \tag{C.26}$$

Using the defining equation (C.20) for $\tilde{X}$ and $\tilde{Y}$, linearity and commutation with stopping of the stochastic integral, the abbreviations from (C.24), the triangle inequality for the Euclidean norm and the seminorm p_t given by (2.1), and finally Jensen's inequality (see (C.18)), it follows that

$$\tilde{u}_N(t) \le 3^{p-1}\mathbb{E}\left[|X_0 - Y_0|^p + p_t^p\left(\int_0^{\cdot} U_s \,\mathrm{d}s\right) + p_t^p(V \cdot B)\right], \quad t \in \mathbb{R}_+. \tag{C.27}$$

[8] In fact, $\beta/2^{p-1}$ instead of β from (C.22) also works for this lemma, but defining β as in (C.22) is useful in other contexts, e.g. in Lemma C.32(b) and Theorem 2.32.

[9] Localization using the stopping times $(\tau_N)_{N\in\mathbb{N}}$ ensures a local integrability condition needed for the application of Grönwall's inequality in the proof of part 134 below. If $p_t(X)$ and $p_t(Y)$ are p-integrable for each $t \in \mathbb{R}_+$, then localization can be skipped. In particular, localization is not required to obtain part (a).

For each $t \in \mathbb{R}_+$, rewriting with (2.1) and using Jensen's inequality (see (C.19)) for $t > 0$,

$$p_t^p\left(\int_0^\cdot U_s \,\mathrm{d}s\right) = \sup_{u\in[0,t]}\left|\int_0^u U_s \,\mathrm{d}s\right|^p \le \left(\int_0^t |U_s|\mathrm{d}s\right)^p \le t^{p-1}\int_0^t |U_s|^p \,\mathrm{d}s. \tag{C.28}$$

Since $V\cdot B$ is a continuous local martingale with quadratic variation process $\operatorname{tr}([V\cdot B]) \overset{\text{uti}}{=} c_{\mathbb{K}}\int_0^\cdot |V_s|^2 \,\mathrm{d}s$, it follows from the Burkholder–Davis–Gundy inequality [53, Theorem 6.32] that

$$\begin{aligned}\mathbb{E}\left[p_t^p(V\cdot B)\right] &\le C_p c_{\mathbb{K}}^{p/2}\mathbb{E}\left[\left(\int_0^t |V_s|^2 \,\mathrm{d}s\right)^{p/2}\right]\\ &\le C_p c_{\mathbb{K}}^{p/2} t^{p/2-1}\mathbb{E}\left[\int_0^t |V_s|^p \,\mathrm{d}s\right],\end{aligned} \tag{C.29}$$

where for the last inequality we used again Jensen's inequality (see (C.19)) with the function $\mathbb{R}_+ \ni x \mapsto x^{p/2}$, which is convex since $p \ge 2$ by assumption. Plugging the estimates (C.28) and (C.29) into (C.27), using Fubini's theorem and applying (C.25) for the last inequality shows that, for each $t \in \mathbb{R}_+$,

$$\begin{aligned}\tilde{u}_N(t) &\le 3^{p-1}\mathbb{E}[|X_0 - Y_0|^p] + 3^{p-1}\int_0^t \mathbb{E}\left[t^{p-1}|U_s|^p + C_p c_{\mathbb{K}}^{p/2} t^{p/2-1}|V_s|^p\right]\mathrm{d}s\\ &\le 3^{p-1}\mathbb{E}[|X_0 - Y_0|^p] + \underbrace{3^{p-1}\left(t^{p-1} + C_p c_{\mathbb{K}}^{p/2} t^{p/2-1}\right)}_{=\beta(t)/2^{p-1} \text{ by (C.22)}}\int_0^t c_s^p u_N(s)\mathrm{d}s\end{aligned} \tag{C.30}$$

with

$$u_N(s) := \mathbb{E}\left[p_s^p\left(X^{\tau_N} - Y^{\tau_N}\right)\right], \quad s \in \mathbb{R}_+. \tag{C.31}$$

For each $t \in \mathbb{R}_+$, $p_t(\tilde{X}^{\tau_N} - \tilde{Y}^{\tau_N}) \nearrow p_t(\tilde{X} - \tilde{Y})$ and similarly $p_t(X^{\tau_N} - Y^{\tau_N}) \nearrow p_t(X - Y)$ pointwise on Ω as $N \to \infty$, hence from the definitions (C.26) and (C.31) of u and $\tilde{u}$,

$$\tilde{u}_N(t) \xrightarrow[N\to\infty]{} \mathbb{E}\big[p_t^p(\tilde{X} - \tilde{Y})\big] \quad \text{and} \quad u_N(t) \xrightarrow[N\to\infty]{} \mathbb{E}\big[p_t^p(X - Y)\big] \tag{C.32}$$

by the monotone convergence theorem. Combining (C.30) with (C.32), another application of monotone convergence yields the claimed bound (C.21).

(b) Let X and Y be two strong solutions of (C.7) (Definition C.15) and assume without loss of generality that $\mathbb{E}[|X_0 - Y_0|^p] < \infty$. The corresponding processes $\tilde{X}$ and $\tilde{Y}$ from part 131 satisfy $X \overset{\text{uti}}{=} \tilde{X}$ and $Y \overset{\text{uti}}{=} \tilde{Y}$, hence the function $\tilde{u}_N$ from (C.26) agrees with u_N from (C.31) for each $N \in \mathbb{N}$. Let μ denote the measure with density $\mathbb{R}_+ \ni s \mapsto c_s^p$ w.r.t. the Lebesgue–Borel measure on $\mathbb{R}_+$ and note that for each $N \in \mathbb{N}$, u_N is bounded by $\mathbb{E}[|X_0 - Y_0|^p] + N^p$ and thus locally integrable w.r.t. μ. The bound (C.30) shows that u_N satisfies the integral inequality needed for the application of the extended Grönwall inequality, see [53, Theorem 8.21], with $I = \mathbb{R}_+$, $\alpha(t) := 3^{p-1}\mathbb{E}[|X_0 - Y_0|^p]$ for all $t \in \mathbb{R}_+$ and the function β given by (C.22). Since α is increasing and both β

and the density c^p of μ are bounded on $[0,t]$ by $\beta(t) \vee c_t^p$ for each $t \in \mathbb{R}_+$, we have by [53, Remark 8.23 (c)]

$$u_N(t) \leq 6^{p-1}\mathbb{E}[|X_0 - Y_0|^p] \exp\big(t(\beta(t) \vee c_t^p)^2\big), \qquad t \in \mathbb{R}_+. \tag{C.33}$$

Combining (C.32) with (C.33), we obtain the claimed bound (C.23). □

Lemma C.32 (Bound on p-th moments of semimartingales given by certain stochastic integrals)**.** *Fix $p \geq 2$, a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, a $\mathbb{K}^n$-valued $\mathcal{F}_0$-measurable initial random vector ξ, an $\mathbb{R}^d$- or $\mathbb{K}^d$-valued standard $(\mathbb{F}, \mathbb{P})$-Brownian motion $B = (B_t)_{t\in\mathbb{R}_+}$, and consider $\mathbb{G} \otimes \mathbb{F}$-progressive coefficients b and σ as in (C.5).*

(a) Let X be a $\mathbb{K}^n$-valued continuous $\mathbb{F}$-adapted process satisfying the integrability condition (C.12). Then the $\mathbb{K}^n$-valued process

$$\tilde{X} := \xi + \int_0^{\cdot} b(s, X, \cdot\,)\,\mathrm{d}s + \int_0^{\cdot} \sigma(s, X, \cdot\,)\,\mathrm{d}B_s \tag{C.34}$$

is a well-defined continuous semimartingale, and for all $t \in \mathbb{R}_+$,

$$\begin{aligned}\mathbb{E}[p_t^p(\tilde{X})] \leq 3^{p-1}\bigg(\mathbb{E}[|\xi|^p] + t^{p-1}\int_0^t \mathbb{E}\left[|b(s,X,\cdot\,)|^p\right]\mathrm{d}s \\ + C_p c_{\mathbb{K}}^{p/2} t^{p/2-1}\int_0^t \mathbb{E}\left[|\sigma(s,X,\cdot\,)|^p\right]\mathrm{d}s\bigg),\end{aligned} \tag{C.35}$$

where C_p originates from the Burkholder–Davis–Gundy inequality [53, Theorem 6.32] and $c_{\mathbb{K}} := \dim_{\mathbb{R}}(\mathbb{K})$.

(b) Assume that b and σ are either

(i) bounded in the second argument (Definition C.26) or

(ii) of linearly bounded growth (Definition C.27).

Then for every $\mathbb{K}^n$-valued continuous $\mathbb{F}$-adapted process X, the process $\tilde{X}$ given by (C.34) is a well-defined $\mathbb{K}^n$-valued continuous semimartingale, and for all $t \in \mathbb{R}_+$,

$$\mathbb{E}\big[p_t^p(\tilde{X})\big] \leq 3^{p-1}\mathbb{E}[|\xi|^p] + \beta(t)\int_0^t D_s^p\,\mathbb{E}[(1 + I\,p_s^p(X))]\,\mathrm{d}s\,, \tag{C.36}$$

with β defined in (C.22), and where in case (i) the function D originates from Definition C.26 and $I := 0$, and in case (ii), D originates from Definition C.27 and $I := 1$.

Proof. The proof is based on ideas from Step 4 of the proof of [53, Theorem 8.27], with modifications to account for a nonzero starting value ξ and the explicit dependence of the coefficients on ω.

(a) By Lemma C.7 (a), the integrands in (C.34) are $\mathbb{F}$-progressive. By assumption, they also satisfy the integrability condition (C.12), hence Remark C.2 implies that $\tilde{X}$ is a continuous semimartingale. By the triangle inequality,

$$p_t(\tilde{X}) \leq |\xi| + p_t\left(\int_0^{\cdot} b(s,X,\cdot\,)\,\mathrm{d}s\right) + p_t\left(\int_0^{\cdot} \sigma(s,X,\cdot\,)\,\mathrm{d}B_s\right). \tag{C.37}$$

From Jensen's inequality for $\mathbb{R}_+ \ni x \mapsto x^p$ (see (C.18)),

$$p_t^p(\tilde{X}) \le 3^{p-1}\left[|\xi|^p + p_t^p\left(\int_0^{\cdot} b(s,X,\,\cdot\,)\,\mathrm{d}s\right) + p_t^p\left(\int_0^{\cdot} \sigma(s,X,\,\cdot\,)\,\mathrm{d}B_s\right)\right]. \tag{C.38}$$

By another application of Jensen's inequality (see (C.19)),

$$p_t^p\left(\int_0^{\cdot} b(s,X,\,\cdot\,)\,\mathrm{d}s\right) \le t^{p-1}\int_0^t |b(s,X,\,\cdot\,)|^p\,\mathrm{d}s\,. \tag{C.39}$$

To obtain a similar estimate for the third term on the right-hand side in (C.37), note that $\int_0^{\cdot}\sigma(s,X,\,\cdot\,)\,\mathrm{d}B_s$ is a continuous local martingale with quadratic variation process $\operatorname{tr}\bigl[\int_0^{\cdot}\sigma(s,X,\,\cdot\,)\,\mathrm{d}B_s\bigr] \overset{\text{uti}}{=} c_{\mathbb{K}}\int_0^{\cdot}|\sigma(s,X,\,\cdot\,)|^2\,\mathrm{d}s$. It follows from the Burkholder–Davis–Gundy inequality [53, Theorem 6.32] that

$$\begin{aligned}\mathbb{E}\left[p_t^p\Bigl(\int_0^{\cdot}\sigma(s,X,\,\cdot\,)\,\mathrm{d}B_s\Bigr)\right] &\le C_p c_{\mathbb{K}}^{p/2}\mathbb{E}\left[\Bigl(\int_0^t |\sigma(s,X,\,\cdot\,)|^2\,\mathrm{d}s\Bigr)^{p/2}\right]\\ &\le C_p c_{\mathbb{K}}^{p/2} t^{p/2-1}\mathbb{E}\left[\int_0^t |\sigma(s,X,\,\cdot\,)|^p\,\mathrm{d}s\right],\end{aligned} \tag{C.40}$$

where for the last inequality we used again Jensen's inequality (see (C.19)) with the function $\mathbb{R}_+ \ni x \mapsto x^{p/2}$, which is convex since $p \ge 2$ by assumption. Combining (C.38), (C.39), (C.40) and using Fubini's theorem yields (C.35).

(b) By Lemma C.7, the integrands in (C.34) are $\mathbb{F}$-progressive. Either the boundedness assumption (i) or continuity of sample paths of X together with assumption (ii) ensure that the stochastic integrals in (C.34) are well-defined in the sense of Definition C.1 (d). Hence by Remark C.2, $\tilde{X}$ is a $\mathbb{K}^n$-valued continuous semimartingale. From subadditivity of $\mathbb{R}_+ \ni x \mapsto x^{1/p}$ for the first, the assumed boundedness (Definition C.26) respectively linearly bounded growth (Definition C.27) of coefficients for the second and Jensen's inequality (see (C.18)) for the third inequality,

$$\begin{aligned}|b(s,X,\,\cdot\,)|^p + |\sigma(s,X,\,\cdot\,)|^p &\le \bigl(|b(s,X,\,\cdot\,)| + |\sigma(s,X,\,\cdot\,)|\bigr)^p\\ &\le D_s^p(1 + I\,p_s(X))^p\\ &\le 2^{p-1}D_s^p(1 + I\,p_s(X)^p)\end{aligned}$$

for all $s \in \mathbb{R}_+$. Combining with (C.35), using also the definition (C.22) of β, the claimed bound (C.36) follows. □

Lemma C.33. *In the setting of Lemma C.32, assume in addition that there is nonrandom dependence of coefficients on the solution, i.e. that b and σ are as in* (C.4)*. Let $\mathbb{H} = (\mathcal{H}_t)_{t\in\mathbb{R}_+}$ be the smallest filtration generated by ξ and B and containing all $\mathbb{P}$-null sets of $\sigma(\xi,(B_t)_{t\in\mathbb{R}_+})$. Then in both parts (a) and (b) of Lemma C.32, if X is $\mathbb{H}$-adapted, so is the process $\tilde{X}$ defined by* (C.34).

Proof. Lemma C.7 (b) and $\mathbb{H}$-progressive measurability of X imply that the integrands in (C.34) are $\mathbb{H}$-progressive. Since the random vector ξ is $\mathcal{H}_0$-measurable and B is also an $(\mathbb{H},\mathbb{P})$-Brownian motion, $\mathbb{H}$-progressive measurability of $\tilde{X}$ follows from Remark C.2 with the underlying filtration $\mathbb{F}$ replaced by $\mathbb{H}$. □

Theorem C.34 (Strong existence and pathwise uniqueness of solutions of classical SDEs, [53, Theorem 8.27])**.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic base (Definition C.5) and consider $\mathbb{G}\otimes\mathbb{F}$-progressive coefficients b and σ as in (C.5) which have the Lipschitz property (Definition C.25). Let $\mu \in \mathcal{P}(\mathbb{K}^n)$ be an initial probability distribution and let B denote an $\mathbb{R}^d$- or $\mathbb{K}^d$-valued standard $(\mathbb{F}, \mathbb{P})$-Brownian motion.*

(a) There is pathwise uniqueness of solutions (Definition C.22) of (C.7).

(b) For the initial distribution μ, there is strong existence of a solution of (C.7) in the sense of Definition C.19, if the functions b and σ are

(i) bounded in the second argument (Definition C.26) or

(ii) of linearly bounded growth (Definition C.27) and $\int_{\mathbb{K}^n} |x|^2 \, \mathrm{d}\mu(x) < \infty$.

For each power $p \in [2, \infty)$, which satisfies $\int_{\mathbb{K}^n} |x|^p \mu(\mathrm{d}x) < \infty$ in case (ii) and for each $t \in \mathbb{R}_+$, the seminorm p_t of the increment process $X - \xi$ is p-integrable. If b and σ are as in (C.4), i.e. in the case of nonrandom dependence, then in both cases above, for a given $\mathcal{F}_0$-measurable initial random vector $\xi\colon \Omega \to \mathbb{K}^n$ with distribution μ, the solution X is progressive w.r.t. the smallest filtration $\mathbb{H} = (\mathcal{H}_t)_{t\in\mathbb{R}_+}$ generated by ξ and B and containing all $\mathbb{P}$-null sets of $\sigma(\xi, (B_t)_{t\in\mathbb{R}_+})$.

Remark C.35 (Bound for the p-th moments of increments in Theorem C.34, [53, Remark 8.29])**.** Equation (C.51) obtained in Step 4 of the proof of Theorem C.34 shows that an $L^p(\mathbb{P})$-bound for the seminorm p_t of the increments is given by

$$\|p_t(X - \xi)\|_{L^p(\mathbb{P})} \le \sum_{k=1}^{\infty} \Big(\frac{(K_t t)^k}{k!} \Big)^{1/p}, \qquad t \in \mathbb{R}_+, \tag{C.41}$$

where the series converges by the ratio test and

$$K_t := \beta(t) \max\left\{ c_t^p, D_t^p \left(1 + I\mathbb{E}[|\xi|^p]\right) \right\}, \qquad t \in \mathbb{R}_+, \tag{C.42}$$

where the function β was defined in (C.22), the function c originates from Definition C.25 and where in case (i) the function D originates from Definition C.26 and $I := 0$, and in case (ii), D originates from Definition C.27 and $I := 1$.

Proof of Theorem C.34. This proof follows closely in structure and wording the proof of [53, Theorem 8.28], with modifications to account for the possibility of explicit dependence of the coefficients on ω.[10]

Step 1 (Pathwise uniqueness). Let ξ and $\tilde{\xi}$ be $\mathcal{F}_0$-measurable initial random vectors with law μ, $\tilde{B}$ another standard $(\mathbb{F}, \mathbb{P})$-Brownian motion, and X and $\tilde{X}$ strong solutions of (C.7) on the set-ups $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$ and $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \tilde{B}, \tilde{\xi})$, respectively. If $\tilde{\xi} \overset{\text{a.s.}}{=} \xi$ and $B \overset{\text{uti}}{=} B$, then by Remark C.18 (a), $\tilde{X}$ is also a strong solution of (C.7) on the set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$. From Lemma C.31 (b), it follows that $\mathbb{E}[p_t^2(X - \tilde{X})] = 0$ for all $t \in \mathbb{R}_+$. Hence $\tilde{X} \overset{\text{uti}}{=} X$, proving part (a).

[10] Furthermore, some results established in the proof of [53, Theorem 8.28] have been slightly generalized and moved into Lemmas C.31 and C.32 for later re-use.

Step 2 (Definition of approximative solutions). To prove part (b), we consider a stochastic extension of the Picard iteration, which is the standard tool to prove the Picard–Lindelöf theorem concerning the existence and uniqueness of solutions of initial value problems in the theory of ordinary differential equations.

Define the pathwise constant process $X^{(0)}$ by $X_t^{(0)} = \xi$ for all $t \in \mathbb{R}_+$, which is a continuous $\mathcal{H}_0$-measurable process and a local martingale by [53, Exercise 4.143(a)]. Using Lemma C.32 (b), it follows inductively that for every $k \in \mathbb{N}_0$, the iterated process

$$X^{(k+1)} := \xi + \int_0^{\cdot} b\big(s, X^{(k)}, \cdot\,\big)\,\mathrm{d}s + \int_0^{\cdot} \sigma\big(s, X^{(k)}, \cdot\,\big)\,\mathrm{d}B_s \tag{C.43}$$

is a well-defined $\mathbb{K}^n$-valued $\mathbb{F}$-progressive continuous semimartingale, which by Lemma C.33 is $\mathbb{H}$-progressive in the case of nonrandom dependence of the coefficients.

Step 3 (The increments $(X^{(k)} - \xi)_{k\in\mathbb{N}}$ form a Cauchy sequence). We claim that, with K_t given by (C.42),

$$\mathbb{E}\Big[p_t^p\big(X^{(k+1)} - X^{(k)}\big)\Big] \le \frac{(K_t t)^{k+1}}{(k+1)!}, \qquad k \in \mathbb{N}_0,\ t \in \mathbb{R}_+. \tag{C.44}$$

Since

$$X^{(1)} - X^{(0)} = \int_0^{\cdot} b(s, X^{(0)}, \cdot\,)\,\mathrm{d}s + \int_0^{\cdot} \sigma(s, X^{(0)}, \cdot\,)\,\mathrm{d}B_s\,, \tag{C.45}$$

applying (C.36) from Lemma C.32 (b) to the right-hand side of (C.45), we have for each $t \in \mathbb{R}_+$

$$\begin{aligned}\mathbb{E}[p_t^p(X^{(1)} - X^{(0)})] &\le \beta(t) \int_0^t D_s^p\,\mathbb{E}\big[(1 + I\,|\xi|^p)\big]\,\mathrm{d}s\\ &\le t\,\beta(t)\,D_t^p\,\mathbb{E}\big[(1 + I\,|\xi|^p)\big] \le K_t\,t,\end{aligned}$$

with K_t given by (C.42). This establishes (C.44) for $k = 0$. We proceed inductively. Fix $k \in \mathbb{N}$ and $t \in \mathbb{R}_+$. From (C.21) in Lemma C.31 (a),

$$\mathbb{E}[p_t^p(X^{(k+1)} - X^{(k)})] \le \beta(t) \int_0^t c_s^p\,\mathbb{E}[p_s^p(X^{(k)} - X^{(k-1)})]\,\mathrm{d}s\,.$$

Since $\mathbb{R}_+ \ni s \mapsto c_s$ is increasing by Definition C.25 of the Lipschitz property, it follows using the definition (C.42) of K_t and the induction hypothesis (C.44) for $k-1$ that

$$\mathbb{E}\Big[p_t^p\big(X^{(k+1)} - X^{(k)}\big)\Big] \le K_t \int_0^t \frac{(K_s s)^k}{k!}\,\mathrm{d}s\,. \tag{C.46}$$

The estimate $K_s \le K_t$ and integration yield the right-hand side of (C.44). For all $j \le l$ in $\mathbb{N}_0$ and $t \in \mathbb{R}_+$, using a telescoping sum, the triangle inequality for p_t and $\|\cdot\|_{L^p(\mathbb{P})}$, and finally (C.44) with an index shift,

$$\begin{aligned}\big\|p_t(X^{(l+1)} - X^{(j)})\big\|_{L^p(\mathbb{P})} &= \bigg\|p_t\bigg(\sum_{k=j}^{l}\big(X^{(k+1)} - X^{(k)}\big)\bigg)\bigg\|_{L^p(\mathbb{P})}\\ &\le \sum_{k=j}^{l}\big\|p_t(X^{(k+1)} - X^{(k)})\big\|_{L^p(\mathbb{P})} \le \sum_{k=j+1}^{\infty}\bigg(\frac{(K_t t)^k}{k!}\bigg)^{1/p}.\end{aligned} \tag{C.47}$$

Since the series on the right-hand side converges by the ratio test, it converges to zero as $j \to \infty$. Hence, restricted to every interval $[0,t]$ with $t \in \mathbb{R}_+$, the processes $\big(X^{(k)} - \xi\big)_{k\in\mathbb{N}}$ form a Cauchy sequence in $L^p\left(\Omega, \mathcal{F}, \mathbb{P}; \left(C\left([0,t], \mathbb{K}^n\right), p_t\right)\right)$. Moreover, in the case of nonrandom dependence, $\mathcal{H}$ can replace $\mathcal{F}$ in the previous sentence.

Step 4 (There exists a limiting continuous progressive process X such that $p_t(X - \xi)$ is p-integrable for each $t \in \mathbb{R}_+$). Let $(a_m)_{m\in\mathbb{N}}$ denote a decreasing sequence in $(0,\infty)$ satisfying $\sum_{m\in\mathbb{N}} a_m < \infty$. Define $j_0 = 0$. For each $m \in \mathbb{N}$ there exists $j_m \in \mathbb{N}$ with $j_m > j_{m-1}$ such that

$$\sum_{k=j_m+1}^{\infty} \left(\frac{(K_m m)^k}{k!}\right)^{1/p} \le a_m^2. \tag{C.48}$$

Define $A_m = \big\{p_m\big(X^{(j_{m+1})} - X^{(j_m)}\big) \ge a_m\big\}$, which is in $\mathcal{F}_m$, and in fact in $\mathcal{H}_m$ in the case of nonrandom dependence. By Jensen's inequality, $\|\cdot\|_{L^1(\mathbb{P})} \le \|\cdot\|_{L^p(\mathbb{P})}$, hence Markov's inequality, (C.47) and (C.48) imply that

$$\mathbb{P}[A_m] \le \frac{1}{a_m}\big\|p_m(X^{(j_{m+1})} - X^{(j_m)})\big\|_{L^p(\mathbb{P})} \le a_m.$$

The set $N := \{A_m \text{ i.o.}\} = \bigcap_{m\in\mathbb{N}} \bigcup_{n\ge m} A_n$ is in $\mathcal{F}_\infty$, and in fact in $\mathcal{H}_\infty$ in the case of nonrandom dependence. Moreover, by the Borel–Cantelli lemma [53, Lemma 17.10(a)], $\mathbb{P}[N] = 0$. Hence for every $\omega \in \Omega$ the functions[11]

$$\tilde{X}^{(j_m)}(\omega) := \xi(\omega) + \Big(X^{(j_m)}(\omega) - \xi(\omega)\Big)\, \mathbb{1}_{N^c}(\omega), \qquad m \in \mathbb{N}_0, \tag{C.49}$$

form a Cauchy sequence in W_n with respect to the seminorm p_t for every $t \in \mathbb{R}_+$. Therefore, they form a Cauchy sequence with respect to the metric ϱ_{ucc} for the topology of compact convergence in W_n given in (2.2), see [53, Exercise 2.121 (c)] for details. By Lemma 2.1(b), the metric space $\left(C\left(\mathbb{R}_+, \mathbb{K}^n\right), \varrho_{\text{ucc}}\right)$ is complete, hence there exists a function $X\colon \Omega \to W_n$ with

$$\tilde{X}^{(j_m)}(\omega) \xrightarrow[m\to\infty]{\text{ucc}} X(\omega), \qquad \omega \in \Omega. \tag{C.50}$$

By assumption, the filtrations $\mathbb{F}$ and $\mathbb{H}$ satisfy the usual completeness condition (Definition C.3). Therefore, the $\mathbb{P}$-null set N is in $\mathcal{F}_0$ respectively in $\mathcal{H}_0$ in the case of nonrandom dependence. Hence for each $m \in \mathbb{N}$ the process $\tilde{X}^{(j_m)}$ is still $\mathbb{F}$- respectively $\mathbb{H}$-progressive and therefore measurable w.r.t. the progressive σ-algebra $\Sigma_{\text{pm}}\left(\mathbb{R}_+, \mathbb{F}\right)$, and in fact w.r.t. $\Sigma_{\text{pm}}\left(\mathbb{R}_+, \mathbb{H}\right)$, in the case of nonrandom dependence, see [53, Exercise 3.36]. Since pointwise convergence transfers the measurability to the limit, see [53, Lemma 15.30], it follows that the limiting process X is also $\mathbb{F}$- respectively $\mathbb{H}$-progressive.

To verify that $X - \xi$ is a p-integrable process satisfying (C.41), note that by Fatou's lemma, the fact that $X^{(0)} \equiv \xi$ and (C.47), for each $t \in \mathbb{R}_+$,

$$\|p_t(X - \xi)\|_{L^p(\mathbb{P})} \le \liminf_{m\to\infty} \|p_t\big(\tilde{X}^{(j_m)} - X^{(0)}\big)\|_{L^p(\mathbb{P})} \le \sum_{k=1}^{\infty}\left(\frac{(K_t t)^k}{k!}\right)^{1/p}. \tag{C.51}$$

[11] This definition of $\tilde{X}^{(j_m)}$ ensures that for the constructed limiting process X, there is no exceptional $\mathbb{P}$-null set in Definition C.15(a).

Step 5 (X is a continuous semimartingale solving the integral equation (C.13)). Using Lemma C.32 (b) for the limiting process X, it follows that the right-hand side of (C.13),

$$\tilde{X} := X_0 + \int_0^{\cdot} b(s, X, \cdot\,)\,\mathrm{d}s + \int_0^{\cdot} \sigma(s, X, \cdot\,)\,\mathrm{d}B_s\,, \tag{C.52}$$

is a well-defined continuous semimartingale, and it remains to show that $X \overset{\text{uti}}{=} \tilde{X}$. By the Lipschitz property (Definition C.25) and (C.50), for all $s \le t$ in $\mathbb{R}_+$,

$$\begin{aligned} &|b(s, X, \cdot\,) - b(s, \tilde{X}^{(j_m)}, \cdot\,)| + |\sigma(s, X, \cdot\,) - \sigma(s, \tilde{X}^{(j_m)}, \cdot\,)| \\ &\qquad \le c_t p_t(X - \tilde{X}^{(j_m)}) \xrightarrow[m\to\infty]{} 0 \quad \text{pointwise for } (t, \omega) \in \mathbb{R}_+ \times \Omega. \end{aligned} \tag{C.53}$$

Therefore, for every $t \in \mathbb{R}_+$,

$$p_t\biggl(\int_0^{\cdot} b(s, X, \cdot\,)\,\mathrm{d}s - \int_0^{\cdot} b(s, \tilde{X}^{(j_m)}, \cdot\,)\,\mathrm{d}s\biggr) \le t\, c_t\, p_t(X - \tilde{X}^{(j_m)}) \xrightarrow[m\to\infty]{} 0 \tag{C.54}$$

pointwise on Ω. Furthermore, it follows from [53, Exercise 5.116(c)] and (C.53) that

$$\begin{aligned} &\operatorname{tr}\biggl(\biggl[\int_0^{\cdot} (\sigma(s, X, \cdot\,) - \sigma(s, \tilde{X}^{(j_m)}, \cdot\,))\,\mathrm{d}B_s\biggr]\biggr) \\ &\qquad\qquad \overset{\text{uti}}{=} c_{\mathbb{K}} \int_0^{\cdot} |\sigma(s, X, \cdot\,) - \sigma(s, \tilde{X}^{(j_m)}, \cdot\,)|^2\,\mathrm{d}s \xrightarrow{\text{ucp}} 0 \quad \text{as } m \to \infty, \end{aligned}$$

where ucp denotes convergence uniformly on compact subsets in probability, see [53, Definition 5.125].[12] By [53, Corollary 5.124], see also [53, Exercise 5.127], for every $t \in \mathbb{R}_+$,

$$p_t\biggl(\int_0^{\cdot} \sigma(s, X, \cdot\,)\,\mathrm{d}B_s - \int_0^{\cdot} \sigma(s, \tilde{X}^{(j_m)}, \cdot\,)\,\mathrm{d}B_s\biggr) \xrightarrow[m\to\infty]{\mathbb{P}} 0 \tag{C.55}$$

For each $m \in \mathbb{N}_0$, using (C.43) with $k = j_m$, (C.49), and invariance up to indistinguishability of stochastic integrals under passage to an indistinguishable integrand, see also [53, Exercise 5.113],

$$X^{(j_m+1)} \overset{\text{uti}}{=} \xi + \int_0^{\cdot} b(s, \tilde{X}^{(j_m)}, \cdot\,)\,\mathrm{d}s + \int_0^{\cdot} \sigma(s, \tilde{X}^{(j_m)}, \cdot\,)\,\mathrm{d}B_s\,.$$

Fix $t \in \mathbb{R}_+$ for a while. Using the definition (C.52) of $\tilde{X}$, the fact that $X_0 = \xi$ by (C.49), (C.50), the previous display, (C.54) and (C.55), it follows that

$$p_t\bigl(X^{(j_m+1)} - \tilde{X}\bigr) \xrightarrow[m\to\infty]{\mathbb{P}} 0. \tag{C.56}$$

For each $m \in \mathbb{N}_0$, using (C.44) with $k = j_m$, and the fact that $\tilde{X}^{(j_m)} \overset{\text{uti}}{=} X^{(j_m)}$ by (C.49),

$$p_t(X^{(j_m+1)} - \tilde{X}^{(j_m)}) \xrightarrow[m\to\infty]{\mathbb{P}} 0. \tag{C.57}$$

[12] For stochastic processes $H, H^1, H^2, \ldots : \mathbb{R}_+ \times \Omega \to \mathbb{K}^n$ which are all right-continuous (or left-continuous), ucp-convergence is is equivalent to $p_t(H^n - H) \xrightarrow[n\to\infty]{\mathbb{P}} 0$ for all $t \in \mathbb{R}_+$, where the seminorms p_t have been defined in (2.1).

Combining (C.50), (C.57) and (C.56), we obtain using the triangle inequality for the seminorm p_t

$$p_t\big(X-\tilde{X}\big) \le p_t(X-\tilde{X}^{(j_m)}) + p_t(\tilde{X}^{(j_m)} - X^{(j_m+1)}) + p_t(X^{(j_m+1)} - \tilde{X}) \xrightarrow[m\to\infty]{\mathbb{P}} 0.$$

Therefore, $p_t(X-\tilde{X}) \overset{\text{a.s.}}{=} 0$. Since $t \in \mathbb{R}_+$ was arbitrary, this implies $X \overset{\text{uti}}{=} \tilde{X}$. From the definition (C.52) of $\tilde{X}$, it follows that X satisfies (C.13) and is therefore a strong solution of (C.5) with initial condition ξ according to Definition C.15. In particular, by Remark C.17, X is a continuous semimartingale. □

C.3. SDEs Driven by Correlated Brownian Motions

This section shows how to pass from an SDE driven by a multidimensional (not necessarily standard) Brownian motion with a suitably specified covariation process to an SDE w.r.t. a multidimensional standard Brownian motion. This addresses a gap in our development of the theory of stochastic differential equations up to this point, where we have treated only the case of standard Brownian motions as integrators. This is a limitation, since mathematical models of continuous time financial markets often involve multiple stochastic processes of interest, e.g. a risk-free interest rate, a stochastic volatility and a number of risky assets (see e.g. Definition 3.1 of the local stochastic volatility model), each of which can be described by its own SDE, with the Brownian motions driving these SDEs often specified to be correlated. In more precise terms, given a stochastic base (Definition C.5 (a)) $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, the market evolution is described by an $\mathbb{R}^n$-valued process $X = (X_t)_{t\in\mathbb{R}_+}$ whose components $X^1, \dots, X^n$ satisfy the SDEs

$$\mathrm{d}X^i_t = b^i(t, X, \cdot\,)\,\mathrm{d}t + \sigma^i(t, X, \cdot\,)\,\mathrm{d}B^i_t, \qquad i \in \{1, \dots, n\}, \tag{C.58}$$

where the coefficients $b^1, \dots, b^n$ and $\sigma^1, \dots, \sigma^n$ are $\mathbb{R}$-valued $\mathbb{G}\otimes\mathbb{F}$-progressive (Definition C.6) processes, each component of $B = (B^1, \dots, B^n)$ is an $\mathbb{R}$-valued standard Brownian motion and the covariation process (see [53, Definition 5.67]) $[B]$ is prescribed.

The advantage of such a specification is that the individual quantities $X^1, \dots, X^n$ of interest follow one-dimensional SDEs, which may be easier to understand and for which existence and uniqueness results with relaxed conditions compared to those in Theorem C.34 are available.[13] A disadvantage of this type of specification in the setting developed in Appendices C.1 and C.2 is that our existence and uniqueness result (Theorem C.34) explicitly assumes the SDE to be driven by a standard Brownian motion.

In the simplest case, which commonly appears in the literature, the specification of $[B]$ is of the form

$$[B]_t = t\rho, \qquad t \in \mathbb{R}_+, \tag{C.59}$$

with a matrix $\rho \in \mathbb{R}^{n\times n}$ which can be viewed as an instantaneous correlation matrix. For generality, we consider (C.58) also for more general prescriptions of $[B]$ than (C.59) and include the case of $\mathbb{C}$-valued coefficients and Brownian motions. The situation of (C.59) is taken up again (and treated elementarily) in Example C.39 below.

We fix for later reference the following assumption.

[13] See e.g. [32, §6.3.1] for a result of this type.

Assumption C.36. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic base (Definition C.5(a)) carrying a $\mathbb{K}^n$-valued $\mathcal{F}_0$-measurable initial random vector ξ as well as a $\mathbb{K}^n$-valued stochastic process $B = (B_t)_{t\in\mathbb{R}_+}$ whose components are $\mathbb{K}$-valued $(\mathbb{F}, \mathbb{P})$-standard Brownian motions. Let $b^1, \dots, b^n$ and $\sigma^1, \dots, \sigma^n$ be $\mathbb{K}$-valued $\mathbb{G} \otimes \mathbb{F}$-progressive (Definition C.6) coefficients and define $b = (b_1, \dots, b_n)$ and $\sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_n)$.

The definition of a strong solution of (C.58) is analogous to the classical case (Definition C.15), with the exception that we relax the requirement that the integrator be a standard Brownian motion.

Definition C.37 (Strong solution of solutions the SDE (C.58))**.** Consider the setting of Assumption C.36. A continuous $\mathbb{K}^n$-valued $\mathbb{F}$-adapted stochastic process X is called a *strong solution* of the system of SDEs (C.58) with initial condition $X_0 \overset{\text{a.s.}}{=} \xi$, if, in this context, parts (a) to (c) from Definition C.15 are satisfied.

Lemma C.41 below allows to write (C.58) as an SDE driven by a standard Brownian motion, provided the covariation process of the original vector of driving Brownian motions is of a suitable form. This is made possible by the following theorem.

Theorem C.38 (Representation of continuous local martingales, [53, Theorem 8.33])**.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic base, $M = (M_t)_{t\in\mathbb{R}_+}$ a $\mathbb{K}^n$-valued continuous local martingale, $V\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^{n\times d}$ a progressive process, and define $c_{\mathbb{K}} = \dim_{\mathbb{R}}(\mathbb{K})$. Assume that*

$$[M] \overset{\text{uti}}{=} c_{\mathbb{K}} \int_0^{\cdot} V_s V_s^{\mathsf{H}}\, \mathrm{d}s\,, \tag{C.60}$$

and, if $\mathbb{K} = \mathbb{C}$, in addition that $[M, M] \overset{\text{uti}}{=} 0$. Then, after a suitable extension of the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ if necessary, there exists a $\mathbb{K}^d$-valued standard $(\mathbb{F}, \mathbb{P})$-Brownian motion W with $V \in L_n(W)$ (see Definition C.1(d)) and

$$M \overset{\text{uti}}{=} M_0 + \int_0^{\cdot} V_s\, \mathrm{d}W_s\,. \tag{C.61}$$

The next example follows partly arguments from the proof of [53, Theorem 8.33].

Example C.39. In this example, we show that the use of Theorem C.38 in Lemma C.41 below can be omitted under some assumptions. Fix a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ (Definition C.5(a)) and let $B = (B_t)_{t\in\mathbb{R}_+}$ be an $\mathbb{R}^n$-valued stochastic process whose components $B^1, \dots, B^n$ are $\mathbb{R}$-valued $(\mathbb{F}, \mathbb{P})$-standard Brownian motions.

(a) Assume that

$$[B] \overset{\text{uti}}{=} \int_0^{\cdot} \rho_s\, \mathrm{d}s\,, \tag{C.62}$$

where $\rho\colon \mathbb{R}_+ \times \Omega \to \mathbb{R}^{n\times n}$ is progressive and takes values in the set of positive definite matrices containing all ones on the diagonal,[14] is pathwise locally integrable w.r.t. the

[14] Since the diagonal of the covariation process $[B]$ contains the quadratic variations $[B^i]$, $i \in \{1, \dots, n\}$ of the component Brownian motions, we necessarily have $\rho_s^{ii} = 1$ for λ-a.e. $s \in \mathbb{R}_+$ and all $i \in \{1, \dots, n\}$ due to Lévy's characterization of Brownian motion ([53, Corollary 7.2]). By [53, Theorem 7.1], the fact that B is a continuous local martingale with deterministic covariation process given by (C.62) implies by itself that B has independent, Gaussian increments with $\mathcal{L}[B_t - B_s] = \mathcal{N}(0, \int_s^t \rho_u\, \mathrm{d}u)$ for all $s < t$ in $\mathbb{R}_+$.

Lebesgue measure and furthermore uniformly positive definite, by which we mean that there exists an $\alpha \in (0,\infty)$ such that[15]

$$\langle x, \rho_s(\omega)x\rangle \geq \alpha |x|^2, \qquad (s,\omega) \in \mathbb{R}_+ \times \Omega,\ x \in \mathbb{R}^n. \tag{C.63}$$

For all $(s,\omega) \in \mathbb{R}_+ \times \Omega$, let $V_s(\omega)$ be the Cholesky factorization of ρ_s, i.e. $V_s(\omega)$ is lower triangular with positive diagonal and $\rho_s = V_s V_s^{\mathsf{T}}$. Note that (C.63) is equivalent to $|V_s x| \geq \sqrt{\alpha}|x|$ for all $s \in \mathbb{R}_+$ and $x \in \mathbb{R}^n$, hence V_s is invertible and[16] $\|V_s^{-1}\| \leq 1/\sqrt{\alpha}$ for all $s \in \mathbb{R}_+$, where $\|\cdot\|$ denotes the matrix 2-norm. Note also that V is a progressive process by Footnote 17 on Page 139. We next verify that $W := V^{-1} \cdot B$ is an $\mathbb{R}^n$-valued $(\mathbb{F},\mathbb{P})$-standard Brownian motion with $V \in L_n(W)$ satisfying also (C.61) for B instead of M. This shows in particular that no extension of the underlying filtered probability space is needed in this simple case. Boundedness of $\|V^{-1}\|$ ensures that $V^{-1} \in L_n(B)$, so $V^{-1} \cdot B$ is a well-defined $\mathbb{R}^n$-valued continuous local martingale and has, by [53, Exercise 6.3 (d)], the covariation process

$$[V^{-1} \cdot B] \overset{\text{uti}}{=} \int_0^{\cdot} V_s^{-1}\, \mathrm{d}[B,B]_s\, (V_s^{-1})^{\mathsf{T}} \overset{\text{(C.62)}}{=} \int_0^{\cdot} V_s^{-1} V_s V_s^{\mathsf{T}} (V_s^{-1})^{\mathsf{T}}\, \mathrm{d}s = \int_0^{\cdot} I_n\, \mathrm{d}s\,,$$

where I_n denotes the $(n \times n)$-identity matrix. By Lévy's characterization of Brownian motion ([53, Corollary 7.2]), $V^{-1} \cdot B$ is an $\mathbb{R}^n$-valued standard Brownian motion. Since $\rho_s = V_s V_s^{\mathsf{T}}$,

$$\|\rho_s\| = \|V_s\|^2, \qquad s \in \mathbb{R}_+, \tag{C.64}$$

and thus pathwise local square integrability of V follows from the assumed local integrability of ρ. Hence $V^{ik}(V^{-1})^{kj}$ is pathwise locally square integrable for all i, j and k in $\{1,\dots,n\}$, and $V \cdot W = V \cdot (V^{-1} \cdot B) = B$ follows by the chain rule for stochastic integrals ([53, Theorem 5.117 (b)]).

(b) Assume that the process ρ from part (a) is pathwise locally bounded, which by (C.62) is equivalent to assuming pathwise local boundedness of V. Consider coefficients b and σ as in Assumption C.36. From Lemma C.41 below, the strong solutions of the SDEs (C.58) and (C.67) coincide.

Remark C.40. (a) Theorem C.38 is generalized in [54, Theorem 4.19] where a representation similar to (C.61) (with a Gaussian martingale instead of Brownian motion as the integrator) is established for every continuous local martingale with a covariation process that is absolutely continuous w.r.t. a given continuous increasing function.

(b) In Theorem C.38, it is implicitly assumed (for well-definedness of (C.60)) that $\mathbb{R}_+ \ni s \mapsto V_s V_s^{\mathsf{H}}$ is pathwise locally integrable w.r.t. the Lebesgue measure. Thus, from (C.60)

[15] By diagonalization of ρ, the condition (C.63) is equivalent to the smallest eigenvalue of $\rho_s(\omega)$ being, for all $(s,\omega) \in \mathbb{R}_+ \times \Omega$, no less than α.

[16] Let $A \in \mathbb{R}^{n\times n}$ be a matrix such that $|Ax| \geq \sqrt{\alpha}|x|$ for all $x \in \mathbb{R}^n$. Clearly this implies that the kernel of A is equal to the trivial subspace $\{0\}$, hence A is nonsingular. Consequently, A^{-1} exists by the rank-nullity theorem. Fix $y \in \mathbb{R}^n$ and let $x \in \mathbb{R}^n$ be such that $y = Ax$. Then $|A^{-1}y| = |A^{-1}Ax| = |x| \leq \frac{1}{\sqrt{\alpha}}|Ax| = |y|$. Since this holds for all $y \in \mathbb{R}^n$, $\|A^{-1}\| \leq \frac{1}{\sqrt{\alpha}}$.

and the Lebesgue differentiation theorem (see [21, Theorem 3.21]), we have that for $\mathbb{P}$-a.e. $\omega \in \Omega$,

$$V(\omega)V^{\mathsf{H}}(\omega) = c_{\mathbb{K}}^{-1}\frac{\partial}{\partial t}[M](\omega), \qquad \lambda\text{-a.e.} \tag{C.65}$$

Equation (C.65) suggests that given a continuous local martingale M as in Theorem C.38 with a covariation process $[M]$ admitting a representation as in (C.61) (but not knowing the particular process V), we can in principle recover VV^{H}, at least $\lambda \otimes \mathbb{P}$-a.e. This is indeed possible, though care has to be taken to handle the exceptional null sets in (C.65), for which we refer to Step 1 of the proof of [54, Theorem 4.19]. Given the process VV^{H}, one can perform, pointwise on $\mathbb{R}_+ \times \Omega$ a Cholesky factorization (see [30, Corollary 7.2.9]) to obtain[17] a progressive process $\tilde{V}$ with $\tilde{V}\tilde{V}^{\mathsf{H}} = VV^{\mathsf{H}}$, and (C.61) remains valid for $\tilde{V}$ instead of V.

Lemma C.41 (Writing (C.58) as an SDE w.r.t. a multidimensional standard Brownian motion)**.** *Consider the setting of Assumption C.36 and assume that there exists a progressive process $V \colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^{n\times d}$ such that (C.60) is satisfied for B in place of M. In addition, if $\mathbb{K} = \mathbb{C}$, assume that $[B, B] \overset{\text{uti}}{=} 0$. Let $X = (X_t)_{t\in\mathbb{R}_+}$ be a $\mathbb{K}^n$-valued stochastic process and denote by W the $\mathbb{K}^d$-valued standard $(\mathbb{F}, \mathbb{P})$-Brownian motion defined on a possibly enlarged probability space from Theorem C.38.*

(a) If X is a strong solution of (C.58) in the sense of Definition C.37 and the process $(t, \omega) \mapsto \sigma(t, X(\omega), \omega)V_t(\omega)$ is pathwise locally square integrable, i.e.

$$\int_0^t |\sigma(s, X, \cdot\,)V_s|^2 \,\mathrm{d}s < \infty, \qquad t \in \mathbb{R}_+, \tag{C.66}$$

then X is a strong solution of the SDE

$$\mathrm{d}X_t = b(t, X, \cdot\,)\,\mathrm{d}t + \sigma(t, X, \cdot\,)V_t\,\mathrm{d}W_t \tag{C.67}$$

in the sense of classical SDEs (Definition C.15).

(b) If X is a strong solution of (C.67) in the sense of classical SDEs (Definition C.15), then X is a strong solution of (C.58) in the sense of Definition C.37.

In particular, if V is pathwise locally bounded, then the strong solutions of (C.58) and (C.67) coincide.

Proof. The proof is essentially an application of the chain rule for stochastic integrals, made slightly tedious only by the fact that questions of stochastic integrability[18] need to be

[17] The Cholesky decomposition $C = LL^{\mathsf{H}}$ of a positive definite matrix C into a product of a lower triangular factor L with its hermitian conjugate L^{H} is unique, but this is not the case when C is merely positive semidefinite. In any case, a decomposition can be computed algorithmically by applying a finite sequence of measurable maps, which (i) yields a concrete choice of the decomposition and (ii) shows that the matrix L is a measurable function of C. Consequently, if C denotes instead a stochastic process taking values in the set of positive semidefinite Hermitian matrices, then the pointwise Cholesky factorization of C is a stochastic process inheriting the measurability properties of C.

[18] See [53, Example 5.119] for failure of the chain rule when a certain integrability condition is not met.

addressed.[19]

Step 1 (Preparation). Recall that by Theorem C.38,

$$V \in L_n(W) \quad \text{and} \quad B \overset{\text{uti}}{=} \int_0^{\cdot} V_s \,\mathrm{d}W_s\,. \tag{C.68}$$

Since σ is a diagonal matrix, (C.66) is equivalent to

$$\int_0^t |\sigma^i(s,X,\,\cdot\,)V_s^{ij}|^2 \,\mathrm{d}s < \infty, \qquad t \in \mathbb{R}_+,\ (i,j) \in \{1,\dots,n\} \times \{1,\dots,d\}. \tag{C.69}$$

Moreover, if (C.69) holds, then by (C.68) and the chain rule for stochastic integrals ([53, Theorem 5.117]), the process $(s,\omega) \mapsto \sigma(s,X,\,\cdot\,)$ is in $L_n(B)$ and

$$\int_0^{\cdot} \sigma(s,X,\,\cdot\,)\,\mathrm{d}B_s \overset{\text{uti}}{=} \int_0^{\cdot} \sigma(s,X,\,\cdot\,)\,\mathrm{d}(V\cdot W)_s \overset{\text{uti}}{=} \int_0^{\cdot} \underbrace{\sigma(s,X,\,\cdot\,)V_s}_{=\hat\sigma(s,X,\,\cdot\,)}\mathrm{d}W_s\,, \tag{C.70}$$

where $\hat\sigma$ is defined in (C.71) below.

Step 2 (Proof of part (a)). The SDE (C.67) can be written as a classical SDE in the usual form (C.7) with $\mathbb{G}\otimes\mathbb{F}$-progressive coefficients b and $\hat\sigma$, where

$$\hat\sigma(t,w,\omega) := \sigma(t,w,\omega)V_t(\omega), \qquad (t,w,\omega) \in \mathbb{R}_+ \times W_n \times \Omega. \tag{C.71}$$

We verify that X satisfies, w.r.t. this classical SDE, parts (a) to (c) from Definition C.15. From the Definition C.37 of a strong solution of (C.58), it is clear that $X_0 \overset{\text{a.s.}}{=} \xi$ and that

$$\int_0^t |b(s,X,\,\cdot\,)|\,\mathrm{d}s < \infty, \qquad t \in \mathbb{R}_+. \tag{C.72}$$

By (C.66), we also have

$$\int_0^t |\hat\sigma(s,X,\,\cdot\,)|^2\,\mathrm{d}s < \infty, \qquad t \in \mathbb{R}_+, \tag{C.73}$$

hence (C.12) holds for $\hat\sigma$ instead of σ. Consequently, Definition C.15 (a) and (b) are satisfied, and it remains only to check that

$$X \overset{\text{uti}}{=} X_0 + \int_0^{\cdot} b(s,X,\,\cdot\,)\,\mathrm{d}s + \int_0^{\cdot} \hat\sigma(s,X,\,\cdot\,)\,\mathrm{d}W_s\,. \tag{C.74}$$

[19] Another subtlety in Lemma C.41 comes from the extension of the underlying probability space. The details of this extension procedure, which we do not recapitulate here, are given in [53, Notation 8.32]. The central observation is that the original filtered probability space, which, in order to to distinguish it from its extension $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$, we denote in this footnote by $(\tilde\Omega,\tilde{\mathcal{F}},\tilde{\mathbb{F}},\tilde{\mathbb{P}})$, is extended by forming the product with another filtered probability space, $(\Omega',\mathcal{F}',\mathbb{F}',\mathbb{P}')$. That is, the extension is given by $\Omega = \tilde\Omega\times\Omega'$, $\mathcal{F} = \tilde{\mathcal{F}}\otimes\mathcal{F}'$, $\mathbb{F} = \tilde{\mathbb{F}}\otimes\mathbb{F}' := (\tilde{\mathcal{F}}_t\otimes\mathcal{F}'_t)_{t\in\mathbb{R}_+}$ and $\mathbb{P} = \tilde{\mathbb{P}}\otimes\mathbb{P}'$, possibly adding null sets to ensure the usual completeness assumption from Definition C.3. Every stochastic process (or random variable) $\tilde Y$ on $(\tilde\Omega,\tilde{\mathcal{F}},\tilde{\mathbb{F}})$ can then be lifted to the extension by defining $Y := \tilde Y\circ\tilde\pi$, where $\tilde\pi\colon \Omega\to\tilde\Omega$, $\omega = (\tilde\omega,\omega')\mapsto\tilde\omega$ denotes the projection onto the first factor, and it is not hard to check that this conserves various process properties such as martingality and independent increments as well as the law of the process.

By assumption,

$$X \overset{\text{uti}}{=} X_0 + \int_0^\cdot b(s, X, \cdot\,)\,\mathrm{d}s + \int_0^\cdot \sigma(s, X, \cdot\,)\,\mathrm{d}B_s\,, \tag{C.75}$$

which using the definition (C.71) of $\hat{\sigma}$ together with (C.70) from Step 1 (which holds since (C.73) ensures (C.69)) implies (C.74).

Step 3 (Proof of part (b)). Assume that X is a strong solution of (C.67). We verify that X satisfies parts (a) to (c) from Definition C.37. From Definition C.15, we obtain the equalities $X_0 \overset{\text{a.s.}}{=} \xi$, (C.66) and (C.72). We have seen in Step 1 that (C.66) implies (C.69). The equations (C.68), (C.69) and the chain rule for stochastic integrals imply that the process $(s,\omega) \mapsto \sigma(s, X(\omega), \omega)$ is in $L_n(V \cdot W) = L_n(B)$, which is equivalent to (C.73) holding for σ instead of $\hat{\sigma}$. This establishes Definition C.37 (a) and (b). It remains only to check part (c), i.e. that (C.75) holds. We have (C.74) by assumption, have already established (C.69) and shown in Step 1 that this implies (C.70). Combining (C.70) and (C.74), we obtain (C.75). □

Example C.42. In this example, we illustrate the application Lemma C.41 on a simple, concrete model of a financial market. Fix a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ (Definition C.5 (a)) and let $B = (B_t)_{t\in\mathbb{R}_+}$ be an $\mathbb{R}^n$-valued stochastic process whose components $B^1, \dots, B^n$ are $\mathbb{R}$-valued $(\mathbb{F}, \mathbb{P})$-standard Brownian motions. Consider a financial market[20] consisting of a triple (a, r, S) of a risk-free interest rate $r = (r_t)_{t\in\mathbb{R}_+}$, a stochastic volatility $a = (a_t)_{t\in\mathbb{R}_+}$ and a risky asset price $S = (S_t)_{t\in\mathbb{R}_+}$, which under a risk-neutral probability measure are described by the SDEs

$$\begin{aligned} \mathrm{d}a_t &= \kappa_a(\bar{a} - a_t)\,\mathrm{d}t + \sigma_a\sqrt{a_t}\,\mathrm{d}B_t^1\,,\\ \mathrm{d}r_t &= \kappa_r(\bar{r} - r_t)\,\mathrm{d}t + \sigma_r\sqrt{r_t}\,\mathrm{d}B_t^2\,,\\ \mathrm{d}S_t &= r_t S_t\,\mathrm{d}t + \sqrt{a_t}S_t\,\mathrm{d}B_t^3\,, \end{aligned} \tag{C.76}$$

with initial values $(r_0, a_0, S_0) \in (0,\infty)^3$, where κ_a, $\bar{a}$, σ_a, κ_r, $\bar{r}$ and σ_r are nonnegative constants and B^1, B^2 and B^3 are $\mathbb{R}$-valued standard Brownian motions such that up to indistinguishability,

$$[B]_t = t\rho, \qquad t \in \mathbb{R}_+, \tag{C.77}$$

where the matrix $\rho \in \mathbb{R}^{3\times 3}$ is positive definite[21] and can be written as

$$\rho = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}.$$

Comparing with Example C.39 (a), we identify the matrix ρ with the process ρ in (C.62) and observe that (C.63) is satisfied. Being symmetric and positive definite, ρ admits a Cholesky

[20] The model in this example is taken from [41, §3.3]

[21] As a symmetric and positive definite matrix, ρ is automatically (this can be seen by Cholesky factorization or diagonalization) the covariance matrix of some random vector. Since ρ has only ones in its diagonal, the components of this random vector all have unit variance, hence ρ is in fact a correlation matrix. Moreover, in this case, $\rho_{12}, \rho_{13}, \rho_{23} \in (-1, 1)$, since full (anti-)correlation of any two random variables in a random vector with correlation matrix ρ implies that ρ is not positive definite, which is ruled out by assumption.

decomposition, $\rho = VV^{\mathsf{T}}$, where the matrix $V = (v_{ij})_{i,j\in\{1,2,3\}}$ is in this case given by

$$V = \begin{pmatrix} 1 & 0 & 0 \\ \rho_{12} & \sqrt{1-\rho_{12}^2} & 0 \\ \rho_{13} & \frac{\rho_{23}-\rho_{12}\rho_{13}}{\sqrt{1-\rho_{12}^2}} & \sqrt{1-\rho_{13}^2 - \frac{(\rho_{23}-\rho_{12}\rho_{13})^2}{1-\rho_{12}^2}} \end{pmatrix}.$$

From Lemma C.41, a triple (a, r, S) of $\mathbb{R}$-valued stochastic processes is a strong solution of (C.76) subject to (C.77) in the sense of Definition C.37 if and only if (a, r, S) is a strong solution in the classical sense (Definition C.15) of (cf. (C.67))

$$\begin{aligned} \mathrm{d}a_t &= \kappa_a(\bar{a} - a_t)\,\mathrm{d}t + \sigma_a\sqrt{a_t}\,\mathrm{d}W_t^1\,, \\ \mathrm{d}r_t &= \kappa_r(\bar{r} - r_t)\,\mathrm{d}t + \sigma_r\sqrt{r_t}\big(v_{21}\,\mathrm{d}W_t^1 + v_{22}\,\mathrm{d}W_t^2\big), \\ \mathrm{d}S_t &= r_tS_t\,\mathrm{d}t + \sqrt{a_t}S_t\big(v_{31}\,\mathrm{d}W_t^1 + v_{32}\,\mathrm{d}W_t^2 + v_{33}\,\mathrm{d}W_t^3\big), \end{aligned} \tag{C.78}$$

where $W = (W^1, W^2, W^3)$ is an $\mathbb{R}^3$-valued standard Brownian motion. Note that the existence of the interest rate process is less clear from (C.78) than from (C.76), since the model (C.76) specifies both the stochastic volatility and interest rate as CIR processes, whose strong existence and pathwise uniqueness can be obtained from results on SDEs driven by a single Brownian motion. See also the discussion in Example 3.3 for details.[22]

C.4. Flows of Classical Stochastic Differential Equations

The main result of this section, Theorem C.43, states that under the assumptions of strong existence and pathwise uniqueness for the SDE (C.6) (see Theorem C.34(b)), the up to indistinguishability unique strong solution of (C.6) on a given set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$ is given by a (deterministic) function of only the initial random vector ξ and the driving Brownian motion B. This function, which is independent of the set-up, has nice measurability properties and is continuous w.r.t. the initial value. It is called the *flow* of the SDE. As an important consequence, we obtain under the conditions of Theorem C.34 uniqueness in law (Definition C.21) of solutions of the classical SDE (C.6) with nonrandom dependence (Corollary C.57).

As in Section 2.1, the space $W_d = C(\mathbb{R}_+, \mathbb{K}^d)$ carries the σ-algebra $\mathcal{G}^d$ and the canonical filtration $\mathbb{G}^d = (\mathcal{G}_t^d)_{t\in\mathbb{R}_+}$ defined by (2.4) and (2.5) for d instead of n, respectively.[23] Denoting by $\mathbb{W}$ the Wiener measure on $(W_d, \mathcal{G}^d)$, let $\overline{\mathbb{G}}^d = (\overline{\mathcal{G}}_t^d)_{t\in\mathbb{R}_+}$ be the filtration where for each $t \in \mathbb{R}_+$, $\overline{\mathcal{G}}_t^d$ is the augmentation of $\mathcal{G}_t^d$ by $(\mathcal{G}^d, \mathbb{W})$-null sets (Definition B.3). The filtered probability space $(W_d, \mathcal{G}^d, \overline{\mathbb{G}}^d, \mathbb{W})$ will serve as the stochastic base in Step 1 of the proof of Theorem C.43 below.

[22] If, on the other hand, in (C.76), a and r are specified to follow instead of the CIR the Hull–White model (see Example 3.3), passage from the original "correlated" SDE to the resulting "decorrelated" analogue of (C.78) yields strong existence and pathwise uniqueness (for (a, r), but not for S) under mild assumptions on the coefficients of the model. Without the switch to an SDE driven by a standard Brownian motion, existence even of the interest rate and drift processes would be, strictly speaking, in question.

[23] In this section, we write $\mathcal{G}^d$ instead of $\mathcal{G}$ and $\mathbb{G}^d$ instead of $\mathbb{G}$ to differentiate the measurable structure on W_d from that on W_n.

On the other hand, solutions of (C.6) can be seen as W_n-valued random variables, with the measurable structure on W_n given by the σ-algebra $\mathcal{G}^n$ and the filtration $\mathbb{G}^n = (\mathcal{G}^n_t)_{t\in\mathbb{R}_+}$ defined in (2.4) and (2.5) respectively. The space W_n carries the topology $\mathcal{O}_{\mathrm{ucc}}$ of compact convergence, which by Lemma 2.1 (d) generates $\mathcal{G}^n$. This topology is generated by every metric of the form (2.2). In addition, for every sequence $(\alpha_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^*_+$ with $\sum_{k\in\mathbb{N}} \alpha_k < \infty$, $\mathcal{O}_{\mathrm{ucc}}$ can also be metrized by

$$\hat{\varrho}_{\mathrm{ucc}}(w, \tilde{w}) := \sum_{k\in\mathbb{N}} \alpha_k (1 \wedge p_k(w - \tilde{w})), \qquad w, \tilde{w} \in W_n. \tag{C.79}$$

Theorem C.43 (Existence and properties of the flow of an SDE under Lipschitz continuity and bounded growth conditions [38, Theorem 8.5] and [51, Theorem 13.1])**.** *Consider $\mathbb{G}^n$-progressive functions b and σ as in (C.4) which have the Lipschitz property (Definition C.25) and are of linearly bounded growth (Definition C.27). There exists a map*

$$F\colon \mathbb{K}^n \times W_d \to W_n, \quad (x, w) \mapsto F_x(w), \tag{C.80}$$

called the flow *of the SDE, with the following properties:*

(i) For every $w \in W_d$, the mapping $\mathbb{K}^n \ni x \mapsto F_x(w) \in W_n$ is continuous w.r.t. the topology of compact convergence[24] in W_n.

(ii) For every $x \in \mathbb{K}^n$, the process $\mathbb{R}_+ \times W_d \ni (t, w) \mapsto F_x(w)_t$ is $\overline{\mathbb{G}}^d$-adapted. Equivalently,

$$F_x^{-1}(\mathcal{G}^n_t) \subseteq \overline{\mathcal{G}}^d_t, \qquad t \in \overline{\mathbb{R}}_+. \tag{C.81}$$

(iii) For every $x \in \mathbb{K}^n$, for every stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and every $\mathbb{R}^d$- or $\mathbb{C}^d$-valued standard $(\mathbb{F}, \mathbb{P})$-Brownian motion B, the process $X = (X_t)_{t\in\mathbb{R}_+}$ defined by

$$X_t = F_x(B)_t, \qquad t \in \mathbb{R}_+,$$

is the pathwise unique strong solution of the SDE (C.6) with initial condition $X_0 \overset{\text{a.s.}}{=} x$.

(iv) For every set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$ (Definition C.5 (b)) with $|\xi| \in \mathcal{L}^2(\mathbb{P})$, the process $X = (X_t)_{t\in\mathbb{R}_+}$ defined by

$$X_t = F_\xi(B)_t, \qquad t \in \mathbb{R}_+,$$

is the pathwise unique strong solution of the SDE (C.6) on the given set-up with the initial condition $X_0 \overset{\text{a.s.}}{=} \xi$. If b and σ are bounded in the second argument (Definition C.26), the assumption $|\xi| \in \mathcal{L}^2(\mathbb{P})$ is superfluous.

Remark C.44. (a) The statement of Theorem C.43 above is based on [38, Theorem 8.5], where the special case $\mathbb{K} = \mathbb{R}$ and $d = 1$ is treated for SDE coefficients with nonrandom and instantaneous rather than pathwise dependence (see Definition C.9) on the solution. In addition, it is assumed that the coefficients are continuous in the time argument. This continuity assumption, which we avoid for now, allows for a relatively direct proof of Theorem C.43 by pathwise approximation arguments. This

[24] Families of compatible metrics are defined in (2.2) and (C.79).

direct proof is of independent interest and also yields strong existence of solutions of SDEs without an integrability condition on the initial condition (even when the coefficients are unbounded). For this reason, a modified version of Theorem C.43 under the added continuity assumption is stated and proved in Proposition C.58 at the end of this chapter.

(b) Results closely related to Theorem C.43 are developed in [51, §V.10-§V.13], also for the case $\mathbb{K} = \mathbb{R}$ and for coefficients which may depend pathwise on the solutions but are required to be $\mathbb{G}^n$-predictable. In addition, a stronger completeness condition than Definition C.3 is imposed on the underlying filtrations in [38] and a yet stronger one in [51]. We will obtain Theorem C.43, the proof of which begins on Page 158, as a consequence of Lemma C.54 and Theorem C.56 below, adapting the corresponding results in [51] to the setting of [53].

The proof of Theorem C.43 requires several auxiliary results. We begin with Lemma C.48 and Lemma C.49, which allow to approximate progressive processes satisfying an integrability condition by simpler, so-called elementary processes.

Definition C.45 (Elementary process [38, Definition 5.2])**.** Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic base. A $\mathbb{K}^d$-valued $\mathbb{F}$-elementary process is a stochastic process of the form

$$\varphi_s(\omega) = \sum_{i=0}^{m-1} \varphi_{(i)}(\omega)\, \mathbb{1}_{(t_i, t_{i+1}]}(s), \qquad (s, \omega) \in \mathbb{R}_+ \times \Omega, \tag{C.82}$$

where $0 = t_0 < \cdots < t_m < \infty$ and each $\varphi_{(i)}$ is a $\mathbb{K}^d$-valued $\mathcal{F}_{t_i}$-measurable random variable bounded in $\mathbb{K}^d$.

Every elementary process is $\mathbb{F}$-adapted and left-continuous, and thus $\mathbb{F}$-predictable. These processes are of interest, because they can be used to approximate more general progressive processes, see Lemma C.48 below.[25]

Definition C.46 ($\|\cdot\|_{L^2(M)}$, the space $L^2(M)$ [38, §5.1])**.** Let $M = (M_t)_{t \in \mathbb{R}_+}$ be a $\mathbb{K}$-valued continuous local martingale defined on a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$.

(a) For every $\mathcal{F} \otimes \mathcal{B}_{\mathbb{R}_+}$-measurable process[26] $H \colon \Omega \times \mathbb{R}_+ \to \mathbb{K}$, let[27]

$$\|H\|_{L^2(M)} := \mathbb{E}\left[\int_0^\infty |H_s|^2 \,\mathrm{d}[M]_s\right]^{1/2}. \tag{C.83}$$

[25] This approximation property is used as a stepping stone for the construction of the one-dimensional stochastic integral in [38, §5]. Parts of the development are retraced here to obtain analogous results in higher dimensions and to include complex-valued processes.

[26] We reverse the usual order of the arguments t and ω here in order to interpret the quadratic variation process $[M]$ more naturally as a σ-finite kernel from $(\Omega, \mathcal{F})$ to $(\mathbb{R}_+ \mathcal{B}_{\mathbb{R}_+})$, see Remark C.47(a) below.

[27] By monotone class arguments, beginning with $H = F \times \mathbb{1}_{(s,t]}$ for $F \in \mathcal{F}$ and $s < t$ in $\mathbb{R}_+$, the $\overline{\mathbb{R}}_+$-valued finite variation integral $\int_0^T |H_s|^2 \,\mathrm{d}[M]_s$ is defined and $\mathcal{F}$-measurable for all bounded $\mathcal{F} \otimes \mathcal{B}_{\mathbb{R}_+}$-measurable H and every $T \in \mathbb{R}_+$. This extends by monotone convergence to unbounded H and $T = \infty$, see also [53, Lemma 5.49], hence (C.83) is well-defined for every $\mathbb{K}$-valued $\mathcal{F} \otimes \mathcal{B}_{\mathbb{R}_+}$-measurable process H.

(b) The space $L^2(M)$ is defined as the set of equivalence classes of progressive processes $H\colon \Omega \times \mathbb{R}_+ \to \mathbb{K}$ such that

$$\|H\|_{L^2(M)} < \infty, \tag{C.84}$$

where two progressive processes H and H' satisfying this condition are identified if $\|H - H'\|_{L^2(M)} = 0$.

Remark C.47. Consider the setting of Definition C.46.

(a) Define the measure $\mu = \mathbb{P} \otimes \kappa_{[M]}$ on $(\Omega \times \mathbb{R}_+, \mathcal{F} \otimes \mathcal{B}_{\mathbb{R}_+})$ as the product of the probability measure $\mathbb{P}$ and the σ-finite transition kernel

$$\kappa_{[M]}\colon \Omega \times \mathcal{B}_{\mathbb{R}_+} \to \overline{\mathbb{R}}_+, \quad (\omega, A) \mapsto \int_{\mathbb{R}_+} \mathbb{1}_A(s)\,\mathrm{d}[M](\omega)_s\,, \tag{C.85}$$

see [53, Theorem 5.93] for a proof of the kernel properties and [53, Lemma 15.133] for details on the product of a measure and a kernel. Equip $\Omega \times \mathbb{R}_+$ with the progressive σ-algebra $\mathcal{P}$ (see [53, Exercise 3.36] or [38, §3.1, p. 43]), which is a sub-σ-algebra of $\mathcal{F} \otimes \mathcal{B}_{\mathbb{R}_+}$. The measure space $(\Omega \times \mathbb{R}_+, \mathcal{P}, \mu)$ is σ-finite, since by progressive measurability of $[M]$, $A_n := \{(\omega, t) \in \Omega \times \mathbb{R}_+ : [M]_t(\omega) \le n\}$, $n \in \mathbb{N}$ defines a sequence of $\mathcal{P}$-measurable sets with $\mu(A_n) \le n$ for every $n \in \mathbb{N}$. Moreover, $A_n \nearrow \Omega \times \mathbb{R}_+$ by continuity of $[M]$. For every $\mathcal{F} \otimes \mathcal{B}_{\mathbb{R}_+}$-measurable $H\colon \Omega \times \mathbb{R}_+ \to \mathbb{K}$, we have, using (C.83) and (C.85) for the first equality,

$$\|H\|_{L^2(M)}^2 = \int_{\Omega\times\mathbb{R}_+} |H|^2\,\mathrm{d}\mathbb{P} \otimes \kappa_{[M]} = \int_{\Omega\times\mathbb{R}_+} |H|^2\,\mathrm{d}\mu = \|H\|_{L^2(\mu)}^2$$

and thus

$$L^2(M) = L^2(\Omega \times \mathbb{R}_+, \mathcal{P}, \mu),$$

i.e. $L^2(M)$ is an ordinary L^2-space w.r.t. the σ-finite measure μ on the progressive σ-algebra $\mathcal{P}$. In fact, $L^2(M)$ is a Hilbert space (see [21, §5.5 and Theorem 6.6]) with inner product

$$\langle H, K\rangle_{L^2(M)} := \int_{\Omega\times\mathbb{R}_+} H\overline{K}\,\mathrm{d}\mu = \mathbb{E}\left[\int_0^\infty H_s \overline{K}_s\,\mathrm{d}[M]_s\right], \qquad H, K \in L^2(M). \tag{C.86}$$

The second equality above has to be interpreted with care: while the Lebesgue integral $\int_{\Omega\times\mathbb{R}_+} H\overline{K}\,\mathrm{d}\mu$ is well-defined and finite by the Cauchy–Schwarz inequality for arbitrary representatives of $H, K \in L^2(M)$, the integral $\int_0^\infty H'_s \overline{K}'_s\,\mathrm{d}[M]_s$ could, for given representatives H' of H and K' of K be ill-defined on an $(\mathcal{F}_\infty, \mathbb{P})$-null set.[28] Nevertheless, we can evaluate $\langle H, K\rangle_{L^2(M)}$ using the right-hand side of (C.86) by passing to indistinguishable modifications H'' of H and K'' of K representing the same respective elements of $L^2(M)$ and for which $\int_0^\infty H''_s \overline{K}''_s\,\mathrm{d}[M]_s$ is $\mathbb{K}$-valued on all of Ω.

[28] A similar problem appears in Fubini's theorem, where a function integrable w.r.t. a product measure $P \otimes Q$ may fail to have marginal integrals w.r.t. P on a Q-null-set.

(b) If a progressive process $H\colon \Omega\times\mathbb{R}_+\to\mathbb{K}$ satisfies (C.84), then $\int_0^\infty |H_s|^2\,\mathrm{d}[M]_s<\infty$, $\mathbb{P}$-a.s. This does not guarantee that H lies in the space $L(M)$ of processes stochastically integrable against M in the sense[29] of Definition C.1 (b), for which H has to satisfy the marginally stronger condition (C.2). However $H\in L^2(M)$ implies that $N:=\{\int_0^\infty |H_s|^2\,\mathrm{d}[M]_s=\infty\}$ is an $(\mathcal{F}_\infty,\mathbb{P})$-null set, and thus the process $H':=H\,\mathbb{1}_{N^c}$ is indistinguishable from H. Since $\mathbb{F}$ satisfies the usual completeness assumption (Definition C.3), H' is $\mathbb{F}$-progressive and is thus identified with H w.r.t. $\|\cdot\|_{L^2(M)}$. By construction, $H'\in L(M)$, hence the stochastic integral $H'\cdot M=\int_0^\cdot H'_s\,\mathrm{d}M_s$ is well-defined and

$$[H'\cdot M]\overset{\text{uti}}{=}\int_0^\cdot |H'_s|^2\,\mathrm{d}[M]_s$$

by [53, Corollary 5.103], which implies in addition that $H'\cdot M$ lies in $\mathcal{M}^2$, the space of continuous martingales started from zero and bounded in $L^2(\mathbb{P})$ and that $|H'\cdot M|^2-[H'\cdot M]$ is a real-valued uniformly integrable martingale. Moreover, we have the well-known Itô isometry

$$\mathbb{E}[|H'\cdot M|_t^2]=\|H'\,\mathbb{1}_{[0,t]}\|_{L^2(M)},\qquad t\in\overline{\mathbb{R}}_+,$$

using again [53, Corollary 5.103] for the case $t\in\mathbb{R}_+$ and martingale convergence [53, Theorem 7.87] together with Doob's L^2-inequality [53, Theorem 4.95] to obtain $(H'\cdot M)_t\xrightarrow[t\to\infty]{L^2(\mathbb{P})}(H'\cdot M)_\infty$ in the case $t=\infty$.

Lemma C.48 (Density of elementary processes [38, Proposition 5.3])**.** *Let $M=(M_t)_{t\in\mathbb{R}_+}$ be a $\mathbb{K}$-valued, square integrable[30] continuous martingale on a stochastic base $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$. Denote by $\mathcal{E}$ the set of $\mathbb{K}$-valued elementary processes (Definition C.45), viewed as a subspace of $L^2(M)$ (Definition C.46 (b)) by identifying processes with $\|\cdot\|_{L^2(M)}$-distance zero. Then $\mathcal{E}$ is dense in $L^2(M)$.*

Proof. Since $L^2(M)$ is a Hilbert space by Remark C.47 (a), it can be decomposed using [53, Theorem 14.12] as the internal direct sum of the closure $\overline{\mathcal{E}}$ of $\mathcal{E}$ w.r.t. $\|\cdot\|_{L^2(M)}$ and its orthogonal complement $(\overline{\mathcal{E}})^\perp$, i.e.

$$L^2(M)=\overline{\mathcal{E}}\oplus(\overline{\mathcal{E}})^\perp.$$

Consequently, $\overline{\mathcal{E}}$ is a proper subspace of $L^2(M)$ if and only if $(\overline{\mathcal{E}})^\perp=\mathcal{E}^\perp\neq\{0\}$. Thus it suffices to show that if $H\in L^2(M)$ is orthogonal to $\mathcal{E}$, then $\|H\|_{L^2(M)}=0$. By Remark C.47 (b), H has a representative in $L(M)$, which we will also denote by H in slight abuse of notation. Using that H then satisfies (C.2) together with the bound $|H|\le 1\vee|H|^2\le 1+|H|^2$,

$$\int_0^t |H_s(\omega)|\,\mathrm{d}[M]_s(\omega)<\infty,\qquad (\omega,t)\in\Omega\times\mathbb{R}_+.\tag{C.87}$$

[29] Note that $H\in L^2(M)$ does guarantee stochastic integrability in the sense of [38, §5.1.2], but this wider sense of integrability is obtained essentially (see the proof of [38, Theorem 5.6]) by passing to an indistinguishable modification H' of H which is in $L(M)$, as described above.

[30] Square integrability of M means that $M_t\in\mathcal{L}^2(\mathbb{P})$ for all $t\in\mathbb{R}_+$.

Using [53, Lemma 5.49], it follows that

$$X_t := \int_0^t H_s \, \mathrm{d}[M]_s \,, \qquad t \in \mathbb{R}_+, \tag{C.88}$$

defines an $\mathbb{F}$-adapted continuous finite variation process. We next verify that X is also an integrable process, i.e. that $X_t \in \mathcal{L}^1(\mathbb{P})$ for all $t \in \mathbb{R}_+$. Square integrability of M implies (see [38, Theorem 4.13] or [53, Exercise 5.88]) that $[M]$ is an integrable process. By the Kunita–Watanabe inequality[31] ([53, Corollary 5.94]), for every $t \in \mathbb{R}_+$,

$$\begin{aligned}\mathbb{E}[|X_t|] \le \mathbb{E}\left[\int_0^t |H_s| \, \mathrm{d}[M]_s\right] &\le \mathbb{E}\left[\int_0^t |H_s|^2 \, \mathrm{d}[M]_s\right]^{1/2} \mathbb{E}[[M]_t]^{1/2} \\ &\le \|H\|_{L^2(M)} \, \mathbb{E}[[M]_t]^{1/2} < \infty.\end{aligned}$$

We next show that $\|H\|_{L^2(M)} = 0$ by establishing $X \overset{\text{uti}}{=} 0$ as a consequence of X being not only a finite variation process, but also an $\mathbb{F}$-martingale. To this end, fix $s < t$ in $\mathbb{R}_+$ and let $F \in \mathcal{F}_s$. Define the elementary process $\varphi \in \mathcal{E}$ by $\varphi_r(\omega) := \mathbb{1}_F(\omega) \, \mathbb{1}_{(s,t]}(r)$ for all $(\omega, r) \in \Omega \times \mathbb{R}_+$. Using that $H \in \mathcal{E}^\perp$ by assumption,

$$0 = \langle H, \varphi \rangle_{L^2(M)} = \mathbb{E}\left[\mathbb{1}_F \int_s^t H_r \, \mathrm{d}[M]_r\right] = \mathbb{E}[\mathbb{1}_F (X_t - X_s)],$$

which implies the fairness property of martingales. By construction, X is adapted and integrable and thus indeed an $\mathbb{F}$-martingale. Since X is additionally a finite variation process, [53, Lemma 5.51] implies that X is indistinguishable from the constant process with value $X_0 = 0$. Let $N := \bigcup_{t \in \mathbb{R}_+ \cap \mathbb{Q}} \{X_t \neq 0\}$, which is an $(\mathcal{F}_\infty, \mathbb{P})$-null set with $X_t(\omega) = 0$ for all $(\omega, t) \in N^{\mathsf{c}} \times \mathbb{R}_+$ by continuity of X. Consequently,

$$\int_0^t H_s(\omega) \, \mathrm{d}[M]_s(\omega) = 0, \qquad (\omega, t) \in \mathbb{R}_+ \times N^{\mathsf{c}}. \tag{C.89}$$

For all $\omega \in \Omega$ and $T \in \mathbb{R}_+$, let $\nu_{\omega,T}$ denote the complex measure with density $H_\cdot(\omega)$ w.r.t. the restriction[32] to $([0,T], \mathcal{B}_{[0,T]})$ of the Lebesgue–Stieltjes measure $\kappa_{[M]}(\omega, \cdot\,)$ on $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$ induced by $[M]_\cdot(\omega)$ (see (C.85)). For all $\omega \in N^{\mathsf{c}}$ and $T \in \mathbb{R}_+$, (C.89) implies that $\nu_{\omega,T}$ vanishes on the π-system $\{[0,t] : t \in [0,T]\}$ which generates $\mathcal{B}_{[0,T]}$. By a Dynkin class argument, see [53, Corollary 15.67], $\nu_{\omega,T}$ equals the zero measure on $\mathcal{B}_{[0,T]}$. Consequently,

$$0 = \int_0^T \overline{H}_s(\omega) \, \mathrm{d}\nu_{\omega,T}\,(s) = \int_0^T |H_s(\omega)|^2 \, \mathrm{d}[M]_s(\omega) \qquad (\omega, T) \in N^{\mathsf{c}} \times \mathbb{R}_+,$$

where we have used the chain rule for measures defined by densities, see [21, Proposition 3.9 (a)] and the definition (C.85) of $\kappa_{[M]}$ in the second equality. By monotone convergence, the previous display extends to the case $T = \infty$, hence $\int_0^\infty |H_s|^2 \, \mathrm{d}[M]_s \overset{\text{a.s.}}{=} 0$ and thus $\|H\|_{L^2(M)} = 0$, proving the claim. □

[31] Alternatively, we can use the Cauchy–Schwarz inequality for the inner product $\langle \cdot\,, \cdot \rangle_{L^2(M)}$ defined in (C.86).

[32] The detour to finite time intervals is necessary, since $\int_0^\infty |H_s(\omega)| \, \mathrm{d}[M]_s$ may be infinite. In this case, $H_\cdot(\omega)$ can fail to be even pseudo-integrable over $\mathbb{R}_+$ w.r.t. $\kappa_{[M]}(\omega, \cdot\,)$ and thus cannot be used as a density to define $\nu_{\omega,\infty}$. On the other hand, the integrability property (C.87) ensures that $\nu_{\omega,T}$ is a well-defined complex measure for all $(\omega, T) \in \Omega \times \mathbb{R}_+$.

Lemma C.48 can be generalized to $\mathbb{K}^d$-valued square integrable continuous martingales, but in the sequel, we will only need the following special case of $\mathbb{K}^d$-valued standard Brownian motion.

Lemma C.49. *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic base and $B = (B_t)_{t\in\mathbb{R}_+}$ a $\mathbb{K}^d$-valued standard $\mathbb{F}$-Brownian motion. For every progressive process $\varphi\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^d$ with*

$$\|\varphi\|_{L^2(B)} \coloneqq \mathbb{E}\left[\int_0^\infty |\varphi_s|^2 \,\mathrm{d}s\right]^{1/2} < \infty, \tag{C.90}$$

there exists a sequence $(\varphi^k)_{k\in\mathbb{N}}$ of $\mathbb{K}^d$-valued elementary processes (Definition C.45) with

$$\|\varphi^k - \varphi\|_{L^2(B)} \xrightarrow[k\to\infty]{} 0.$$

Proof. It suffices to show that for every $\varepsilon > 0$, there exists a $\mathbb{K}^d$-valued elementary process $\tilde{\varphi}$ with

$$\|\tilde{\varphi} - \varphi\|^2_{L^2(B)} < \varepsilon. \tag{C.91}$$

Fix $\varepsilon > 0$ and consider the case $\mathbb{K} = \mathbb{C}$ without loss of generality. Each component $B^1, \dots, B^d$ of B is a $\mathbb{C}$-valued Brownian motion, hence we have $[B^j]_t = 2t$ for all $j \in \{1, \dots, d\}$ and all $t \in \mathbb{R}_+$, see e.g. [53, Exercise 5.75]. By (C.90), it follows that $\varphi^j \in L^2(B^j)$ for each $j \in \{1, \dots, d\}$, which by Lemma C.48 implies the existence of $\mathbb{C}$-valued elementary processes $\tilde{\varphi}^j$, $j \in \{1, \dots, d\}$, with

$$\mathbb{E}\left[\int_0^\infty |\tilde{\varphi}^j_s - \varphi^j_s|^2 \,\mathrm{d}s\right] < \frac{\varepsilon}{d}, \qquad j \in \{1, \dots, d\}. \tag{C.92}$$

By passing to a common refinement, we may assume that the elementary processes $\tilde{\varphi}^j$, $j \in \{1, \dots, d\}$ have the same intervals of constancy in time (see Definition C.45). Thus $\tilde{\varphi} \coloneqq (\tilde{\varphi}^1, \dots, \tilde{\varphi}^d)$ is a $\mathbb{K}^d$-valued elementary process. Since $|\tilde{\varphi} - \varphi|^2 = \sum_{j=1}^d |\tilde{\varphi}^j - \varphi^j|^2$, it follows with (C.92) that

$$\mathbb{E}\left[\int_0^\infty |\tilde{\varphi}_s - \varphi_s|^2 \,\mathrm{d}s\right] < \varepsilon. \qquad \square$$

Definition C.50 (The space $L^2(B)$)**.** Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic base and $B = (B_t)_{t\in\mathbb{R}_+}$ a $\mathbb{K}^d$-valued standard $\mathbb{F}$-Brownian motion. The space $L^2(B)$ is defined as the set of equivalence classes of progressive processes $H\colon \Omega \times \mathbb{R}_+ \to \mathbb{K}^d$ such that (see (C.90))

$$\|H\|_{L^2(B)} < \infty, \tag{C.93}$$

where two progressive processes H and H' satisfying this condition are identified if $\|H - H'\|_{L^2(B)} = 0$.

Remark C.51. In the setting of Lemma C.49, let $\varphi\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^d$ be a progressive process with $\|\varphi\|_{L^2(B)} < \infty$ and let $(\varphi^k)_{k\in\mathbb{N}}$ be a sequence of $\mathbb{K}^d$-valued elementary processes with $\|\varphi^k - \varphi\|_{L^2(B)} \xrightarrow[k\to\infty]{} 0$. We have $\varphi^k \in L(B)$ for each $k \in \mathbb{N}$ by boundedness of elementary processes, and furthermore

$$\int_0^t |\varphi_s|^2(\omega) \,\mathrm{d}s < \infty, \qquad (t, \omega) \in \mathbb{R}_+ \times \Omega,$$

i.e. $\varphi \in L(B)$ (Definition C.1(d)), after possibly passing to an indistinguishable modification of φ. For each $k \in \mathbb{N}$, the quadratic variation process of the stochastic integral $(\varphi^k - \varphi) \cdot B$ is given by

$$[(\varphi^k - \varphi) \cdot B] \overset{\text{uti}}{=} c_{\mathbb{K}} \int_0^{\cdot} |\varphi^k_s - \varphi_s|^2 \,\mathrm{d}s\,,$$

with $c_{\mathbb{K}} \coloneqq \dim_{\mathbb{R}}(\mathbb{K})$, see [53, Exercise 5.116]. Hence $\|\varphi^k - \varphi\|_{L^2(B)} \xrightarrow[k\to\infty]{} 0$ implies $\mathbb{E}[[(\varphi^k - \varphi) \cdot B]_\infty] \xrightarrow[k\to\infty]{} 0$. Therefore, by [53, Corollary 5.124],

$$\sup_{t\in\mathbb{R}_+} \left| \int_0^t (\varphi^k_s - \varphi_s)\,\mathrm{d}B_s \right| \xrightarrow[k\to\infty]{\mathbb{P}} 0. \tag{C.94}$$

Our next objective is Lemma C.54 below, for which the following set-up (Definition C.5) for stochastic integration will be used.

Definition C.52 (Canonical set-up [51, §V.10])**.** Let $\mathbb{G}^d = (\mathcal{G}^d_t)_{t\in\mathbb{R}_+}$ be the canonical filtration on $(W_d, \mathcal{G}^d)$, see (2.4) and (2.5), and let $\mathbb{W}$ denote the d-dimensional Wiener measure on $(W_d, \mathcal{G}^d)$. On

$$\Omega \coloneqq \mathbb{K}^n \times W_d,$$

define the projections

$$\begin{aligned} &\xi\colon \Omega \to \mathbb{K}^n, \quad \xi(x,w) = x, \\ &B_t\colon \Omega \to \mathbb{K}^d, \quad B_t(x,w) = w_t, \qquad t \in \mathbb{R}_+. \end{aligned} \tag{C.95}$$

Let

$$\begin{aligned} \mathcal{F}_t &\coloneqq \sigma(\xi, B_s : s \in [0,t]) = \mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{G}^d_t, \qquad t \in \mathbb{R}_+, \\ \mathcal{F} &\coloneqq \mathcal{F}_\infty \coloneqq \sigma(\xi, B_s : s \in \mathbb{R}_+) = \mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{G}^d, \end{aligned} \tag{C.96}$$

and let $\mu \in \mathcal{P}(\mathbb{K}^n)$ be an initial distribution. Define

$$\mathbb{P} = \mu \otimes \mathbb{W}$$

and $\overline{\mathbb{F}} \coloneqq (\bar{\mathcal{F}}_t)_{t\in\mathbb{R}_+}$, where for each $t \in \mathbb{R}_+$, $\bar{\mathcal{F}}_t$ denotes the augmentation of $\mathcal{F}_t$ by $(\mathcal{F}_\infty, \mathbb{P})$-null sets (Definition B.3). On the stochastic base $(\Omega, \mathcal{F}, \overline{\mathbb{F}}, \mathbb{P})$, B is a standard $(\overline{\mathbb{F}}, \mathbb{P})$-Brownian motion[33] and ξ is a $\mathbb{K}^n$-valued $\mathcal{F}_0$-measurable random vector with $\mathcal{L}[\xi] = \mu$. The set-up $(\Omega, \mathcal{F}, \overline{\mathbb{F}}, \mathbb{P}, B, \xi)$ is called the *canonical set-up with initial distribution* μ.[34]

We continue with a preparatory result needed for Lemma C.54 below.

Lemma C.53. *Let $\mu \in \mathcal{P}(\mathbb{K}^n)$ and $(\Omega, \mathcal{F}, \overline{\mathbb{F}}, \mathbb{P}, \xi, B)$ be the canonical set-up with initial distribution μ, and let $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}}, \tilde{\xi}, \tilde{B})$ be a set-up (Definition C.5(b)) with a filtration $\tilde{\mathbb{F}} = (\tilde{\mathcal{F}}_t)_{t\in\mathbb{R}_+}$ and $\mathcal{L}_{\tilde{\mathbb{P}}}[\tilde{\xi}] = \mu$. Let $\tilde{\mathcal{F}}_\infty \coloneqq \sigma(\tilde{\mathcal{F}}_t : t \in \mathbb{R}_+)$.*

(a) The map $\tilde{\Omega} \ni \tilde{\omega} \mapsto (\tilde{\xi}(\tilde{\omega}), \tilde{B}(\tilde{\omega})) \in \Omega$ is $\tilde{\mathcal{F}}_t$-$\bar{\mathcal{F}}_t$-measurable for each $t \in \overline{\mathbb{R}}_+$, and

$$\mathcal{L}_{\tilde{\mathbb{P}}}[(\tilde{\xi}, \tilde{B})] = \mu \otimes \mathbb{W} = \mathcal{L}_{\mathbb{P}}[(\xi, B)]. \tag{C.97}$$

[33] By definition of the filtration $\mathbb{F}$, see (C.96), B is a standard $(\mathbb{F}, \mathbb{P})$-Brownian motion. By Remark C.4(a), it is also an $\overline{\mathbb{F}}$-Brownian motion.

[34] We differ in notation from [51], where $\mathcal{F}_t$ is denoted by $\mathcal{F}^\circ_t$, and $\bar{\mathcal{F}}_t$ by $\mathcal{F}_t$.

(b) Let $\varphi\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^n$ be $\overline{\mathbb{F}}$-progressive. The composition

$$\tilde{\varphi}\colon \mathbb{R}_+ \times \tilde{\Omega} \ni (s, \tilde{\omega}) \mapsto \varphi(s, (\tilde{\xi}(\tilde{\omega}), \tilde{B}(\tilde{\omega})))$$

is $\tilde{\mathbb{F}}$-progressive.

Proof. (a) Define $g\colon \tilde{\Omega} \to \Omega$, $g(\tilde{\omega}) = (\tilde{\xi}(\tilde{\omega}), \tilde{B}(\tilde{\omega}))$ and recall that by Definition C.52, $\mathcal{F}_t = \sigma(\xi, B_s : s \in [0,t] \cap \mathbb{R})$ for all $t \in \overline{\mathbb{R}}_+$. For every $t \in \overline{\mathbb{R}}_+$, we have

$$\xi \circ g = \tilde{\xi} \quad \text{and} \quad B_s \circ g = \tilde{B}_s, \quad s \in [0,t] \cap \mathbb{R},$$

and all of these maps are $\tilde{\mathcal{F}}_t$-measurable. Hence g is $\tilde{\mathcal{F}}_t$-$\mathcal{F}_t$-measurable for every $t \in \overline{\mathbb{R}}_+$. In particular, we can view $(\tilde{\xi}, \tilde{B})$ as an Ω-valued random variable measurable w.r.t. $\tilde{\mathcal{F}}_\infty$-$\mathcal{F}_\infty$, which we use to next verify (C.97). By Definition C.5 (b) of a set-up, $\tilde{\xi}$ is $\tilde{\mathcal{F}}_0$-measurable and $\tilde{B}$ is a standard $(\tilde{\mathbb{F}}, \tilde{\mathbb{P}})$-Brownian motion and thus independent of $\tilde{\mathcal{F}}_0$ w.r.t. $\tilde{\mathbb{P}}$. Hence $\tilde{B}$ is independent from $\tilde{\xi}$ w.r.t. $\tilde{\mathbb{P}}$, implying

$$\mathcal{L}_{\tilde{\mathbb{P}}}[(\tilde{\xi}, \tilde{B})] = \mathcal{L}_{\tilde{\mathbb{P}}}[\tilde{\xi}] \otimes \mathcal{L}_{\tilde{\mathbb{P}}}[\tilde{B}] = \mu \otimes \mathbb{W} = \mathcal{L}_{\mathbb{P}}[(\xi, B)],$$

and thus (C.97). It remains to show that g is $\tilde{\mathcal{F}}_t$-$\bar{\mathcal{F}}_t$-measurable for all $t \in \overline{\mathbb{R}}_+$. Let $t \in \overline{\mathbb{R}}_+$ and $\bar{F} \in \bar{\mathcal{F}}_t$. Since $\bar{\mathcal{F}}_t$ is the augmentation of $\mathcal{F}_t$ by $(\mathcal{F}_\infty, \mathbb{P})$-null sets (Definition B.3), $\bar{F}$ can be written using Lemma B.4 as $\bar{F} = (F \cup N) \setminus N'$ for an $F \in \mathcal{F}_t$ and $\mathbb{P}$-null sets $N, N' \in \mathcal{F}_\infty$. With (C.97) and the already established $\tilde{\mathcal{F}}_\infty$-$\mathcal{F}_\infty$-measurability of g, we obtain

$$\tilde{\mathbb{P}}[g^{-1}(N)] = \tilde{\mathbb{P}}[(\tilde{\xi}, \tilde{B}) \in N] = 0 \qquad \text{and} \qquad \tilde{\mathbb{P}}[g^{-1}(N')] = \tilde{\mathbb{P}}[(\tilde{\xi}, \tilde{B}) \in N'] = 0.$$

Hence $g^{-1}(N)$ and $g^{-1}(N')$ are $(\tilde{\mathcal{F}}_\infty, \tilde{\mathbb{P}})$-null sets. Since $\tilde{\mathbb{F}}$ satisfies the usual completeness assumption (Definition C.3), we have $g^{-1}(N), g^{-1}(N') \in \tilde{\mathcal{F}}_t$. Using $\tilde{\mathcal{F}}_t$-$\mathcal{F}_t$-measurability of g, we obtain

$$g^{-1}(\bar{F}) = (g^{-1}(F) \cup g^{-1}(N)) \setminus g^{-1}(N') \in \tilde{\mathcal{F}}_t.$$

(b) Let $G\colon \mathbb{R}_+ \times \tilde{\Omega} \to \mathbb{K}^n$, $G(s, \tilde{\omega}) \coloneqq (s, g(\tilde{\omega}))$. We have $\tilde{\varphi} = \varphi \circ G$ and

$$\tilde{\varphi}\restriction_{[0,t]\times\tilde{\Omega}} = \varphi\restriction_{[0,t]\times\Omega} \circ G\restriction_{[0,t]\times\tilde{\Omega}}, \qquad t \in \mathbb{R}_+.$$

For each $t \in \mathbb{R}_+$, $\varphi\restriction_{[0,t]\times\Omega}$ is $\mathcal{B}_{[0,t]} \otimes \bar{\mathcal{F}}_t$-measurable by $\overline{\mathbb{F}}$-progressive measurability of φ, hence it suffices to show that G is $(\mathcal{B}_{[0,t]} \otimes \tilde{\mathcal{F}}_t)$-$(\mathcal{B}_{[0,t]} \otimes \bar{\mathcal{F}}_t)$-measurable for each $t \in \mathbb{R}_+$. But this follows directly from the fact that g is $\tilde{\mathcal{F}}_t$-$\bar{\mathcal{F}}_t$-measurable for all $t \in \mathbb{R}_+$ by part (a). ☐

Lemma C.54 ([51, Lemma 10.1])**.** *Let $\mu \in \mathcal{P}(\mathbb{K}^n)$ and let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B)$ be the canonical set-up with initial distribution μ (Definition C.52). Let $\varphi\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^d$ be an $\overline{\mathbb{F}}$-progressive map with*

$$\int_0^t |\varphi_s(\xi, B)|^2 \, \mathrm{d}s < \infty, \qquad t \in \mathbb{R}_+. \tag{C.98}$$

Then there exists a function $\Phi\colon \mathbb{K}^n \times W_d \to W_1$ which is $(\mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{G}^d)$-$\mathcal{G}^1$-measurable and satisfies

$$\int_0^\cdot \varphi_s(\xi, B)\,\mathrm{d}B_s \overset{\text{uti}}{=} \Phi(\xi, B). \tag{C.99}$$

Moreover, on every set-up $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}}, \tilde{\xi}, \tilde{B})$ where $\tilde{\xi}$ has law μ,

$$\int_0^\cdot \varphi_s(\tilde{\xi}, \tilde{B})\,\mathrm{d}\tilde{B}_s \overset{\text{uti}}{=} \Phi(\tilde{\xi}, \tilde{B}). \tag{C.100}$$

Remark C.55 (Well-definedness of the stochastic integrals in Lemma C.54)**.** Consider the setting of Lemma C.54.

(a) Since φ is $\overline{\mathbb{F}}$-progressive, (C.98) implies that $\varphi(\xi, B) \in L(B)$ (see Definition C.1(d)), hence the stochastic integral $\int_0^\cdot \varphi_s(\xi, B)\,\mathrm{d}B_s$ in (C.99) is well-defined. Given an alternative set-up $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}}, \tilde{\xi}, \tilde{B})$, we have by Lemma C.53(b) that $\mathbb{R}_+ \times \tilde{\Omega} \ni (s, \tilde{\omega}) \mapsto \varphi_s(\tilde{\xi}(\tilde{\omega}), \tilde{B}(\tilde{\omega}))$ is $\tilde{\mathbb{F}}$-progressive. Moreover, since (ξ, B) is the identity on $\Omega = \mathbb{K}^n \times W_d$ by (C.95) and since $(\tilde{\xi}, \tilde{B})$ takes values in Ω, (C.98) implies that

$$\int_0^t |\varphi_s(\tilde{\xi}, \tilde{B})|^2\,\mathrm{d}s < \infty, \qquad t \in \mathbb{R}_+. \tag{C.101}$$

Hence $\varphi(\tilde{\xi}, \tilde{B}) \in L(\tilde{B})$, i.e. the stochastic integral $\int_0^\cdot \varphi_s(\tilde{\xi}, \tilde{B})\,\mathrm{d}\tilde{B}_s$ in (C.100) is also well-defined.[35]

(b) Given an $\overline{\mathbb{F}}$-progressive map φ as in Lemma C.54, the existence of a function $\Phi\colon \mathbb{K}^n \times W_d \to W_1$ satisfying (C.99) follows, using that by (C.96), $\mathcal{F} = \mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{G}^d$, already from $\mathcal{F}$-$\mathcal{G}^1$-measurability of the stochastic integral $\int_0^\cdot \varphi_s(\xi, B)\,\mathrm{d}B_s$ and Doob's factorization lemma B.8(a). However, it is not a priori clear why the map Φ obtained in this way should satisfy (C.100).

Proof of Lemma C.54. The main idea is to prove the lemma first for the case where φ is an elementary process and subsequently extend to processes in $L^2(B)$ by using Lemma C.49. The extension to progressive processes satisfying (C.98) then happens by localization. In each case, the map Φ is first constructed on the canonical set-up $(\Omega, \mathcal{F}, \overline{\mathbb{F}}, \mathbb{P}, \xi, B)$ from Definition C.52. Subsequently, verify in each case (C.100), i.e. that the map Φ also works on arbitrary alternative set-ups. Crucially, this property is straightforward to show for elementary processes and conserved when complicating the form of φ.

Throughout this proof, fix an alternative set-up $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}}, \tilde{\xi}, \tilde{B})$.

Step 1 (Lemma C.54 holds for elementary φ). Assume that φ is a $\mathbb{K}^d$-valued $\overline{\mathbb{F}}$-elementary process (Definition C.45), i.e. φ can be written as

$$\varphi_s(\omega) = \sum_{i=0}^{m-1} \varphi_{(i)}(\omega)\,\mathbb{1}_{(t_i, t_{i+1}]}(s), \qquad (s, \omega) \in \mathbb{R}_+ \times \Omega, \tag{C.102}$$

[35] Some authors (see e.g. [38, §5.1.2] or [51, Theorem 30.7]) use a weaker integrability requirement, where the pathwise local square integrability condition (C.98) is only required to hold almost surely. In this case, (C.98) and the fact that $\mathcal{L}_{\tilde{\mathbb{P}}}[(\tilde{\xi}, \tilde{B})] = \mathcal{L}_{\mathbb{P}}[(\xi, B)]$ (see Lemma C.53(a)) allow to similarly conclude that (C.101) holds $\tilde{\mathbb{P}}$-almost surely.

with $0 = t_0 < \cdots < t_m < \infty$ and where each $\varphi_{(i)}$ is bounded and $\bar{\mathcal{F}}_{t_i}$-measurable. By definition C.52 of the canonical set-up, $\bar{\mathcal{F}}_{t_i}$ is the augmentation of $\mathcal{F}_{t_i}$ by the $(\mathcal{F}, \mathbb{P})$-null sets for each $i \in \{0, \dots, m\}$. There exist by Lemma B.7 $\mathbb{K}^d$-valued random vectors $\varphi'_{(0)}, \dots, \varphi'_{(m-1)}$ with $\varphi'_{(i)}$ measurable w.r.t. $\mathcal{F}_{t_i}$ and $\varphi_{(i)} \overset{\text{a.s.}}{=} \varphi'_{(i)}$ for each $i \in \{0, \dots, m-1\}$. Consequently,

$$\int_0^{\cdot} \varphi_s \, \mathrm{d}B_s \overset{\text{uti}}{=} \sum_{i=0}^{m-1} \varphi_{(i)} \left(B^{t_{i+1}} - B^{t_i}\right) \overset{\text{uti}}{=} \sum_{i=0}^{m-1} \varphi'_{(i)} \left(B^{t_{i+1}} - B^{t_i}\right), \tag{C.103}$$

where the first equality is due to the elementary form of φ, see also [53, Remark 5.101].[36] Let

$$\Phi \colon \Omega \to W_1, \quad \Phi(\omega) \coloneqq \sum_{i=0}^{m-1} \varphi'_{(i)}(\omega) \left(B^{t_{i+1}}(\omega) - B^{t_i}(\omega)\right) \tag{C.104}$$

and note that Φ is $\mathcal{F}$-$\mathcal{G}^1$-measurable, by $\mathcal{F}$-$\mathcal{B}_{\mathbb{K}}$-measurability of each projection $\Phi_t \colon \omega \mapsto \sum_{i=0}^{m-1} \varphi'_{(i)}(\omega)(B_{t_{i+1}\wedge t}(\omega) - B_{t_i \wedge t}(\omega))$, $t \in \mathbb{R}_+$. By definition, $\Omega = \mathbb{K}^n \times W_d$ and $\mathcal{F} = \mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{G}^d$ (see (C.96)), hence Φ is $(\mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{G}^d)$-$\mathcal{G}^1$-measurable. Moreover, (ξ, B) is the identity on Ω, and thus (C.99) follows from (C.103) and (C.104).

We next establish (C.100). Let $N \coloneqq \bigcup_{i=0}^{m-1} \{\varphi_{(i)} \neq \varphi'_{(i)}\}$, which is an $(\mathcal{F}, \mathbb{P})$-null set. Since $\mathcal{L}_{\tilde{\mathbb{P}}}[(\tilde{\xi}, \tilde{B})] = \mathcal{L}_{\mathbb{P}}[(\xi, B)]$ by Lemma C.53(a),

$$\tilde{\mathbb{P}}[(\tilde{\xi}, \tilde{B}) \in N] = \mathbb{P}[(\xi, B) \in N] = \mathbb{P}[N] = 0. \tag{C.105}$$

Using that $\varphi_{(i)}$ is $\bar{\mathcal{F}}_{t_i}$-$\mathcal{B}_{\mathbb{K}^n}$-measurable for each $i \in \{0, \dots, m-1\}$ together with the fact that $(\tilde{\xi}, \tilde{B})$ is $\tilde{\mathcal{F}}_t$-$\bar{\mathcal{F}}_t$-measurable for all $t \in \mathbb{R}_+$ by Lemma C.53(a), it follows that $\varphi(\tilde{\xi}, \tilde{B})$ is an $\tilde{\mathbb{F}}$-elementary process and thus in particular in $L(\tilde{B})$. Consequently,

$$\begin{aligned}\int_0^{\cdot} \varphi_s(\tilde{\xi}, \tilde{B}) \, \mathrm{d}\tilde{B}_s &\overset{\text{uti}}{=} \sum_{i=0}^{m-1} \varphi_{(i)}(\tilde{\xi}, \tilde{B}) \big(\tilde{B}^{t_{i+1}} - \tilde{B}^{t_i}\big) \\ &\overset{\text{uti}}{=} \sum_{i=0}^{m-1} \varphi'_{(i)}(\tilde{\xi}, \tilde{B}) \big(\tilde{B}^{t_{i+1}} - \tilde{B}^{t_i}\big) = \Phi(\tilde{\xi}, \tilde{B}),\end{aligned}$$

where the first equality is due to the elementary form (C.102) of φ (see again [53, Remark 5.101]), the second holds since $\varphi_{(i)}(\tilde{\xi}, \tilde{B}) = \varphi'_{(i)}(\tilde{\xi}, \tilde{B})$ for all $i \in \{0, \dots, m-1\}$ away from the $\tilde{\mathbb{P}}$-null-set $\{(\tilde{\xi}, \tilde{B}) \in N\}$, see (C.105), and the third is due to the definition (C.104) of Φ.

Step 2 (Lemma C.54 holds for all $\varphi \in L^2(B)$). Assume that $\varphi \in L^2(B)$, i.e. (see Definition C.50) $\mathbb{E}[\int_0^\infty |\varphi_s|^2 \, \mathrm{d}s] < \infty$. By Lemma C.49, there exists a sequence $(\varphi^k)_{k \in \mathbb{N}}$ of $\overline{\mathbb{F}}$-elementary processes (Definition C.45) on $(\Omega, \mathcal{F}, \mathbb{P})$ with

$$\|\varphi^k - \varphi\|_{L^2(B)} \xrightarrow[k \to \infty]{} 0. \tag{C.106}$$

[36] In the terminology of [53], the elementary processes from Definition C.45 are a subset of the $\mathbb{F}$-predictable step processes ([53, Definition 5.1]), whose integral is defined in [53, Definition 5.7].

By (C.94) in Remark C.51, this implies

$$\sup_{t\in\mathbb{R}_+}\left|\int_0^t(\varphi_s^k-\varphi_s)\,\mathrm{d}B_s\right| \xrightarrow[k\to\infty]{\mathbb{P}} 0,$$

and by passing to a subsequence, we can assume that $(\varphi^k\cdot B).(\omega)\to(\varphi\cdot B).(\omega)$ uniformly on $\mathbb{R}_+$, for all $\omega\in\Omega$ away from a $\mathbb{P}$-null set $N\in\mathcal{F}$. By Step 1, using that $\mathcal{F}=\mathcal{B}_{\mathbb{K}^n}\otimes\mathcal{G}^d$ by (C.96), there exists for each $k\in\mathbb{N}$ an $\mathcal{F}$-$\mathcal{G}^1$ measurable map $\Phi_k\colon\mathbb{K}^n\times W_d\to W_1$ with

$$\int_0^\cdot\varphi_s^k(\xi,B)\,\mathrm{d}B_s\overset{\text{uti}}{=}\Phi_k(\xi,B),$$

and which also satisfies (C.100). Viewing $\varphi^k\cdot B$ and Φ_k as $\mathcal{F}$-measurable, W_1-valued random variables for each $k\in\mathbb{N}$, let $N':=\bigcup_{k\in\mathbb{N}}\{\varphi^k\cdot B\neq\Phi_k\}$, which is an $(\mathcal{F},\mathbb{P})$-null set.[37] Furthermore, $(\mathbb{1}_{(N\cup N')^\mathrm{c}}\,\Phi_k)_{k\in\mathbb{N}}$ is a sequence of $\mathcal{F}$-$\mathcal{G}^1$-measurable maps which converges pointwise on Ω w.r.t. the topology of uniform convergence, and thus also w.r.t. the topology $\mathcal{O}_\text{ucc}$ of compact convergence in W_1. Since $\mathcal{O}_\text{ucc}$ is metrizable (see (2.2)) and generates $\mathcal{G}^1$ by Lemma 2.1(d),

$$\Phi:=\operatorname*{ucc-lim}_{k\to\infty}\mathbb{1}_{(N\cup N')^\mathrm{c}}\,\Phi_k \tag{C.107}$$

is not only well defined, but also $\mathcal{F}$-$\mathcal{G}^1$-measurable by Lemma B.1. In addition,

$$\begin{aligned}\Phi(\xi,B)&=\operatorname*{ucc-lim}_{k\to\infty}\mathbb{1}_{(N\cup N')^\mathrm{c}}(\xi,B)\ \Phi_k(\xi,B)\\&\overset{\text{uti}}{=}\operatorname*{ucc-lim}_{k\to\infty}\mathbb{1}_{N^\mathrm{c}}(\xi,B)\int_0^\cdot\varphi_s^k(\xi,B)\,\mathrm{d}B_s\overset{\text{uti}}{=}\int_0^\cdot\varphi_s(\xi,B)\,\mathrm{d}B_s\,,\end{aligned}$$

where the last equality is due to the uniform convergence of $\varphi^k\cdot B$ to $\varphi\cdot B$ away from the null set N. This establishes the existence of a map Φ satisfying (C.99) and it remains to verify (C.100). Since $\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{F}$-measurability of φ implies that $\Omega=\mathbb{K}^n\times W_d\ni(x,w)\mapsto\int_0^\infty|\varphi_s(x,w)|^2\,\mathrm{d}s$ is $\mathcal{F}$-measurable, it follows using (C.90) and (C.97) that

$$\|\tilde\varphi\|^2_{L^2(\tilde B)}=\tilde{\mathbb{E}}\left[\int_0^\infty|\varphi_s(\tilde\xi,\tilde B)|^2\,\mathrm{d}s\right]=\mathbb{E}\left[\int_0^\infty|\varphi_s(\xi,B)|^2\,\mathrm{d}s\right]=\|\varphi\|^2_{L^2(B)}<\infty. \tag{C.108}$$

By Lemma C.53(b)

$$\tilde\varphi\colon\mathbb{R}_+\times\tilde\Omega\mapsto\mathbb{K}^d,\quad\tilde\varphi_s(\tilde\omega):=\varphi_s(\tilde\xi(\tilde\omega),\tilde B(\tilde\omega)),$$

is $\tilde{\mathbb{F}}$-progressive, which together with (C.108) and Definition C.50 implies $\tilde\varphi\in L^2(\tilde B)$. For each $k\in\mathbb{N}$ define $\tilde\varphi^k$ as the composition of the elementary process φ^k with $(\tilde\xi,\tilde B)$ in the second argument. Then $\tilde\varphi^k\in L^2(\tilde B)$ for all $k\in\mathbb{N}$, and the same arguments as those leading to (C.108) show that (C.106) implies

$$\|\tilde\varphi-\tilde\varphi^k\|_{L^2(\tilde B)}=\|\varphi-\varphi^k\|_{L^2(B)}\xrightarrow[k\to\infty]{}0.$$

[37] Let $X=(X_t)_{t\in\mathbb{R}_+}$ and $Y=(Y_t)_{t\in\mathbb{R}_+}$ be continuous stochastic processes on a probability space $(\Omega,\mathcal{F},\mathbb{P})$. Since the σ-algebra $\mathcal{G}^1$ on W_1 is generated by the canonical projections, we can view X and Y as random variables taking values in $(W_1,\mathcal{G}^1)$, and $\{X\neq Y\}=\bigcup_{t\in\mathbb{R}_+\cap\mathbb{Q}}\{X_t\neq Y_t\}\in\mathcal{F}$ by continuity of sample paths.

Using (C.94) from Remark C.51 for $(\tilde{B}, \tilde{\mathbb{P}})$ instead of $(B, \mathbb{P})$ and passing once more to a subsequence, we can assume that $(\tilde{\varphi}^k \cdot \tilde{B}).(\tilde{\omega}) \xrightarrow[k\to\infty]{} (\tilde{\varphi} \cdot \tilde{B}).(\tilde{\omega})$ uniformly on $\mathbb{R}_+$, away from a $\tilde{\mathbb{P}}$-null set $\tilde{N} \in \tilde{\mathcal{F}}_\infty$. By construction of the maps Φ_k,

$$\tilde{\varphi}^k \cdot \tilde{B} \overset{\text{uti}}{=} \Phi_k(\tilde{\xi}, \tilde{B}) \qquad \text{w.r.t. } \tilde{\mathbb{P}}\,, \quad k \in \mathbb{N},$$

hence $\hat{N} := \bigcup_{k\in\mathbb{N}}\{\tilde{\varphi}^k \cdot \tilde{B} \neq \Phi_k(\tilde{\xi}, \tilde{B})\}$ is an $(\tilde{\mathcal{F}}, \tilde{\mathbb{P}})$-null set. By (C.107), $\mathbb{1}_{(N\cup N')^{\mathsf{c}}}\,\Phi_k \to \Phi$ pointwise on $\Omega = \mathbb{K}^n \times W_d$ w.r.t. the topology of compact convergence in W_1. It follows that w.r.t. $\tilde{\mathbb{P}}$,

$$\begin{aligned}\int_0^{\cdot} \varphi_s(\tilde{\xi}, \tilde{B})\,\mathrm{d}\tilde{B}_s &\overset{\text{uti}}{=} \underset{k\to\infty}{\text{ucc-lim}}\, \mathbb{1}_{\tilde{N}^{\mathsf{c}}} \int_0^{\cdot} \tilde{\varphi}^k_s\,\mathrm{d}\tilde{B}_s \\ &\overset{\text{uti}}{=} \underset{k\to\infty}{\text{ucc-lim}}\, \mathbb{1}_{(\tilde{N}\cup\hat{N})^{\mathsf{c}}}\,\Phi_k(\tilde{\xi}, \tilde{B}) \\ &\overset{\text{uti}}{=} \Phi(\tilde{\xi}, \tilde{B}),\end{aligned}$$

where the first equality fails[38] at most on $\tilde{N}$, the second fails at most on $\hat{N}$, and the third fails at most on $\{(\tilde{\xi}, \tilde{B}) \in N \cup N'\}$, which is also a $\tilde{\mathbb{P}}$-null set by (C.97). Hence Φ satisfies (C.100).

Step 3 (Lemma C.54 holds for every progressive process φ satisfying (C.98)). Fix a progressive process φ satisfying (C.98) and define

$$\tau_k := \inf\left\{t \in \mathbb{R}_+ : \int_0^t |\varphi_s|^2\,\mathrm{d}s \geq k\right\}, \qquad k \in \mathbb{N}.$$

The process $(t, \omega) \mapsto \int_0^t |\varphi_s(\omega)|^2\,\mathrm{d}s$ is $\overline{\mathbb{F}}$-adapted, $\mathbb{R}_+$-valued by (C.98) and in addition continuous. Hence $(\tau_k)_{k\in\mathbb{N}}$ is a sequence of $\overline{\mathbb{F}}$-stopping times by [53, Lemma 3.53], and in fact $\tau_k \nearrow \infty$ as $k \to \infty$ by continuity. Since $(s, \omega) \mapsto \mathbb{1}_{s\leq\tau_k(\omega)}$ is $\overline{\mathbb{F}}$-adapted, left-continuous and thus $\overline{\mathbb{F}}$-progressive for each $k \in \mathbb{N}$,

$$\varphi^k_s(\omega) := \mathbb{1}_{s\leq\tau_k(\omega)}\,\varphi_s(\omega), \qquad (s, \omega) \in \mathbb{R}_+ \times \Omega,\ k \in \mathbb{N}. \tag{C.109}$$

defines a sequence of $\overline{\mathbb{F}}$-progressive processes. For every $k \in \mathbb{N}$, $\int_0^\infty |\varphi^k_s|^2\,\mathrm{d}s \leq k$ and thus $\varphi^k \in L^2(B)$. Since the stochastic integral commutes with stopping (see [53, Lemma 5.122]), $\varphi^k \cdot B \overset{\text{uti}}{=} (\varphi \cdot B)^{\tau_k}$ for all $k \in \mathbb{N}$, hence $N := \bigcup_{k\in\mathbb{N}}\{\varphi^k \cdot B \neq (\varphi \cdot B)^{\tau_k}\}$ is an $(\mathcal{F}, \mathbb{P})$-null set. Then $\tau_k \nearrow \infty$ implies that

$$\int_0^{\cdot} \varphi^k_s\,\mathrm{d}B_s \xrightarrow[k\to\infty]{\text{ucc}} \int_0^{\cdot} \varphi_s\,\mathrm{d}B_s \qquad \text{on } N^{\mathsf{c}}. \tag{C.110}$$

By Step 2, there exists for each $k \in \mathbb{N}$ an $\mathcal{F}$-$\mathcal{G}^1$ measurable map $\Phi_k \colon \mathbb{K}^n \times W_d \to W_1$ with

$$\int_0^{\cdot} \varphi^k_s\,\mathrm{d}B_s \overset{\text{uti}}{=} \Phi_k, \tag{C.111}$$

[38] In fact, $\int_0^{\cdot} \tilde{\varphi}^k_s\,\mathrm{d}\tilde{B}_s$ may not converge at all on $\tilde{N}$ as $k \to \infty$.

and which also satisfies (C.100) for φ^k and Φ_k instead of φ and Φ, i.e.

$$\int_0^{\cdot} \varphi^k_s(\tilde{\xi},\tilde{B})\,\mathrm{d}\tilde{B}_s \overset{\text{uti}}{=} \Phi_k(\tilde{\xi},\tilde{B}), \qquad k\in\mathbb{N}. \tag{C.112}$$

Let $N' := \bigcup_{k\in\mathbb{N}}\{\varphi^k\cdot B\neq\Phi_k\}$, which is a $\mathbb{P}$-null set in[39] $\mathcal{F}$. From (C.110) and (C.111), it follows that the sequence $(\Phi_k\,\mathbb{1}_{(N\cup N')^{\mathsf{c}}})_{k\in\mathbb{N}}$ of $\mathcal{F}$-measurable functions converges pointwise on Ω w.r.t. the topology $\mathcal{O}_{\mathrm{ucc}}$ of uniform convergence on W_1. We redefine Φ_k to equal $\Phi_k\,\mathbb{1}_{(N\cup N')^{\mathsf{c}}}$ for each $k\in\mathbb{N}$, which preserves (C.111) and, since $\{(\tilde{\xi},\tilde{B})\in N\cup N'\}$ is a $\tilde{\mathbb{P}}$-null set by Lemma C.53(a), also (C.112). Let

$$\Phi := \underset{k\to\infty}{\text{ucc-lim}}\,\Phi_k. \tag{C.113}$$

By metrizability of $\mathcal{O}_{\mathrm{ucc}}$ (see (2.2)) and Lemma B.1, Φ is $\mathcal{F}$-$\mathcal{G}^1$-measurable. Combining the previous display with (C.111) and (C.110),

$$\begin{aligned}\Phi(\xi,B) &= \underset{k\to\infty}{\text{ucc-lim}}\,\Phi_k(\xi,B)\\ &\overset{\text{uti}}{=} \underset{k\to\infty}{\text{ucc-lim}}\,\mathbb{1}_{N^{\mathsf{c}}}(\xi,B)\int_0^{\cdot}\varphi^k_s(\xi,B)\,\mathrm{d}B_s \overset{\text{uti}}{=} \int_0^{\cdot}\varphi_s(\xi,B)\,\mathrm{d}B_s\,.\end{aligned}$$

Hence Φ satisfies (C.99), and it remains to verify (C.100). Let

$$\tilde{\tau}_k := \tau_k\circ(\tilde{\xi},\tilde{B}) = \inf\left\{t\in\mathbb{R}_+ : \int_0^t |\varphi_s(\tilde{\xi},\tilde{B})|^2\,\mathrm{d}s \ge k\right\}, \qquad k\in\mathbb{N}. \tag{C.114}$$

Since τ_k is an $\overline{\mathbb{F}}$-stopping time for each $k\in\mathbb{N}$ and $\tilde{\Omega}\ni\tilde{\omega}\mapsto(\tilde{\xi}(\tilde{\omega}),\tilde{B}(\tilde{\omega}))\in\Omega$ is $\tilde{\mathcal{F}}_t$-$\bar{\mathcal{F}}_t$-measurable for all $t\in\mathbb{R}_+$ by Lemma C.53(a), $(\tilde{\tau}_k)_{k\in\mathbb{N}}$ is a sequence of $\tilde{\mathbb{F}}$-stopping times. Moreover since $\tau_k\nearrow\infty$, we have $\tilde{\tau}_k\nearrow\infty$ as $k\to\infty$. Let

$$\begin{aligned}&\tilde{\varphi}\colon \mathbb{R}_+\times\tilde{\Omega}\mapsto\mathbb{K}^d, \quad \tilde{\varphi}_s(\tilde{\omega}) := \varphi_s(\tilde{\xi}(\tilde{\omega}),\tilde{B}(\tilde{\omega})), \qquad (s,\tilde{\omega})\in\mathbb{R}_+\times\tilde{\Omega},\\ &\tilde{\varphi}^k\colon \mathbb{R}_+\times\tilde{\Omega}\mapsto\mathbb{K}^d, \quad \tilde{\varphi}^k_s(\tilde{\omega}) := \varphi^k_s(\tilde{\xi}(\tilde{\omega}),\tilde{B}(\tilde{\omega})), \qquad (s,\tilde{\omega})\in\mathbb{R}_+\times\tilde{\Omega},\ k\in\mathbb{N}.\end{aligned}$$

By the previous display, (C.109) and (C.114),

$$\tilde{\varphi}^k_s(\tilde{\omega}) = \mathbb{1}_{s\le\tilde{\tau}_k(\tilde{\omega})}\,\tilde{\varphi}_s(\tilde{\omega}), \qquad (s,\tilde{\omega})\in\mathbb{R}_+\times\tilde{\Omega},\ k\in\mathbb{N}.$$

Since each $\tilde{\tau}_k$ is an $\tilde{\mathbb{F}}$-stopping time, [53, Lemma 5.122] implies that $\tilde{\varphi}^k\cdot\tilde{B}\overset{\text{uti}}{=}(\tilde{\varphi}\cdot\tilde{B})^{\tilde{\tau}_k}$ for all $k\in\mathbb{N}$. Let $\tilde{N} := \bigcup_{k\in\mathbb{N}}\{\tilde{\varphi}^k\cdot\tilde{B}\neq(\tilde{\varphi}\cdot\tilde{B})^{\tilde{\tau}_k}\}$ which is a $\tilde{\mathbb{P}}$-null set in $\tilde{\mathcal{F}}_\infty$. Since $\tilde{\tau}_k\nearrow\infty$,

$$\int_0^{\cdot}\tilde{\varphi}^k_s\,\mathrm{d}\tilde{B}_s \xrightarrow[k\to\infty]{\text{ucc}} \int_0^{\cdot}\tilde{\varphi}_s\,\mathrm{d}\tilde{B}_s \qquad \text{on } \tilde{N}^{\mathsf{c}}. \tag{C.115}$$

It follows that

$$\Phi(\tilde{\xi},\tilde{B}) \underset{(C.113)}{=} \underset{k\to\infty}{\text{ucc-lim}}\,\Phi_k(\tilde{\xi},\tilde{B}) \underset{(C.112)}{\overset{\text{uti}}{=}} \underset{k\to\infty}{\text{ucc-lim}}\,\mathbb{1}_{\tilde{N}^{\mathsf{c}}}\int_0^{\cdot}\tilde{\varphi}^k_s\,\mathrm{d}\tilde{B}_s \underset{(C.115)}{\overset{\text{uti}}{=}} \int_0^{\cdot}\tilde{\varphi}_s\,\mathrm{d}\tilde{B}_s\,,$$

i.e. Φ satisfies (C.100). □

[39] As before, measurability of N' follows from continuity of $\varphi^k\cdot B$, $k\in\mathbb{N}$ and Φ_k when viewed as stochastic processes, see Footnote 37 on Page 153 for details.

Theorem C.56 ([51, Theorem 10.4]). *Let $\mu \in \mathcal{P}(\mathbb{K}^n)$ be an initial distribution and assume that on the canonical set-up $(\Omega, \mathcal{F}, \overline{\mathbb{F}}, \mathbb{P}, \xi, B)$ with initial distribution μ (Definition C.52), the SDE (C.6) has a strong solution*[40]

$$X \overset{\text{uti}}{=} \xi + \int_0^{\cdot} b(s, X)\,\mathrm{d}s + \int_0^{\cdot} \sigma(s, X)\,\mathrm{d}B_s \tag{C.116}$$

in the sense of Definition C.15.

(a) There exists a map $F_\mu \colon \mathbb{K}^n \times W_d \to W_n$ with

$$F_\mu^{-1}(\mathcal{G}_t^n) \subseteq \bar{\mathcal{F}}_t, \quad t \in \mathbb{R}_+, \tag{C.117}$$

and such that

$$X = F_\mu(\xi, B). \tag{C.118}$$

(b) On every set-up $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}}, \tilde{\xi}, \tilde{B})$ where $\tilde{\xi}$ has law μ, the process $\tilde{X} := F_\mu(\tilde{\xi}, \tilde{B})$ is a strong solution (Definition C.15) of (C.6). In particular,

$$\tilde{X} \overset{\text{uti}}{=} \tilde{\xi} + \int_0^{\cdot} b(s, \tilde{X})\,\mathrm{d}s + \int_0^{\cdot} \sigma(s, \tilde{X})\,\mathrm{d}\tilde{B}_s \quad \textit{w.r.t. } \tilde{\mathbb{P}}. \tag{C.119}$$

Proof. (a) Viewing X as a map from $\Omega = \mathbb{K}^n \times W_d$ to W_n, let

$$F_\mu := X. \tag{C.120}$$

Since (ξ, B) is the identity map on Ω, (C.118) holds trivially. Moreover, since X is $\overline{\mathbb{F}}$-adapted and the filtration $\mathbb{G}^n = (\mathcal{G}_t^n)_{t \in \mathbb{R}_+}$ on $(W_n, \mathcal{G}^n)$ is generated by the canonical projections on W_n (see (2.5)), F_μ is $\bar{\mathcal{F}}_t$-$\mathcal{G}_t^n$-measurable for each $t \in \mathbb{R}_+$, which shows (C.117).

(b) Fix an alternative set-up $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}}, \tilde{\xi}, \tilde{B})$ with $\mathcal{L}_{\tilde{\mathbb{P}}}[\tilde{\xi}] = \mu$ and let

$$\tilde{X} := F_\mu(\tilde{\xi}, \tilde{B}). \tag{C.121}$$

We verify that $\tilde{X}$ is continuous, $\tilde{\mathbb{F}}$-adapted and has satisfies parts (a) to (c) from Definition C.15. Clearly, $\tilde{X}$ has continuous sample paths. By Lemma C.53 (a) and (C.117), $\tilde{X} = F_\mu \circ (\tilde{\xi}, \tilde{B})$ is $\tilde{\mathcal{F}}_t$-$\mathcal{G}_t^n$-measurable for each $t \in \mathbb{R}_+$, which implies that $\tilde{X}$ is $\tilde{\mathbb{F}}$-adapted. From (C.116), $N := \{F_\mu(\xi, B)_0 \neq \xi\} = \{X_0 \neq \xi\}$ is an $(\mathcal{F}_0, \mathbb{P})$-null set. By (C.97) in Lemma C.53 (a), $\tilde{\mathbb{P}}[(\tilde{\xi}, \tilde{B}) \in N] = 0$, hence

$$\tilde{X}_0 \overset{\text{a.s.}}{=} F_\mu(\tilde{\xi}, \tilde{B}) = \tilde{\xi}, \qquad \text{w.r.t. } \tilde{\mathbb{P}}.$$

Since (ξ, B) is the identity on $\Omega = \mathbb{K}^n \times W_d$, the integrability condition (C.12) for $X = F_\mu(\xi, B)$ implies that

$$\int_0^t |b(s, F_\mu(x, w))| + |\sigma(s, F_\mu(x, w))|^2\,\mathrm{d}s < \infty,$$

[40] The concept of a strong solution in the sense of Definition C.15 differs from [51], where a strong solution of (C.6) is defined as a function $F \colon \mathbb{K}^n \times W_d \to W_n$ as given in Theorem C.56.

for all $(t, x, w) \in \mathbb{R}_+ \times \mathbb{K}^n \times W_d$. With (C.121), it follows that

$$\int_0^t |b(s, \tilde{X})| + |\sigma(s, \tilde{X})|^2 \,\mathrm{d}s < \infty, \quad \text{for all } t \in \mathbb{R}_+.$$

Hence the stochastic integrals on the right-hand side of (C.119) are well-defined. To show that $\tilde{X}$ is a strong solution of (C.6), it remains by Definition C.15(c) only to establish (C.119). Define

$$\varphi^i\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^d, \quad \varphi^i_s(\omega) := \sigma^i(s, F_\mu(\omega)), \qquad i \in \{1, \dots, n\}. \tag{C.122}$$

By (C.120), $\mathbb{R}_+ \times \Omega \ni (s, \omega) \mapsto F_\mu(\omega)_s = X_s(\omega) \in \mathbb{K}^n$ is $\overline{\mathbb{F}}$-progressive. By Definition C.8, $\mathbb{R}_+ \times W_n \ni (s, w) \mapsto \sigma(s, w)$ is $\mathbb{G}^n$-progressive, hence φ^i is $\overline{\mathbb{F}}$-progressive for each $i \in \{1, \dots, n\}$ by Lemma C.7(b). Since X is a strong solution on the canonical set-up, σ satisfies the pathwise square integrability condition (C.12), hence

$$\int_0^t |\varphi^i_s(\xi, B)|^2 \,\mathrm{d}s = \int_0^t |\sigma^i(s, X)|^2 \,\mathrm{d}s < \infty, \qquad t \in \mathbb{R}_+,\ i \in \{1, \dots, n\}.$$

By Lemma C.54, using that $\mathcal{F}_\infty = \mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{G}^d$, see (C.96), there exists for each $i \in \{1, \dots, n\}$ an $\mathcal{F}_\infty$-$\mathcal{G}^1$-measurable map $\Phi^i\colon \mathbb{K}^n \times W_d \to W_1$ with

$$\int_0^\cdot \varphi^i_s(\xi, B) \,\mathrm{d}B_s \overset{\text{uti}}{=} \Phi^i(\xi, B), \quad \text{w.r.t. } \mathbb{P}, \quad \text{and} \tag{C.123}$$

$$\int_0^\cdot \varphi^i_s(\tilde{\xi}, \tilde{B}) \,\mathrm{d}\tilde{B}_s \overset{\text{uti}}{=} \Phi^i(\tilde{\xi}, \tilde{B}), \quad \text{w.r.t. } \tilde{\mathbb{P}}. \tag{C.124}$$

Let $\Phi := (\Phi^1, \dots, \Phi^n)$, which is $\mathcal{F}_\infty$-$\mathcal{G}^n$-measurable by $\mathcal{F}_\infty$-$\mathcal{G}^1$ measurability of each component map. By (C.124), (C.122) and (C.121)

$$\Phi(\tilde{\xi}, \tilde{B}) \overset{\text{uti}}{=} \int_0^\cdot \sigma(s, \tilde{X}) \,\mathrm{d}\tilde{B}_s\,, \quad \text{w.r.t. } \tilde{\mathbb{P}}. \tag{C.125}$$

Let $G\colon \mathbb{R}_+ \times \Omega \to \mathbb{K}^n$ be defined by

$$G_t(x, w) := F_\mu(x, w)_t - x - \Phi(x, w)_t - \int_0^t b(s, F_\mu(x, w)) \,\mathrm{d}s\,, \tag{C.126}$$

for all $(t, x, w) \in \mathbb{R}_+ \times \mathbb{K}^n \times W_d$, and observe that by (C.120), (C.122) and (C.123) for the first and (C.116) for the second equality,

$$G_\cdot(\xi, B) \overset{\text{uti}}{=} X - \xi - \int_0^\cdot \sigma(s, X) \,\mathrm{d}B_s - \int_0^\cdot b(s, X) \,\mathrm{d}s \overset{\text{uti}}{=} 0 \quad \text{w.r.t. } \mathbb{P}\,. \tag{C.127}$$

Let $N \in \mathcal{F}$ be a $\mathbb{P}$-null set outside of which all equalities in (C.127) hold. From (C.97), $\tilde{\mathbb{P}}[(\tilde{\xi}, \tilde{B}) \in N] = 0$ and thus

$$G_\cdot(\tilde{\xi}, \tilde{B}) \overset{\text{uti}}{=} 0 \quad \text{w.r.t. } \tilde{\mathbb{P}}. \tag{C.128}$$

Using (C.121) and (C.125), (C.119) follows. □

As a corollary, we obtain weak uniqueness of solutions of (C.6) provided the coefficients are such that pathwise uniqueness holds and that there exists a strong solution on the canonical set-up. In fact, by the Yamada–Watanabe Theorem, weak existence together with pathwise uniqueness implies weak uniqueness, see [64, Proposition 1] for the original result, which applies to the setting of instantaneous and time-homogeneous dependence (Definition C.9), and [62, Theorem 12.4.4] for a proof in the pathwise setting.

Corollary C.57 (Uniqueness in law for classical SDEs with nonrandom dependence)**.** *Consider $\mathbb{G}^n$-progressive coefficients b and σ as in* (C.4)*. If there is pathwise uniqueness of solutions of the SDE* (C.6) *for the initial distribution $\mu \in \mathcal{P}(\mathbb{K}^n)$ and there exists a strong solution on the canonical set-up*[41] *with initial distribution μ (Definition C.52), then there is weak uniqueness (Definition C.21) of solutions.*

Proof. Let X be a strong solution of (C.6) on the canonical setup $(\Omega, \mathcal{F}, \overline{\mathbb{F}}, \mathbb{P}, \xi, B)$ with initial distribution μ and $\tilde{Y}$ a strong solution of (C.6) on a given alternative set-up $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}}, \tilde{\xi}, \tilde{B})$ with $\mathcal{L}_{\tilde{\mathbb{P}}}[\tilde{\xi}] = \mu$. By Theorem C.56, there exists an $\mathcal{F}_\infty$-$\mathcal{G}^n$-measurable map $F_\mu \colon \mathbb{K}^n \times W_d \to W_n$ with the property that $X \overset{\text{uti}}{=} F_\mu(\xi, B)$ w.r.t. $\mathbb{P}$ and $\tilde{X} := F_\mu(\tilde{\xi}, \tilde{B})$ is a strong solution of (C.6) on the alternative set-up. Due to (C.97) in Lemma C.53(a), $\mathcal{L}_{\mathbb{P}}[X] = \mathcal{L}_{\tilde{\mathbb{P}}}[\tilde{X}]$. By pathwise uniqueness, $\tilde{Y} \overset{\text{uti}}{=} \tilde{X}$ w.r.t. $\tilde{\mathbb{P}}$, hence

$$\mathcal{L}_{\tilde{\mathbb{P}}}[\tilde{Y}] = \mathcal{L}_{\tilde{\mathbb{P}}}[\tilde{X}] = \mathcal{L}_{\mathbb{P}}[X] = \mathbb{P} \circ F_\mu^{-1}$$

follows. □

The proof of Theorem C.43 below is based on the proofs of [38, Theorem 8.5] and [51, Theorem 13.1].

Proof of Theorem C.43. Each of the properties (i) to (iv) is shown in a separate step. In Step 1, the map F is constructed as a continuous modification of a family of solutions using Lemma C.31 and the Kolmogorov–Chentsov continuity criterion. In Step 2, the measurability property (ii) is established using that adaptedness w.r.t. a filtration satisfying the usual completeness assumption (Definition C.3) survives under passage to a modification. In Step 3, Theorem C.56 is used to show that the map F can be used to obtain strong solutions for deterministic initial conditions, i.e. property (iii). This is extended to arbitrary initial conditions (property (iv)) in Step 4 using an approximation argument based on the continuity of F in the first argument.

Step 1 (Construction of the map F, property (i)). Let $p = 1 + \dim_{\mathbb{R}}(\mathbb{K}^n)$ and fix the stochastic base $(W_d, \mathcal{G}^d, \overline{\mathbb{G}}^d, \mathbb{W})$ described at the beginning of Appendix C.4. On this base, the canonical process $\Pi = (\Pi_t)_{t \in \mathbb{R}_+}$ is a d-dimensional standard $(\overline{\mathbb{G}}^d, \mathbb{W})$-Brownian motion[42], and the constant maps $W_d \ni w \mapsto x \in \mathbb{K}^n$, $x \in \mathbb{K}^n$ are $\mathcal{G}_0$-measurable random vectors. Hence for every $x \in \mathbb{K}^n$, there exists by Theorem C.34 a solution X^x with initial condition $X_0^x \overset{\text{a.s.}}{=} x$ of the SDE (C.6) with Π as the driving Brownian motion. By Lemma C.31 (b),

$$\mathbb{E}[p_t^p(X^x - X^y)] \le 6^{p-1}|x-y|^p \exp\left(t(\beta(t) \vee c_t^p)^2\right), \qquad t \in \mathbb{R}_+,\ x, y \in \mathbb{K}^n, \qquad \text{(C.129)}$$

[41] This is the case in particular if the Lipschitz and boundedness conditions of Theorem C.34 are satisfied

[42] The fact that a standard Brownian motion is defined as a Brownian motion having initial value almost surely equal to zero rather than identically equal to zero (see [53, Definition 2.66]) is crucial here.

with $c\colon \mathbb{R}_+ \to \mathbb{R}_+$ from Definition C.25 and β defined in (C.22). Let $(\alpha_k)_{k\in\mathbb{N}}$ be a summable sequence in $\mathbb{R}_+^*$ with

$$\alpha_k \le k^{-2}\left(6^{p-1}\exp(k(\beta(k)\vee c_k^p)^2)\right)^{-1}, \qquad k\in\mathbb{N}, \tag{C.130}$$

and let $\hat{\varrho}_{\mathrm{ucc}}$ be the corresponding metric for compact convergence in W_n defined by (C.79). Dropping the minimum with the constant 1 in (C.79), using Jensen's inequality for the function $\mathbb{R}_+ \ni x \mapsto x^p$ and the probability measure $(\sum_{k\in\mathbb{N}}\alpha_k)^{-1}\sum_{k\in\mathbb{N}}\alpha_k\delta_k$ on $(\mathbb{N},\mathfrak{P}(\mathbb{N}))$ and finally applying the monotone convergence to exchange sum and expectation, it follows that for all $x,y\in\mathbb{K}^n$,

$$\begin{aligned}\mathbb{E}[\hat{\varrho}_{\mathrm{ucc}}(X^x,X^y)^p] &\le \mathbb{E}\left[\left(\sum_{k\in\mathbb{N}}\alpha_k\,p_k(X^x-X^y)\right)^p\right]\\ &\le \left(\sum_{k\in\mathbb{N}}\alpha_k\right)^{p-1}\sum_{k\in\mathbb{N}}\alpha_k\,\mathbb{E}[p_k^p(X^x-X^y)].\end{aligned} \tag{C.131}$$

Combining (C.131) with the bounds (C.129) and (C.130), it follows that there exists a constant $C\in\mathbb{R}_+$ such that

$$\mathbb{E}[\hat{\varrho}_{\mathrm{ucc}}(X^x,X^y)^p] \le C|x-y|^p, \qquad x,y\in\mathbb{K}^n. \tag{C.132}$$

By the Kolmogorov–Chentsov continuity criterion [53, Theorem 2.108] with $T=\mathbb{R}^n$ if $\mathbb{K}=\mathbb{R}$ and $T=\mathbb{R}^{2n}\cong\mathbb{C}^n$ if $\mathbb{K}=\mathbb{C}$, fixing $a=p$, $b=1$ and $d'=\dim_{\mathbb{R}}(\mathbb{K}^n)$ (such that $b+d'=1+\dim_{\mathbb{R}}(\mathbb{K}^n)=p$), the stochastic process $(X^x)_{x\in\mathbb{K}^n}$ taking values in the Polish space $(W_n,\hat{\varrho}_{\mathrm{ucc}})$ has a modification $(\tilde{X}^x)_{x\in\mathbb{K}^n}$ with continuous[43] sample paths. Let

$$F_x(w) \coloneqq \tilde{X}^x(w) = (\tilde{X}_t^x(w))_{t\in\mathbb{R}_+}, \qquad x\in\mathbb{K}^n,\ w\in W_d. \tag{C.133}$$

Property (i) then follows from continuity of $x\mapsto\tilde{X}^x$ w.r.t. the metric $\hat{\varrho}_{\mathrm{ucc}}$.

Step 2 (Property (ii)). By construction, $(\tilde{X}^x)_{x\in\mathbb{K}^n}$ is a modification of $(X^x)_{x\in\mathbb{K}^n}$. Hence for each $x\in\mathbb{K}^n$, $\tilde{X}^x$ is $\mathcal{G}^d$-$\mathcal{G}^n$-measurable and

$$\tilde{X}^x = X^x, \qquad \mathbb{W}\text{-a.s.} \tag{C.134}$$

Fixing $x\in\mathbb{K}^n$, it follows that $(t,w)\mapsto\tilde{X}_t^x(w)=F_x(w)_t$ is a $\mathcal{G}^d$-measurable stochastic process, which is indistinguishable from X^x w.r.t. $\mathbb{W}$. Since X^x is $\overline{\mathbb{G}}^d$-adapted and the filtration $\overline{\mathbb{G}}^d$ satisfies the usual completeness assumption (Definition C.3), $(t,w)\mapsto F_x(w)_t$ is $\overline{\mathbb{G}}^d$-adapted by Corollary B.6 (b). This adaptedness property is equivalent to (C.81), since by (2.5), $\mathbb{G}^n$ is generated by the canonical projections.

Step 3 (Property (iii)). Fix again $x\in\mathbb{K}^n$ and let $F_{\delta_x}\colon\mathbb{K}^n\times W_d\to W_n$ denote the map from Theorem C.56 (a) associated with the SDE (C.6) and the initial distribution δ_x. By Theorem C.56 (b) applied to the set-up $(W_d,\mathcal{G}^d,\overline{\mathbb{G}}^d,\mathbb{W},\Pi,\xi)$ with $\xi(w)\coloneqq x$ (see also the beginning of

[43] In fact, the Kolmogorov–Chentsov continuity criterion yields that $x\mapsto\tilde{X}^x$ is locally uniformly Hölder continuous w.r.t. the given metric $\hat{\varrho}_{\mathrm{ucc}}$ on W_n for all exponents $c\in(0,(1+\dim_{\mathbb{R}}(\mathbb{K}^n))^{-1})$.

Step 1), $F_{\delta_x}(x,\Pi)$ is a strong solution of (C.6). By assumption, there is pathwise uniqueness of solutions, hence

$$F_{\delta_x}(x,\Pi) \overset{\text{uti}}{=} X^x \overset{\text{uti}}{=} F_x(\Pi) \quad \text{w.r.t. } \mathbb{W}, \tag{C.135}$$

where the second equality is due to (C.133) and (C.134). Let N be a $(\mathcal{G}^d, \mathbb{W})$-null set outside of which equality holds in (C.135). Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be an alternative stochastic base and $B = (B_t)_{t\in\mathbb{R}_+}$ a $\mathbb{K}^d$-valued standard $(\mathbb{F},\mathbb{P})$-Brownian motion. By Theorem C.56 (b), $\tilde{X} := F_{\delta_x}(x,B)$ solves the SDE (C.6) with initial condition $\tilde{X}_0 = x$, $\mathbb{P}$-a.s. Since $\mathbb{P}[B\in N] = \mathbb{W}[N] = 0$, (C.135) implies that

$$\tilde{X} = F_{\delta_x}(x,B) \overset{\text{uti}}{=} F_x(B) \quad \text{w.r.t. } \mathbb{P},$$

i.e. $F_x(B)$ is indistinguishable from the solution $\tilde{X}$. Since B is an $\mathbb{F}$-Brownian motion, $\Omega \ni \omega \mapsto B(\omega) \in W_d$ is $\mathcal{F}_t$-$\mathcal{G}_t^d$-measurable for each $t\in\mathbb{R}_+$. The filtration $\mathbb{F}$ satisfies the usual completeness assumption (Definition C.3), therefore arguments as in the proof of Lemma C.53 (a) show that $\omega \mapsto B(\omega)$ is in fact $\mathcal{F}_t$-$\overline{\mathcal{G}}_t^d$-measurable for all $t\in\mathbb{R}_+$. From (C.81), it follows that $F_x(B)$ is $\mathbb{F}$-adapted, which implies using Remark C.18 (c) that $F_x(B)$ is a strong solution of (C.6) satisfying the initial condition $F_x(B)_0 \overset{\text{a.s.}}{=} x$, thereby proving (iii).

Step 4 (Property (iv)). Fix a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$ and define

$$X := F_\xi(B). \tag{C.136}$$

We will show that X is a strong solution (C.6) with the initial condition $X_0 \overset{\text{a.s.}}{=} \xi$ by showing that it is the limit of a sequence of approximate solutions. Let

$$\mathbb{D}^n := \begin{cases} \{2^{-m}(a+b\mathrm{i}) : m\in\mathbb{N},\ a,b\in\mathbb{Z}^n\} & \text{if } \mathbb{K}=\mathbb{C},\\ \{2^{-m}a : m\in\mathbb{N},\ a\in\mathbb{Z}^n\} & \text{if } \mathbb{K}=\mathbb{R}.\end{cases}$$

For each $m\in\mathbb{N}$, let $\varphi_m\colon \mathbb{K}^n \to \mathbb{D}^n$ be a measurable map such that $|\varphi_m(x) - x| < 2^{-m}$ for all $x\in\mathbb{K}^n$. We first verify that

$$X^{(m)} := F_{\varphi_m(\xi)}(B)$$

solves (C.6) with the initial condition $X_0^{(m)} \overset{\text{a.s.}}{=} \varphi_m(\xi)$ for each $m\in\mathbb{N}$. Indeed, $X^{(m)} = \sum_{p\in\mathbb{D}^n} \mathbb{1}_{\varphi_m(\xi)=p} F_p(B)$ is $\mathbb{F}$-adapted with continuous sample paths since $F_p(B)$ has these properties for each $p\in\mathbb{D}^n$ by Step 3 and ξ is $\mathcal{F}_0$-measurable. By assumption, b and σ satisfy one of the boundedness assumptions from Theorem C.34, which ensures together with continuity of sample paths of $X^{(m)}$ that the integrability conditions (C.12) are satisfied for $X^{(m)}$. Moreover, using that $F_p(B)$ solves (C.6) with initial condition $F_p(B)_0 \overset{\text{a.s.}}{=} p$ for every $p\in\mathbb{D}^n$ by Step 3, we obtain

$$\begin{aligned} X^{(m)}\mathbb{1}_{\{\varphi_m(\xi)=p\}} &= \mathbb{1}_{\{\varphi_m(\xi)=p\}} F_p(B)\\ &\overset{\text{uti}}{=} \mathbb{1}_{\{\varphi_m(\xi)=p\}}\left(p + \int_0^\cdot b(s,F_p(B))\,\mathrm{d}s + \int_0^\cdot \sigma(s,F_p(B))\,\mathrm{d}B_s\right)\\ &= \mathbb{1}_{\{\varphi_m(\xi)=p\}}\left(\varphi_m(\xi) + \int_0^\cdot b(s,F_{\varphi_m(\xi)}(B))\,\mathrm{d}s + \int_0^\cdot \sigma(s,F_{\varphi_m(\xi)}(B))\,\mathrm{d}B_s\right)\\ &= \mathbb{1}_{\{\varphi_m(\xi)=p\}}\left(\varphi_m(\xi) + \int_0^\cdot b(s,X^{(m)})\,\mathrm{d}s + \int_0^\cdot \sigma(s,X^{(m)})\,\mathrm{d}B_s\right), \end{aligned}$$

for all $p \in \mathbb{D}^n$. Hence for every $m \in \mathbb{N}$, $X^{(m)}$ is a solution of (C.6) with initial condition $\varphi_m(\xi)$. By (i), $x \mapsto F_x(w)$ is continuous w.r.t. the topology $\mathcal{O}_{\text{ucc}}$ for every $w \in W_d$. Since $\varphi_m(\xi) \to \xi$ pointwise on Ω as $m \to \infty$, it follows using (C.136) that

$$X(\omega) = \operatorname*{ucc-lim}_{m\to\infty} X^{(m)}(\omega), \qquad \omega \in \Omega.$$

By metrizability of $\mathcal{O}_{\text{ucc}}$ and Lemma B.1, X is $\mathcal{F}_\infty$-measurable. By the assumptions on b and σ and Theorem C.34, the SDE (C.6) has a strong solution $\tilde{X}$ with initial condition $\tilde{X}_0 \overset{\text{a.s.}}{=} \xi$ on the given set-up. Hence to show that X is also a solution, it suffices by Remark C.18 (c) to show that $X \overset{\text{uti}}{=} \tilde{X}$. For fixed $t \in \mathbb{R}_+$, using (C.23), we obtain using that $|\xi - \varphi_m(\xi)| < 2^{-m}$ for all $m \in \mathbb{N}$ that

$$p_t^p(\tilde{X} - X^{(m)}) \xrightarrow[m\to\infty]{L^1} 0.$$

By passing to a subsequence, we may assume that

$$p_t(\tilde{X} - X^{(m)}) \xrightarrow[m\to\infty]{\text{a.s.}} 0.$$

Since $p_t(X - X^{(m)}) \xrightarrow[m\to\infty]{} 0$ by compact convergence of $X^{(m)}$ to X, it follows that $p_t(\tilde{X} - X) \overset{\text{a.s.}}{=} 0$. Since $t \in \mathbb{R}_+$ was arbitrary, $X \overset{\text{uti}}{=} \tilde{X}$ follows. □

The main ingredient in the proof of Theorem C.43 (iii) was Theorem C.56, whose proof required considerable preparatory effort. Under the assumption that $t \mapsto b(t, w)$ and $t \mapsto \sigma(t, w)$ are continuous for all $w \in W_n$ in addition to the usual Lipschitz property (Definition C.25),[44] Theorem C.43 (iii) and Theorem C.43 (iv) can be shown more directly by pathwise approximation arguments. A nice consequence of this proof is that on every set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$, $F_\xi(B)$ is the pathwise unique solution of (C.6), even if the initial random vector ξ is merely $\mathbb{K}^n$-valued and $\mathcal{F}_0$-measurable, and not necessarily in $\mathcal{L}^2(\mathbb{P})$, which slightly improves the existence statement from Theorem C.34. We state and prove this as a separate proposition.

Proposition C.58 (Existence and properties of the flow of SDEs with Lipschitz continuous and continuous in time coefficients [38, Theorem 8.5])**.** *Consider coefficients b and σ as in (C.4) which have the Lipschitz property (Definition C.25) and are continuous in the first argument, meaning that $t \mapsto b(t, w)$ and $t \mapsto \sigma(t, w)$ are continuous for all $w \in W_n$.*[45] *Then there exists a map*

$$F\colon \mathbb{K}^n \times W_d \to W_n, \quad (x, w) \mapsto F_x(w)$$

with the properties (i), (ii) and (iii) from Theorem C.43. In addition,

(iv') For every set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$ (Definition C.5 (b)), the process $X = (X_t)_{t\in\mathbb{R}_+}$ defined by

$$X_t = F_\xi(B)_t, \qquad t \in \mathbb{R}_+,$$

[44] In this case, the bounded growth condition (Definition C.27) is satisfied automatically by Remark C.28.

[45] In fact, continuity in the time argument is required only for σ and could be replaced by a linearly bounded growth condition as in Definition C.27 for the drift coefficient b.

is the up to indistinguishability unique strong solution of the SDE (C.6) *on the given set-up with the initial condition* $X_0 \overset{\text{a.s.}}{=} \xi$.[46]

Proof. [38, Theorem 8.5] By Remark C.28 continuity in the first argument together with Lipschitz continuity in the second argument implies that b and σ are of linearly bounded growth (Definition C.27). Therefore, parts (i) and (ii) follow exactly as in Step 1 and Step 2 of the proof of Theorem C.43, in the setting of which we now prove the remaining two parts.

Step 1 (Property (iii)). Fix a stochastic base $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ a filtration, an $\mathbb{R}^d$- or $\mathbb{C}^d$-valued standard $(\mathbb{F}, \mathbb{P})$-Brownian motion B and let $x \in \mathbb{K}^n$. Since the coefficients b and σ have the Lipschitz property (Definition C.25) and are continuous in the time argument, they are $\mathbb{G}^n$-progressive by Lemma C.29 (b). Hence they are admissible coefficients for the SDE (C.6). By Theorem C.34 (a), there is pathwise uniqueness of solutions of this SDE, so it remains to show that $F_x(B)$ is a strong solution of (C.6) with initial condition $F_x(B)_0 \overset{\text{a.s.}}{=} x$ (Definition C.15).

Since $W_d \ni w \mapsto F_x(w)$ is W_n-valued, $F_x(B)$ has continuous sample paths. This additionally shows that the pathwise integrability condition (C.12) is satisfied for $F_x(B)$ instead of X, using also that b and σ are of linearly bounded growth by Remark C.28. Adaptedness of $F_x(B)$ follows exactly as at the end of Step 3 of the proof of Theorem C.43, using that $\omega \mapsto B(\omega)$ is $\mathcal{F}_t$-$\overline{\mathcal{G}}^d_t$-measurable for all $t \in \mathbb{R}_+$. It remains to show that parts (a) and (c) of Definition C.15 are satisfied for $F_x(B)$ instead of X. By construction (see (C.133) and (C.134) in Step 1 of the proof of Theorem C.43),

$$F_x \overset{\text{uti}}{=} X^x \qquad \text{w.r.t. } \mathbb{W}, \tag{C.137}$$

and X^x solves the SDE (C.6) on the stochastic base $(W_d, \mathcal{G}^d, \overline{\mathbb{G}}^d, \mathbb{W})$ described at the beginning of Appendix C.4 with the canonical process $\Pi = (\Pi_t)_{t\in\mathbb{R}_+}$ as the driving Brownian motion. From (C.137), using also that indistinguishable integrands have indistinguishable stochastic integrals[47] (see [53, Exercise 5.113]),

$$\begin{aligned} F_x \overset{\text{uti}}{=} X^x &\overset{\text{uti}}{=} x + \int_0^{\cdot} b(s, X^x(\Pi))\,\mathrm{d}s + \int_0^{\cdot} \sigma(s, X^x(\Pi))\,\mathrm{d}\Pi_s \\ &\overset{\text{uti}}{=} x + \int_0^{\cdot} b(s, F_x(\Pi))\,\mathrm{d}s + \int_0^{\cdot} \sigma(s, F_x(\Pi))\,\mathrm{d}\Pi_s \qquad \text{w.r.t. } \mathbb{W}. \end{aligned} \tag{C.138}$$

Since $F_x(w)_0 = x$ for $\mathbb{W}$-a.e. $w \in W_d$ and $\mathcal{L}_{\mathbb{P}}[B] = \mathbb{W}$, it follows that $F_x(B)_0 \overset{\text{a.s.}}{=} x$. Hence it remains only to establish that

$$F_x(B) \overset{\text{uti}}{=} x + \int_0^{\cdot} b(s, F_x(B))\,\mathrm{d}s + \int_0^{\cdot} \sigma(s, F_x(B))\,\mathrm{d}B_s\,. \tag{C.139}$$

[46] Compared to Theorem C.34 and Theorem C.43 (iv), we get in Proposition C.58 strong existence without square integrability of the initial distribution, if in addition to the Lipschitz property (Definition C.25), there is continuity of coefficients in the time argument. By Remark C.28, continuity in time implies the linearly bounded growth condition (Definition C.27), hence the improvement is only in the integrability requirements on the initial distribution.

[47] The maps $(s, w) \mapsto \sigma(s, F_x(w))$ and $(s, w) \mapsto b(s, F_x(w))$ are $\overline{\mathbb{G}}^d$-progressive by Lemma C.7 and pathwise locally bounded by continuity in the time argument. Hence the stochastic integrals in (C.138) are well defined in the sense of Definition C.1.

For each $k \in \mathbb{N}$, and all $(t, x, w) \in \mathbb{R}_+ \times \mathbb{K}^n \times W_d$ let

$$H_k(x, w)_t := \sum_{\ell=0}^{2^k-1} \sigma(2^{-k}\ell t, F_x(w)) \left(w_{2^{-k}(\ell+1)t} - w_{2^{-k}\ell t}\right). \tag{C.140}$$

Using that $s \mapsto \sigma(s, F_x(w))$ is continuous for all $w \in W_d$, we have by an approximation result[48] for stochastic integrals ([38, Proposition 5.9]) that for every $t \in \mathbb{R}_+$ and every sequence $(n_k)_{k\in\mathbb{N}}$ in $\mathbb{N}$ converging to infinity,

$$H_{n_k}(x, w)_t \xrightarrow[k\to\infty]{\mathbb{W}} \int_0^t \sigma(s, F_x(\Pi))\, \mathrm{d}\Pi_s\,, \tag{C.141}$$

where $\xrightarrow{\mathbb{W}}$ denotes convergence in probability w.r.t. the Wiener measure on W_d. By passing to a subsequence, we can assume that the convergence is $\mathbb{W}$-a.s. Fixing $t \in \mathbb{R}_+$, it follows with (C.138) that for all w outside of a $(\mathcal{G}^d, \mathbb{W})$-null set N in W_d,

$$F_x(w)_t = x + \int_0^t b(s, F_x(w))\, \mathrm{d}s + \lim_{k\to\infty} H_{n_k}(x, w)_t \tag{C.142}$$

Since $\mathcal{L}_{\mathbb{P}}[B] = \mathbb{W}$, we have $\mathbb{P}[B \in N] = 0$, hence we can substitute B for w in (C.142) to obtain that

$$\begin{aligned} F_x(B)_t &\overset{\text{a.s.}}{=} x + \int_0^t b(s, F_x(B))\, \mathrm{d}s + \lim_{k\to\infty} H_{n_k}(x, B)_t \\ &\overset{\text{a.s.}}{=} x + \int_0^t b(s, F_x(B))\, \mathrm{d}s + \int_0^t \sigma(s, F_x(B))\, \mathrm{d}B_s \qquad \text{w.r.t. } \mathbb{P}, \end{aligned} \tag{C.143}$$

where the last equality follows by another application of the approximation argument by which we obtained (C.141). Since $t \in \mathbb{R}_+$ was arbitrary, $F_x(B)$ is a modification of $x + \int_0^{\cdot} b(s, F_x(B))\, \mathrm{d}s + \int_0^{\cdot} \sigma(s, F_x(B))\, \mathrm{d}B_s$, which by continuity of both processes establishes (C.139).[49]

Step 2 (Property (iv')). Fix a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$. Pathwise uniqueness of solutions follows as in the previous step from Theorem C.34 (a), so we need to check only that $F_\xi(B)$ is a strong solution of (C.6) with initial condition $F_\xi(B)_0 \overset{\text{a.s.}}{=} \xi$. We verify first adaptedness and continuity of $F_\xi(B)$ and then check that $F_\xi(B)$ satisfies parts (a) to (c) of Definition C.15.

Since F is W_n-valued, $F_\xi(B)$ has continuous sample paths. For adaptedness, fix $t \in \mathbb{R}_+$. By (iii), $F_x(B)_t$ is $\mathcal{F}_t$-measurable. Since $x \mapsto F_x(B(\omega))_t$ is continuous for all $\omega \in \Omega$, it follows with Lemma B.10 that $(x, \omega) \mapsto F_x(B(\omega))_t$ is $\mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{F}_t$-measurable. Since $\omega \mapsto (\xi(\omega), \omega)$ is $\mathcal{F}_t$-$(\mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{F}_t)$-measurable, $F_\xi(B)_t$ is $\mathcal{F}_t$-measurable. Since this holds for all $t \in \mathbb{R}_+$, $F_\xi(B)$ is adapted.

[48] This is also a direct consequence of continuity in the time argument and the dominated convergence theorem for stochastic integrals [53, Theorem 5.129].

[49] Since (C.143) holds for all $x \in \mathbb{K}^n$, it is tempting to simply substitute ξ for x in this equation and conclude. This is not valid, since the exceptional null set in (C.143) can depend on x, so a more elaborate argument is needed.

By part (iii), there exists an $(\mathcal{F},\mathbb{P})$-null set N such that

$$F_x(B(\omega))_0 = x, \qquad (x,\omega) \in D \times N^{\mathsf{c}}, \tag{C.144}$$

where $D := \mathbb{Q}^n$ if $\mathbb{K} = \mathbb{R}$ and $D := \mathbb{Q}^n + \mathrm{i}\mathbb{Q}^n$ if $\mathbb{K} = \mathbb{C}$. By part (i), $x \mapsto F_x(B(\omega))_0$ is continuous for all $\omega \in \Omega$, hence (C.144) implies that

$$F_x(B(\omega))_0 = x, \qquad (x,\omega) \in \mathbb{K}^n \times N^{\mathsf{c}},$$

which shows that $F_\xi(B)_0 \overset{\text{a.s.}}{=} \xi$. The integrability condition from Definition C.15(b) is satisfied by continuity of sample paths of $F_\xi(B)$, using also that b and σ are of linearly bounded growth as observed in Step 1 of this proof. It remains to show that

$$F_\xi(B) \overset{\text{uti}}{=} \xi + \int_0^\cdot b(s, F_\xi(B))\,\mathrm{d}s + \int_0^\cdot \sigma(s, F_\xi(B))\,\mathrm{d}B_s\,. \tag{C.145}$$

Define $G\colon \mathbb{K}^n \times W_d \to W_n$ and $H\colon \mathbb{K}^n \times W_d \to W_n$ by

$$\begin{aligned} G(x,w) &= \int_0^\cdot b(s, F_x(w))\,\mathrm{d}s\,, && \text{and} \\ H(x,w) &= F_x(w) - x - G(x,w), && (x,w) \in \mathbb{K}^n \times W_d. \end{aligned}$$

Using Lemma B.10 and the already established properties (i) and (ii) of F, the map $(x,w) \mapsto F_x(w)$ is $(\mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{G}^d)$-$\mathcal{G}^n$-measurable.[50] Using also the $\mathcal{G}^n$-$\mathcal{G}^n$-measurability of the Lebesgue–Stieltjes integral $W_n \ni w \mapsto \int_0^\cdot b(s,w)\,\mathrm{d}s \in W_n$ (see [53, Lemma 5.49(c)]), it follows that G is $(\mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{G}^d)$-$\mathcal{G}^n$ measurable. Therefore, the same is true for H. Consequently, (C.145) is equivalent to[51]

$$H(\xi, B) \overset{\text{uti}}{=} \int_0^\cdot \sigma(s, F_\xi(B))\,\mathrm{d}B_s\,. \tag{C.146}$$

We will establish (C.146) by proving, for each $t \in \mathbb{R}_+$, convergence in probability of $H_k(\xi,B)_t$, see (C.140), as $k \to \infty$, to both $H(\xi,B)_t$ and $\int_0^t \sigma(s, F_\xi(B))\,\mathrm{d}B_s$. Fix $t \in \mathbb{R}_+$. First, by another application of the previous approximation argument,

$$H_k(\xi,B)_t \xrightarrow[k\to\infty]{\mathbb{P}} \int_0^t \sigma(s, F_\xi(B))\,\mathrm{d}B_s\,. \tag{C.147}$$

Next, note that by (C.138), for all $x \in \mathbb{K}^n$

$$H(x,\Pi) \overset{\text{uti}}{=} \int_0^\cdot \sigma(s, F_x(\Pi))\,\mathrm{d}\Pi_s \qquad \text{w.r.t. } \mathbb{W}. \tag{C.148}$$

Using continuity of $s \mapsto \sigma(s,\tilde{w})$ for all $\tilde{w} \in W_n$, approximation arguments as in the previous step together with (C.148) show that

$$H_k(x,\Pi)_t \xrightarrow[k\to\infty]{\mathbb{W}} H(x,\Pi)_t \tag{C.149}$$

[50] See the proof of Corollary C.59(a) for details.

[51] Note that the stochastic integral on the right-hand side of (C.146) is a well-defined continuous local martingale, since σ is $\mathbb{G}^n$-progressive and of linearly bounded growth, as observed at the beginning of Step 1 of the proof.

for every $x \in \mathbb{K}^n$. We next establish that this implies

$$H_k(\xi, B)_t \xrightarrow[k\to\infty]{\mathbb{P}} H(\xi, B)_t. \tag{C.150}$$

Indeed, using that the $\mathcal{F}_0$-measurable random vector ξ is independent of B and applying Lemma B.11, we have for every $\varepsilon > 0$

$$\begin{aligned} \mathbb{P}[|H_k(\xi, B)_t - H(\xi, B)_t| \geq \varepsilon] &= \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{|H_k(\xi,B)_t - H(\xi,B)_t| \geq \varepsilon\}} \mid \sigma(\xi)]] \\ &= \mathbb{E}[\mathbb{P}[|H_k(x, B)_t - H(x, B)_t| \geq \varepsilon]\big|_{x=\xi}] \\ &= \int_{\mathbb{K}^n} \mathbb{P}[|H_k(x, B)_t - H(x, B)_t| \geq \varepsilon] \, \mathrm{d}\mathbb{P}_\xi(x)\,. \end{aligned} \tag{C.151}$$

From (C.149) and the fact that $\mathcal{L}_{\mathbb{P}}[B] = \mathbb{W}$, we have $\mathbb{P}[|H_k(x, B)_t - H(x, B)_t| \geq \varepsilon] \xrightarrow[k\to\infty]{} 0$ for all $x \in \mathbb{K}^n$. Using dominated convergence in (C.151) then yields (C.150). Combining (C.147) with (C.150) and using uniqueness up to equality a.s. of limits in probability implies that

$$H(\xi, B)_t \overset{\text{a.s.}}{=} \int_0^t \sigma(s, F_\xi(B)) \, \mathrm{d}B_s\,.$$

The preceding display holds for all $t \in \mathbb{R}_+$, hence continuity of sample paths of $H(\xi, B)$ and $\int_0^\cdot \sigma(s, F_\xi(B)) \, \mathrm{d}B_s$ implies indistinguishability of these processes and thus (C.146). □

We next state several corollaries relating to continuity and measurability properties of the flow (Corollary C.59), distributional properties of SDE solutions (Corollaries C.60 and C.61) and the Markov property of solutions (Theorem C.62).

Corollary C.59 (Joint continuity and measurability of the flow). *In the setting of Theorem C.43, denote by $F\colon \mathbb{K}^n \times W_d \to W_n$ the flow of the SDE (C.6).*

(a) For every $t \in \overline{\mathbb{R}}_+$, the map F is $(\mathcal{B}_{\mathbb{K}^n} \otimes \overline{\mathcal{G}}_t^d)$-$\mathcal{G}_t^n$-measurable.

(b) For each $w \in W_d$,

$$\mathbb{R}_+ \times \mathbb{K}^n \ni (t, x) \mapsto F_x(w)_t \in \mathbb{K}^n$$

is continuous w.r.t. the product $\mathcal{O}_{\mathbb{R}_+} \otimes \mathcal{O}_{\mathbb{K}^n}$ of Euclidean topologies on $\mathbb{R}_+ \times \mathbb{K}^n$.

(c) The map

$$\mathbb{R}_+ \times \mathbb{K}^n \times W_d \ni (t, x, w) \mapsto F_x(w)_t \in \mathbb{K}^n$$

is measurable w.r.t. $(\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{G}^d)$-$\mathcal{B}_{\mathbb{K}^n}$.

Proof. (a) By (C.81), $w \mapsto F_x(w)$ is $\overline{\mathcal{G}}_t^d$-$\mathcal{G}_t^n$-measurable for each fixed $x \in \mathbb{K}^n$. Using that $x \mapsto F_x(w)$ is continuous for all $w \in W_d$ by Theorem C.43(i), Lemma B.10 implies $(\mathcal{B}_{\mathbb{K}^n} \otimes \overline{\mathcal{G}}_t^d)$-$\mathcal{G}_t^n$-measurability of F.

(b) Fix $w \in W_d$, $(t, x) \in \mathbb{R}_+ \times \mathbb{K}^n$, $\varepsilon > 0$ and let $(t_n, x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_+ \times \mathbb{K}^n$ converging to (t, x). Since $t_n \to t$, there exists an $n_0 \in \mathbb{N}$ with $t_n \leq t + 1$ for all $n \geq n_0$.

By continuity of $x \mapsto F_x(w)$ w.r.t. the topology $\mathcal{O}_{\mathrm{ucc}}$ of compact convergence in W_n, there exists an $n_1 \geq n_0$ such that

$$\sup_{s\in[0,t+1]} |F_{x_n}(w)_s - F_x(w)_s| < \varepsilon/2, \qquad n \geq n_1. \tag{C.152}$$

Since F is W^n-valued, $s \mapsto F_x(w)_s$ is continuous, there exists an $n_2 \geq n_1$ such that

$$|F_x(w)_{t_n} - F_x(w)_t| < \varepsilon/2, \qquad n \geq n_2. \tag{C.153}$$

By the triangle inequality, (C.152) and (C.153),

$$|F_{x_n}(w)_{t_n} - F_x(w)_t| \leq |F_{x_n}(w)_{t_n} - F_x(w)_{t_n}| + |F_x(w)_{t_n} - F_x(w)_t| < \varepsilon$$

for all $n \geq n_2$.

(c) By part (b), $(t, x, w) \mapsto F_x(w)_t$ is continuous in (t, x) for each fixed $w \in W_d$. By Theorem C.43(ii), $W_d \ni w \mapsto F_x(w)_t$ is $\mathcal{G}^d$-$\mathcal{B}_{\mathbb{K}^n}$-measurable for all fixed $(t, x) \in \mathbb{R}_+ \times \mathbb{K}^n$. Hence $(t, x, w) \mapsto F_x(w)_t$ is $(\mathcal{B}_{\mathcal{O}_{\mathbb{R}_+}\otimes\mathcal{O}_{\mathbb{K}^n}} \otimes \mathcal{G}^d)$-$\mathcal{B}_{\mathbb{K}^n}$-measurable by Lemma B.10. Since $\mathcal{B}_{\mathcal{O}_{\mathbb{R}_+}\otimes\mathcal{O}_{\mathbb{K}^n}} = \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{K}^n}$, see [53, Theorem 15.53], the claim follows. □

Corollary C.60 (Regular conditional distribution of SDE solutions given the initial condition)**.** *Consider $\mathbb{G}^n$-progressive functions b and σ as in* (C.4) *which satisfy the Lipschitz continuity and bounded growth conditions for pathwise uniqueness and strong existence of solutions from Theorem C.34. Let $F\colon \mathbb{K}^n \times W_d \to W_n$ be the flow of the SDE* (C.6) *from Theorem C.43, and let X be a strong solution w.r.t. a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B)$. Then X, viewed as a random variable taking values in $(W_n, \mathcal{G}^n)$, has a regular conditional distribution*[52] *given X_0, which is given by*

$$\kappa\colon \Omega \times \mathcal{G}^n \to [0,1], \quad \kappa(\omega, G) := \mathbb{P}[X^x \in G]\big|_{x=X_0(\omega)},$$

where for each $x \in \mathbb{K}^n$, X^x is a strong solution of (C.6) *with initial condition $X_0^x \overset{\text{a.s.}}{=} x$.*

Proof. Since the coefficients b and σ have the Lipschitz property and are of linearly bounded growth by assumption, Corollary C.57 implies that the maps $x \mapsto \mathbb{P}[X^x \in G]$, $G \in \mathcal{G}^n$, do not depend on the particular family $(X^x)_{x\in\mathbb{K}^n}$ of strong solutions. Hence we may assume w.l.o.g. that for each $x \in \mathbb{K}^n$, $X^x = F_x(B)$. By Theorem C.43(iv),

$$X \overset{\text{uti}}{=} F_\xi(B) \overset{\text{uti}}{=} F_{X_0}(B). \tag{C.154}$$

Fix $A \in \mathcal{B}_{\mathbb{K}^n}$ and $G \in \mathcal{G}^n$. By (C.154),

$$\begin{aligned}\mathbb{P}\left[X \in G \,|\, \sigma(X_0)\right] &\overset{\text{a.s.}}{=} \mathbb{P}\left[F_{X_0}(B) \in G \,|\, \sigma(X_0)\right] \\ &\overset{\text{a.s.}}{=} \mathbb{P}\left[F_x(B) \in G\right]\big|_{x=X_0} = \mathbb{P}[X^x \in G]\big|_{x=X_0} = \kappa(\,\cdot\,, G),\end{aligned}$$

[52] See [36, §8.3] for details on regular conditional distributions.

where the second equality follows using Lemma B.11 by independence of the $\mathcal{F}_0$-measurable random vector X_0 and the $\mathbb{F}$-Brownian motion B. Hence κ is a conditional distribution of X given X_0. It remains to show regularity, i.e. that κ is a stochastic kernel.

Clearly, $\kappa(\omega, \cdot\,)$ is a probability measure for all $\omega \in \Omega$, so it remains to show measurability of $\omega \mapsto \kappa(\omega, G)$ for every fixed $G \in \mathcal{G}^n$. If $f\colon W_n \to \mathbb{R}$ is a bounded $\mathcal{O}_{\mathrm{ucc}}$-continuous function, then $\mathbb{K}^n \ni x \mapsto f(X^x)$ is continuous by definition of X^x and Theorem C.43 (i). Since $\mathcal{G}^n = \sigma(\mathcal{O}_{\mathrm{ucc}})$ by Lemma 2.1 (d), $\mathcal{O}_{\mathrm{ucc}}$-continuity of f implies that f is $\mathcal{G}^n$-$\mathcal{B}_{\mathbb{R}}$-measurable. Hence $\mathbb{K}^n \times \Omega \ni (x, \omega) \mapsto f(X^x(\omega))$ is continuous in x and measurable in ω, and thus $\mathcal{B}_{\mathbb{K}^n} \otimes \mathcal{F}$-measurable by Lemma B.10. Hence $\mathbb{K}^n \ni x \mapsto \mathbb{E}[f(X^x)]$ is measurable. By monotonously approximating indicator function of closed sets with bounded continuous functions (see e.g. [36, Lemma 13.10]), it follows that $x \mapsto \mathbb{E}[\mathbb{1}_C(X^x)] = \mathbb{P}[X^x \in C]$ is measurable for all closed $C \subseteq W_n$. Since the closed subsets of W_n form a π-system generating $\mathcal{G}^n$, it follows using [53, Lemma 15.132] that $x \mapsto \mathbb{P}[X^x \in G]$ is measurable for all $G \in \mathcal{G}^n$. This implies measurability of $\omega \mapsto \kappa(\omega, G)$ for all $G \in \mathcal{G}^n$ and thus the kernel property of κ. $\square$

Corollary C.61 (Law of the solution of an SDE as a mixture distribution)**.** *Consider $\mathbb{G}^n$-progressive functions b and σ as in (C.4) which satisfy the Lipschitz continuity and bounded growth conditions for pathwise uniqueness and strong existence of solutions from Theorem C.34. Fix $\mu \in \mathcal{P}(\mathbb{K}^n)$ and let X be a strong solution of the SDE (C.6) with initial distribution μ w.r.t. a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B)$. Then*

$$\mathbb{P}[X \in A] = \int_{\mathbb{K}^n} \mathbb{P}[X^x \in A]\,\mathrm{d}\mu(x)\,, \qquad A \in \mathcal{G}^n,$$

where for each $x \in \mathbb{K}^n$, X^x is a strong solution of (C.6) with initial condition $X_0^x \overset{\text{a.s.}}{=} x$.

Proof. Let $F\colon \mathbb{K}^n \times W_d \to W_n$ be the flow of the SDE (C.6) from Theorem C.43. By Theorem C.43 (iv), we may assume that $X = F_\xi(B)$ and $X^x = F_x(B)$ for each $x \in \mathbb{K}^n$. Using Corollary C.60 and denoting by κ the regular conditional distribution of X given X_0, it follows that for each $A \in \mathcal{G}^n$,

$$\begin{aligned}\mathbb{P}[X \in A] &= \mathbb{E}[\mathbb{P}[X \in A \mid X_0]] \\ &= \mathbb{E}\big[\mathbb{P}[X^x \in A]\big|_{x=X_0}\big] = \int_{\mathbb{K}^n} \mathbb{P}[X^x \in A]\,\mathrm{d}\mu(x)\,.\end{aligned}$$ $\square$

In the case of nonrandom, time-homogeneous dependence of the SDE coefficients on the solutions, see Definition C.9, solutions of (C.9) are Markov processes, provided the coefficients are Lipschitz continuous. We will not go into details about Markov processes here and instead refer to [38, §8] and [53, §9] for details.

Theorem C.62 (Markov property of solutions of classical SDEs with time-homogeneous coefficients [38, Theorem 8.6])**.** *Let $b\colon \mathbb{K}^n \to \mathbb{K}^n$ and $\sigma\colon \mathbb{K}^n \to \mathbb{K}^{n\times d}$ be Lipschitz continuous and let $X = (X_t)_{t\in\mathbb{R}_+}$ be a strong solution of the SDE (C.10) on a set-up $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$. Then X is an $(\mathbb{F}, \mathbb{P})$-Markov process with transition semigroup $Q = (Q_t)_{t\in\mathbb{R}_+}$, where*

$$Q_t(x, \cdot\,) \coloneqq \mathcal{L}[X_t^x], \qquad (t, x) \in \mathbb{R}_+ \times \mathbb{K}^n, \tag{C.155}$$

and for each $x \in \mathbb{K}^n$, $X^x = (X_t^x)_{t\in\mathbb{R}_+}$ is a solution of (C.10) with initial condition $X_0^x \overset{\text{a.s.}}{=} x$.

Proof. [38, Theorem 8.6] Lipschitz continuity and independence of the time argument ensure by Proposition C.58 existence of the flow $F\colon \mathbb{K}^n \times W_d \to W_n$ associated with the SDE. By Proposition C.58 (iii), the family of probability measures $(Q_t)_{t\in\mathbb{R}_+}$ is well-defined and independent of the stochastic base and the Brownian motion associated with the family of solutions $(X^x)_{x\in\mathbb{K}^n}$, which we can assume to be $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and B, respectively, w.l.o.g. For every bounded measurable map $f\colon \mathbb{K}^n \to \mathbb{K}$, let

$$Q_t f(x) := \mathbb{E}[f(X_t^x)] = \int_{\mathbb{K}^n} f(y)\,\mathrm{d}Q_t(\,\cdot\,, \mathrm{d}y)\,, \qquad (t,x) \in \mathbb{R}_+ \times \mathbb{K}^n,$$

and note that since $X^x \overset{\text{uti}}{=} F_x(B)$, we also have

$$Q_t f(x) = \mathbb{E}[f(F_x(B)_t)] = \int_{\mathbb{K}^n} f(F_x(w)_t)\,\mathrm{d}\mathbb{W}(w)\,. \tag{C.156}$$

We next verify that for all bounded measurable $f\colon \mathbb{K}^n \to \mathbb{K}$,

$$\mathbb{E}[f(X_{s+t})|\mathcal{F}_s] = Q_t f(X_s), \qquad s,t \in \mathbb{R}_+. \tag{C.157}$$

Fix $s \in \mathbb{R}_+$. Then for all $t \in \mathbb{R}_+$,

$$X_{s+t} \overset{\text{a.s.}}{=} X_s + \int_s^{s+t} b(X_r)\,\mathrm{d}r + \int_s^{s+t} \sigma(X_r)\,\mathrm{d}B_r\,. \tag{C.158}$$

Let

$$X'_t := X_{s+t}, \quad \mathcal{F}'_t := \mathcal{F}_{s+t}, \quad B'_t := B_{s+t} - B_s, \qquad t \in \mathbb{R}_+.$$

Then $(\Omega, \mathcal{F}, \mathbb{F}', \mathbb{P})$ with $\mathbb{F}' := (\mathcal{F}'_t)_{t\in\mathbb{R}_+}$ is a stochastic base (Definition C.5 (a)) and B' is a d-dimensional standard $(\mathbb{F}', \mathbb{P})$-Brownian motion. By a pathwise approximation result for stochastic integrals (see [38, Proposition 5.9]), it can be shown that

$$\int_s^{s+t} \sigma(X_r)\,\mathrm{d}B_r \overset{\text{a.s.}}{=} \int_0^t \sigma(X'_u)\,\mathrm{d}B'_u\,, \qquad t \in \mathbb{R}_+,$$

where the stochastic integral on the right-hand side is w.r.t. the filtration $\mathbb{F}'$. Comparison with (C.158) yields that

$$X' \overset{\text{uti}}{=} X_s + \int_0^{\cdot} b(X'_u)\,\mathrm{d}u + \int_0^{\cdot} \sigma(X'_u)\,\mathrm{d}B'_u\,,$$

where the equality is up to indistinguishability by continuity of both sides of the previous display. Therefore, X' is a strong solution of (C.10) on the set-up $(\Omega, \mathcal{F}, \mathbb{F}', \mathbb{P}, B', X_s)$ with the $\mathcal{F}'_0$-measurable initial random vector $X'_0 = X_s$. By Proposition C.58 (iv') it follows that $X' \overset{\text{uti}}{=} F_{X_s}(B')$. Consequently, for every $t \in \mathbb{R}_+$ and every bounded measurable function $f\colon \mathbb{K}^n \to \mathbb{K}$,

$$\begin{aligned}\mathbb{E}[f(X_{s+t})|\mathcal{F}_s] = \mathbb{E}[f(X'_t)|\mathcal{F}_s] &\overset{\text{a.s.}}{=} \mathbb{E}[f(F_{X_s}(B')_t))|\mathcal{F}_s] \\ &\overset{\text{a.s.}}{=} \mathbb{E}[f(F_x(B')_t)]\big|_{x=X_s} = Q_t f(X_s),\end{aligned}$$

where we used in the third equality Lemma B.11 together with the fact that the standard $\mathbb{F}'$-Brownian motion B' is independent of $\mathcal{F}'_0 = \mathcal{F}_s$, while X_s is $\mathcal{F}_s$-measurable, and obtained the last equality by (C.156).

It remains to verify that $Q = (Q_t)_{t\in\mathbb{R}_+}$ is a transition semigroup. The representation (C.156) together with the joint measurability property of the flow (see Corollary C.59(c)) show that $\mathbb{R}_+ \times \mathbb{K}^n \ni (t, x) \mapsto Q_t f(x)$ is measurable for all bounded measurable $f : \mathbb{K}^n \to \mathbb{K}$, implying that Q_t is a transition kernel for all $t \in \mathbb{R}_+$. For the Chapman–Kolmogorov equations, it suffices to observe that by applying (C.157) to X^x, we obtain for all $s, t \in \mathbb{R}_+$, all bounded measurable $f : \mathbb{K}^n \to \mathbb{K}$ and all $x \in \mathbb{K}^n$

$$\begin{aligned} Q_{s+t} f(x) &= \mathbb{E}[f(X^x_{s+t})] = \mathbb{E}[\mathbb{E}[f(X^x_{s+t})|\mathcal{F}_s]] \\ &= \mathbb{E}[Q_t f(X^x_s)] = \int_{\mathbb{K}^n}\int_{\mathbb{K}^n} f(z) Q_t(y, \mathrm{d}z) Q_s(x, \mathrm{d}y) = Q_s Q_t f(x), \end{aligned}$$

which proves the claim. □

D. Wasserstein Spaces

In this appendix, we first (Appendix D.1) review basic definitions and selected important results from the theory of Wasserstein spaces, which provide important topological and measurable structures underpinning McKean SDEs.. Subsequently, in Appendix D.2, we establish related measurability results for random measures which are used to establish the propagation of chaos (see Section 2.4).

D.1. Selected Facts about Wasserstein Spaces

Definition D.1 (Wasserstein space)**.** [61, Definition 6.4] Let (S, d) be a metric space, and $p \in [1, \infty)$. The Wasserstein space $\mathcal{P}_p(S)$ of order p is defined as the set of probability measures on S with a finite p-th moment, i.e.

$$\mathcal{P}_p(S) = \Big\{ \mu \in \mathcal{P}(S) : \int_S d(x, x_0)^p \, \mathrm{d}\mu(x) < \infty \Big\}, \tag{D.1}$$

where $x_0 \in S$ is arbitrary.[1]

Note that $\mathcal{P}_p(S) = \mathcal{P}(S)$ whenever the metric d is bounded.

Definition D.2 (Wasserstein distance)**.** [61, Definition 6.1] Let (S, d) be a separable metric space, and let $p \in [1, \infty)$. The Wasserstein distance of order p between two probability measures[2] μ and ν on S is defined by

$$\mathcal{W}_p(\mu, \nu) = \left(\inf_{\pi \in \mathcal{P}(\mu,\nu)} \int_{S \times S} d(x, y)^p \, \mathrm{d}\pi(x, y) \right)^{1/p}, \tag{D.2}$$

where $\mathcal{P}(\mu, \nu)$ denotes the set of all probability measures π on $(S \times S, \mathcal{B}_S \otimes \mathcal{B}_S)$ with marginals μ and ν.

The following example extends an observation made in [33, Equation (5)] for the case $S = \mathbb{R}^n$ and $p = 2$.

Example D.3 (Wasserstein distance of empirical measures)**.** Let (S, d) be a metric space, fix $p \in [1, \infty)$ and let $x_1, \ldots, x_N$ and $y_1, \ldots, y_N$ be points in S. Then

$$\mathcal{W}_p\left(\frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \frac{1}{N} \sum_{i=1}^N \delta_{y_i} \right) \leq \left(\frac{1}{N} \sum_{i=1}^N d(x_i, y_i)^p \right)^{1/p} \tag{D.3}$$

[1] By the triangle inequality for d, the definition of $\mathcal{P}_p(S)$ does not depend on the point $x_0 \in S$.

[2] For general probability measures μ and ν in $\mathcal{P}(S)$, the distance $\mathcal{W}_p(\mu, \nu)$ defined by (D.2) may be infinite. When μ and ν are restricted to lie in $\mathcal{P}_p(S)$, $\mathcal{W}_p(\mu, \nu)$ is finite by the triangle inequality for the metric d and Jensen's inequality (C.18). In this case, $\mathcal{W}_p$ is a metric on $\mathcal{P}_p(S)$, see Lemma D.5 below.

as a consequence of the definition (D.2) of $\mathcal{W}_p$ and the observation that the measure $\pi := \frac{1}{N}\sum_{i=1}^N \delta_{(x_i,y_i)}$ is a coupling of $\frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ and $\frac{1}{N}\sum_{i=1}^N \delta_{y_i}$.

Remark D.4 (Wasserstein distances are increasing in the order p)**.** [61, Remark 6.6] For all $p, q \in [1,\infty)$ with $p \leq q$, taking the q-th power in (D.2) and applying Jensen's inequality with the convex function $x \mapsto x^{q/p}$ shows that

$$\mathcal{W}_p \leq \mathcal{W}_q. \tag{D.4}$$

Lemma D.5 (Wasserstein distances are metrics)**.** *[60, Example 6.3] For every Polish space (S,d) and every $p \in [1,\infty)$, $\mathcal{W}_p$ defines a metric on $\mathcal{P}_p(S)$.*

Polish spaces appear in Lemma D.5, because the proof of Lemma D.5 uses regular conditional distributions to establish the triangle inequality.

Theorem D.6 (Relation of the Wasserstein metrics to weak convergence, [60, Theorem 7.12])**.** *Let (S,d) be a Polish space, $p \in [1,\infty)$, $(\mu_k)_{k\in\mathbb{N}}$ a sequence of probability measures in $\mathcal{P}_p(S)$, and let $\mu \in \mathcal{P}(S)$. Then, the following four statements are equivalent:*

(a) $\mathcal{W}_p(\mu_k, \mu) \underset{k\to\infty}{\longrightarrow} 0$.

(b) $\mu_k \to \mu$ weakly, and (μ_k) satisfies the following "tightness" condition: for some (and thus every) $x_0 \in S$,

$$\lim_{R\to\infty} \limsup_{k\to\infty} \int_{d(x,x_0)\geq R} d(x,x_0)^p \, \mathrm{d}\mu_k(x) = 0.$$

(c) $\mu_k \underset{k\to\infty}{\longrightarrow} \mu$ weakly, and there is convergence of the moment of order p: for some (and thus every) $x_0 \in S$,

$$\int_S d(x,x_0)^p \, \mathrm{d}\mu_k(x) \underset{k\to\infty}{\longrightarrow} \int_S d(x,x_0)^p \, \mathrm{d}\mu(x)\,.$$

(d) Whenever a continuous function φ on S satisfies, for some $x_0 \in S$, some $C \in \mathbb{R}_+$ and all $x \in S$, the bound $|\varphi(x)| \leq C[1 + d(x_0,x)^p]$, then

$$\int_S \varphi \, \mathrm{d}\mu_k \underset{k\to\infty}{\longrightarrow} \int_S \varphi \, \mathrm{d}\mu\,.$$

Remark D.7. Each of the four equivalent conditions in Theorem D.6 implies that $\mu \in \mathcal{P}_p(S)$. To see this, assume w.l.o.g. that $\mathcal{W}_p(\mu_k,\mu) \to 0$ as $k \to \infty$. Then there exists an index $k_0 \in \mathbb{N}$ and (see (D.2)) a probability measure $\pi \in \mathcal{P}(\mu, \mu_{k_0})$ with

$$\int_{S\times S} d(x,y)^p \, \mathrm{d}\pi(x,y) < \infty. \tag{D.5}$$

Fix $x_0 \in S$. By the law of the unconscious statistician, the triangle inequality for the metric d, and Jensen's inequality (C.18),

$$\begin{aligned}
\int_S d(x,x_0)^p \,\mathrm{d}\mu(x) &= \int_{S\times S} d(x,x_0)^p \,\mathrm{d}\pi(x,y) \\
&\le \int_{S\times S} \big(d(x,y)+d(y,x_0)\big)^p \,\mathrm{d}\pi(x,y) \\
&\le 2^{p-1}\int_{S\times S} d(x,y)^p + d(y,x_0)^p \,\mathrm{d}\pi(x,y) \\
&= 2^{p-1}\int_{S\times S} d(x,y)^p \,\mathrm{d}\pi(x,y) + 2^{p-1}\int_{S\times S} d(y,x_0)^p \,\mathrm{d}\mu_{k_0}(y) \\
&< \infty,
\end{aligned}$$

where we have used (D.5) and the fact that $\mu_{k_0} \in \mathcal{P}_p(S)$ (see Definition D.1) to conclude finiteness.

Theorem D.8 ($\mathcal{W}_p$ induces a Polish topology on $\mathcal{P}_p(S)$)**.** *[61, Theorem 6.18] Let (S,d) be a Polish space and $p \in [1,\infty)$. Then $(\mathcal{P}_p(S), \mathcal{W}_p)$ is a Polish space.*

The following proposition is a variation for Wasserstein spaces of a famous result, also known as the Glivenko–Cantelli theorem, due to V. S. Varadarajan [59]. It states that the empirical measures associated with a sequence $(\xi_i)_{i\in\mathbb{N}}$ of pairwise independent and identically distributed random variables taking values in a Polish space S converge[3] weakly almost surely to $\mathcal{L}[\xi_1]$. This result will be an important tool in the proof that the particle approximation for the McKean SDE converges (Theorem 2.56).

Theorem D.9 (Glivenko–Cantelli theorem, convergence of empirical measures of i.i.d. random variables w.r.t. the Wasserstein metric, [33, Lemma 4])**.** *Let (S,d) be a non-empty Polish space, $p \in [1,\infty)$ and equip $\mathcal{P}_p(S)$ with the Wasserstein metric $\mathcal{W}_p$ of order p. Fix $\mu \in \mathcal{P}_p(S)$, let $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of S-valued, pairwise independent and identically distributed random variables with law μ, and define the sequence*

$$\mu^N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_i}, \qquad N \in \mathbb{N}, \tag{D.6}$$

of corresponding empirical measures. Let $x_0 \in S$ be an arbitrary point. Then for all $N \in \mathbb{N}$,

$$\mathbb{E}[\mathcal{W}_p^p(\mu^N,\mu)] \le 2^p \int_S d(x,x_0)^p \,\mathrm{d}\mu(x) \tag{D.7}$$

and

$$\mathcal{W}_p^p(\mu^N,\mu) \xrightarrow[N\to\infty]{} 0 \qquad \textit{a.s. and in } L^1. \tag{D.8}$$

[3] Note that while Theorem D.9 below assumes that $\mathcal{L}[\xi_1] \in \mathcal{P}_p(S)$, the proof gives the a.s. weak convergence also without this moment condition.

Proof. [33, Lemma 4] By Etemadi's strong law of large numbers [36, Theorem 5.17], for every μ-integrable function $f\colon S \to \mathbb{K}$,

$$\int_S f \,\mathrm{d}\mu^N = \frac{1}{N}\sum_{i=1}^N f(\xi_i) \xrightarrow[N\to\infty]{\text{a.s.}} \int_S f\,\mathrm{d}\mu\,. \tag{D.9}$$

By [4, Proposition 7.19], the weak topology on $\mathcal{P}_p(S)$ is induced by a countable subfamily $\mathcal{D} \subseteq C_\mathrm{b}(S)$. Then by (D.9), $\int_S f\,\mathrm{d}\mu^N \xrightarrow[N\to\infty]{} \int_S f\,\mathrm{d}\mu$ for all $f \in \mathcal{D}$ away from a $\mathbb{P}$-null set. By [4, Lemma 7.6], it follows that $\mu^N \to \mu$ weakly almost surely. Since $\mu \in \mathcal{P}_p(S)$, the function $x \mapsto d(x,x_0)^p$ is μ-integrable by Definition D.1, and we obtain

$$\int_S d(x,x_0)^p \,\mathrm{d}\mu^N(x) = \frac{1}{N}\sum_{i=1}^N d(\xi_i,x_0)^p \xrightarrow[N\to\infty]{\text{a.s.}} \int_S d(x,x_0)^p\,\mathrm{d}\mu(x)\,. \tag{D.10}$$

From Theorem D.6, (c) $\Rightarrow$ (a), $\mathcal{W}_p(\mu^N,\mu) \xrightarrow[N\to\infty]{} 0$ almost surely. For the bound (D.7), fix $N \in \mathbb{N}$ and observe that for every coupling π of μ^N and μ, using Jensen's inequality (see (C.18)) for the second inequality,

$$\begin{aligned}\mathcal{W}_p^p(\mu^N,\mu) &\le \int_{S\times S} d(x,y)^p\,\mathrm{d}\pi(x,y) \le 2^{p-1}\int_{S\times S}\big(d(x,x_0)^p + d(y,x_0)^p\big)\,\mathrm{d}\pi(x,y)\\ &= 2^{p-1}\int_S d(x,x_0)^p\,\mathrm{d}\mu^N(x) + 2^{p-1}\int_S d(y,x_0)^p\,\mathrm{d}\mu(y)\\ &= \frac{2^{p-1}}{N}\sum_{i=1}^N d(\xi_i,x_0)^p + 2^{p-1}\int_S d(x,x_0)^p\,\mathrm{d}\mu(x)\,.\end{aligned} \tag{D.11}$$

Taking the expectation[4] in the previous display yields (D.7).

It remains to show that $\mathcal{W}_p^p(\mu^N,\mu)$ converges to zero in L^1 as $N \to \infty$. We have already shown almost sure convergence, hence it suffices by Vitali's convergence theorem [53, Theorem 4.60] to establish uniform integrability of $\{\mathcal{W}_p^p(\mu^N,\mu)\}_{N\in\mathbb{N}}$. For each $N \in \mathbb{N}$, let Z_N denote the right-hand side of (D.11). By (D.10), $Z_N \xrightarrow{\text{a.s.}} 2^p \int_S d(x,x_0)^p\,\mathrm{d}\mu(x)$ as $N \to \infty$. Moreover, $Z_N \ge 0$ and $\mathbb{E}[Z_N] = 2^p\int_S d(x,x_0)^p\,\mathrm{d}\mu(x)$ for all $N \in \mathbb{N}$. Therefore Z_N converges in L^1 to $2^p\int_S d(x,x_0)^p\,\mathrm{d}\mu(x)$ as $N \to \infty$ by Scheffé's lemma [63, §5.10]. Hence $\{Z_N\}_{N\in\mathbb{N}}$ is uniformly integrable by another application of Vitali's convergence theorem. With the bound (D.11), uniform integrability of $\{\mathcal{W}_p^p(\mu^N,\mu)\}_{N\in\mathbb{N}}$ follows. □

For the case $S = \mathbb{K}^d$, a stronger independence condition and under an additional moment condition, explicit rates of convergence of $\mathbb{E}[\mathcal{W}_p^p(\mu^N,\mu)]$ in Theorem D.9 can be given. This will be useful to obtain a corresponding rate of convergence of the particle approximation for solutions of McKean SDEs (Theorem 2.58).

Theorem D.10 (Rate of convergence in Theorem D.9 in the i.i.d. case when $S = \mathbb{K}^d$, [23, Theorem 1])**.** *Let $\mu \in \mathcal{P}(\mathbb{K}^d)$ and let $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random vectors with*

[4] To be precise, we should also verify that $\Omega \ni \omega \mapsto \mathcal{W}_p^p(\mu^N,\mu) \in \mathbb{R}_+$ is measurable. This follows from Lemma D.16 below, whose proof is independent of the present lemma.

law μ. Fix $p \in [1,\infty)$ and denote by d_p the Wasserstein metric of order p on $\mathcal{P}_p(\mathbb{K}^d)$ (see Definition 2.4). Assume that

$$M_q(\mu) \coloneqq \int_{\mathbb{K}^d} |x|^q \,\mathrm{d}\mu(x) < \infty \tag{D.12}$$

for some $q > p$ and let the sequence $(\mu^N)_{N\in\mathbb{N}}$ of empirical measures be defined as in (D.6). There exists a constant C depending only on p, d and q such that for all $N \in \mathbb{N}$,

$$\mathbb{E}\big[d_p^p(\mu^N,\mu)\big] \le C\, M_q^{p/q}(\mu)\,\alpha_{d_{\mathbb{K}},p,q}(N), \tag{D.13}$$

where for all $N \in \mathbb{N}$,

$$\alpha_{d_{\mathbb{K}},p,q}(N) \coloneqq \begin{cases} N^{-1/2} + N^{-(q-p)/q} & \text{if } p > d_{\mathbb{K}}/2 \text{ and } q \neq 2p, \\ N^{-1/2}\log(1+N) + N^{-(q-p)/q} & \text{if } p = d_{\mathbb{K}}/2 \text{ and } q \neq 2p, \\ N^{-p/d_{\mathbb{K}}} + N^{-(q-p)/q} & \text{if } p \in [1, d_{\mathbb{K}}/2) \text{ and } q \neq d_{\mathbb{K}}/(d_{\mathbb{K}}-p), \end{cases} \tag{D.14}$$

with $d_{\mathbb{K}} \coloneqq \dim_{\mathbb{R}}(\mathbb{K}^d)$.

For some of the cases in (D.14), explicit bounds for the constant C in (D.13) can be given. The following result is a specialization and simplification of [22, Theorem 1], using also the bounds from [22, Equation (2) and §2.4].

Theorem D.11 (Upper bounds for the constant C in (D.13), [22, Theorem 1])**.** *For all $x > 0$ and $q > s > 0$, define*

$$H(x,s,q) = \Big(x\frac{q-s}{s} + (1+x)\Big(\frac{q}{s}\Big)^{q/(q-s)}\Big)^{s/q}\frac{q}{q-s}, \tag{D.15}$$

and let $d_{\mathbb{K}} \coloneqq \dim_{\mathbb{R}}(\mathbb{K})$. Under the assumptions of Theorem D.10, the following holds:

(a) If $p > d_{\mathbb{K}}/2$ and $q > 2p$, let[5]

$$\kappa_{d_{\mathbb{K}},p} \coloneqq \frac{2^{d_{\mathbb{K}}/2-1}}{1-2^{d_{\mathbb{K}}/2-p}} \qquad \text{and} \qquad \theta_{d_{\mathbb{K}},p,q} \coloneqq H\Big(\frac{1}{2\kappa_{d_{\mathbb{K}},p}}, 2p, q\Big). \tag{D.16}$$

(b) If $p = d_{\mathbb{K}}/2$ and $q > 2p$, let

$$\kappa_{d_{\mathbb{K}},p,N} \coloneqq \frac{2^{p-1}}{p\log 2}\Big(\log\Big(\big(2^{1-p}-2^{1-2p}\big)\sqrt{N}\Big)\Big)^+ + \frac{2^{p-1}}{1-2^{-p}}, \tag{D.17}$$

and

$$\theta_{d_{\mathbb{K}},p,q,N} \coloneqq H\Big(\frac{1}{2\kappa_{d_{\mathbb{K}},p,N}}, 2p, q\Big).$$

[5] The constant $\kappa_{d_{\mathbb{K}},p}$ corresponds to $\kappa^{(\infty)}_{d,p}$ in [22, Theorem 1]. We present bounds chosen for simplicity instead of optimality, i.e., improved bounds can be obtained from a careful study of [22].

(c) If $p \in (0, d_{\mathbb{K}}/2)$ and $q > d_{\mathbb{K}}p/(d_{\mathbb{K}} - p)$, let

$$\kappa_{d_{\mathbb{K}},p} := \frac{2^{p-2p/d_{\mathbb{K}}}\left(1 - 2^{-d_{\mathbb{K}}/2}\right)^{1-2p/d_{\mathbb{K}}}}{1 - 2^{p-d_{\mathbb{K}}/2}} \qquad \text{and} \qquad \theta_{d_{\mathbb{K}},p,q} := H\Big(\frac{2^{-2p/d_{\mathbb{K}}}}{\kappa_{d_{\mathbb{K}},p}}, \frac{d_{\mathbb{K}}p}{d_{\mathbb{K}} - p}, q\Big). \tag{D.18}$$

With the definitions from (D.16) to (D.18), we have for all $N \in \mathbb{N}$

$$\mathbb{E}[d_p^p(\mu_N, \mu)] \le M_q^{p/q}(\mu)\tilde{\alpha}_{d_{\mathbb{K}},p,q}(N), \tag{D.19}$$

where

$$\tilde{\alpha}_{d_{\mathbb{K}},p,q}(N) = 2^p d_{\mathbb{K}}^{p/2} \begin{cases} \kappa_{d_{\mathbb{K}},p}\theta_{d_{\mathbb{K}},p,q}N^{-1/2} & p > d_{\mathbb{K}}/2 \text{ and } q > 2p, \\ \kappa_{d_{\mathbb{K}},p,N}\theta_{d_{\mathbb{K}},p,q,N}N^{-1/2} & p = d_{\mathbb{K}}/2 \text{ and } q > 2p, \\ \kappa_{d_{\mathbb{K}},p}\theta_{d_{\mathbb{K}},p,q}N^{-p/d_{\mathbb{K}}} & p \in [1, d_{\mathbb{K}}/2) \text{ and } q > d_{\mathbb{K}}p/(d_{\mathbb{K}} - p). \end{cases} \tag{D.20}$$

Table D.1 shows specific values of the products $2^p d_{\mathbb{K}}^{p/2}\kappa_{d_{\mathbb{K}},p}\theta_{d_{\mathbb{K}},p,q}$ multiplying the N-dependent terms in the first and third case of (D.20).

	$p = 1$			$p = 2$			$p = 3$	$p = 4$
d	$q = 2.1$	$q = 3$	$q = 4$	$q = 4.1$	$q = 5$	$q = 6$	$q = 7$	$q = 9$
1	323	38	22	698	74	40	183	453
2	—	—	—	2194	237	128	709	2360
3	119	57	41	7029	771	422	1933	7231
4	72	39	29	—	—	—	5026	18 478

Table D.1. Specific values of the products $2^p d_{\mathbb{K}}^{p/2}\kappa_{d_{\mathbb{K}},p}\theta_{d_{\mathbb{K}},p,q}$ multiplying the N-dependent terms in the first and third case of (D.20), rounded up.

For the remainder of this section, fix $p \in [1, \infty)$.

Lemma D.12 (Tightness of transfer plans [61, Lemma 4.4])**.** *Let X and Y be metric spaces. Let $\mathcal{Q}_X \subseteq \mathcal{P}(X)$ and $\mathcal{Q}_Y \subseteq \mathcal{P}(Y)$ be tight subsets of $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ respectively, and equip $X \times Y$ with the product σ-algebra $\mathcal{B}_X \otimes \mathcal{B}_Y$. Then the set $\mathcal{P}(\mathcal{Q}_X, \mathcal{Q}_Y)$ of all probability measures in $\mathcal{P}(X \times Y)$ with first marginal in $\mathcal{Q}_X$ and second marginal in $\mathcal{Q}_Y$ is tight in $\mathcal{P}(X \times Y)$.*

Lemma D.13 (Bound on the Wasserstein distance of weak limits, [61, Remark 6.12], [45, Equation (2.5)])**.** *Let (S, d) be a Polish space and $(\mu_n)_{n\in\mathbb{N}}$ and $(\nu_n)_{n\in\mathbb{N}}$ sequences in $\mathcal{P}_p(S)$ which converge weakly to probability measures μ and ν in $\mathcal{P}(S)$, respectively. Then*

$$\mathcal{W}_p(\mu, \nu) \le \liminf_{n\to\infty} \mathcal{W}_p(\mu_n, \nu_n). \tag{D.21}$$

Proof. [6]For each $n \in \mathbb{N}$, there exists by the definition (D.2) of the Wasserstein distance $\mathcal{W}_p$ a coupling $\pi_n \in \mathcal{P}(\mu_n, \nu_n)$ of μ_n and ν_n with

$$\left(\int_{S\times S} d(x, y)^p \,\mathrm{d}\pi_n(x, y)\right)^{1/p} \le \mathcal{W}_p(\mu_n, \nu_n) + \frac{1}{n}.$$

[6] I thank my advisor for suggesting a major simplification of the proof of Lemma D.13.

In particular, $\liminf_{n\to\infty}\left(\int_{S\times S} d(x,y)^p \,\mathrm{d}\pi_n(x,y)\right)^{1/p} \le \liminf_{n\to\infty} \mathcal{W}_p(\mu_n,\nu_n)$. The reverse inequality is trivial by (D.2), hence

$$\liminf_{n\to\infty}\left(\int_{S\times S} d(x,y)^p \,\mathrm{d}\pi_n(x,y)\right)^{1/p} = L, \tag{D.22}$$

where

$$L \coloneqq \liminf_{n\to\infty} \mathcal{W}_p(\mu_n,\nu_n). \tag{D.23}$$

By passing to a subsequence on the left-hand side in (D.22) and relabeling indices, we may assume

$$\lim_{n\to\infty}\left(\int_{S\times S} d(x,y)^p \,\mathrm{d}\pi_n(x,y)\right)^{1/p} = L. \tag{D.24}$$

Weak convergence of $(\mu_n)_{n\in\mathbb{N}}$ and $(\nu_n)_{n\in\mathbb{N}}$ implies using Prokhorov's theorem [53, Theorem 8.67] and Lemma D.12 tightness of $(\pi_n)_{n\in\mathbb{N}}$. By another application of Prokhorov's theorem, passing once more to a subsequence and relabeling, we may assume that $(\pi_n)_{n\in\mathbb{N}}$ converges weakly to a probability measure $\pi \in \mathcal{P}(S\times S)$. We next check that $\pi\in\mathcal{P}(\mu,\nu)$. Fix $f \in C_{\mathrm{b}}(S)$. Since π_n has first marginal μ_n for all $n\in\mathbb{N}$ and $\mu_n \xrightarrow[n\to\infty]{} \mu$ weakly, we obtain

$$\begin{aligned}\int_{S\times S} f(x)\,\mathrm{d}\pi(x,y) &= \lim_{n\to\infty}\int_{S\times S} f(x)\,\mathrm{d}\pi_n(x,y)\\ &= \lim_{n\to\infty}\int_S f(x)\,\mathrm{d}\mu_n(x) = \int_S f(x)\,\mathrm{d}\mu(x)\,.\end{aligned}$$

Since f was arbitrary, [5, Theorem 1.3] implies that $\pi(\,\cdot\,\times S) = \mu$. The same arguments establish that $\pi(S\times\,\cdot\,) = \nu$, hence $\pi\in\mathcal{P}(\mu,\nu)$.

For each $m\in\mathbb{N}$, $c_m(x,y) \coloneqq d(x,y)^p \wedge m$ defines a bounded continuous function on $S\times S$, and $c_m \nearrow d^p$ pointwise as $m\to\infty$. Using[7] that $\pi\in\mathcal{P}(\mu,\nu)$ and the definition (D.2) of the Wasserstein distance $\mathcal{W}_p$ in the first, monotone convergence in the second, weak convergence of $(\pi_n)_{n\in\mathbb{N}}$ to π in the third and finally (D.24) in the last line,

$$\begin{aligned}\mathcal{W}_p(\mu,\nu) &\le \left(\int_{S\times S} d(x,y)^p\,\mathrm{d}\pi(x,y)\right)^{1/p}\\ &= \lim_{m\to\infty}\left(\int_{S\times S} c_m(x,y)\,\mathrm{d}\pi(x,y)\right)^{1/p}\\ &= \lim_{m\to\infty}\lim_{n\to\infty}\left(\int_{S\times S} c_m(x,y)\,\mathrm{d}\pi_n(x,y)\right)^{1/p}\\ &\le \liminf_{n\to\infty}\left(\int_{S\times S} d(x,y)^p\,\mathrm{d}\pi_n(x,y)\right)^{1/p}\\ &= L.\end{aligned}$$

By (D.23) (wherein (μ_n) and (ν_n) refer to the original sequence), the claim follows. □

[7] Alternatively, we could conclude directly by the Portmanteau theorem, specifically the characterization of weak convergence using lower bounded lower semicontinuous functions, see [53, Theorem 8.50(h)].

In Section 3.1, see in particular (3.25), the following observation is of interest.

Example D.14 (The conditional expectation is not continuous w.r.t. the Wasserstein distance)**.** Consider the unit cube $S = [0,1]^2$ and denote by X and Y the canonical projections on the first and second coordinate respectively. We show that[8]

$$\mathcal{P}_p(S) \ni \pi \mapsto \mathbb{E}_\pi[Y|X=x] \in \mathbb{R},$$

where $x \in [0,1]$ is fixed, is not continuous w.r.t. the Wasserstein metric $\mathcal{W}_p$ on $\mathcal{P}_p(S)$. To see this, assume w.l.o.g. that $x=0$, let μ be the uniform distribution on $[0,1]$ and define

$$\pi_n \coloneqq \mu \otimes \kappa_n, \qquad n \in \mathbb{N}, \tag{D.25}$$

where $\otimes$ is interpreted in the sense of products of probability measures and stochastic kernels (see [36, Corollary 14.26]) and

$$\kappa_n(x,\,\cdot\,) \coloneqq \begin{cases} \delta_0 & \text{if } x \in [0,\frac{1}{n}], \\ \mu & \text{if } x \in (\frac{1}{n},1], \end{cases} \qquad n \in \mathbb{N},$$

with δ_0 the Dirac measure at zero in $[0,1]$. It is easy to check that each κ_n defines a stochastic kernel, hence π_n is a probability measure on S for each $n \in \mathbb{N}$. Let $\pi \coloneqq \mu^{\otimes 2}$ be the uniform distribution on S. By compactness of S, the map

$$S \ni (x,y) \mapsto |(x,y)|^p \in \mathbb{R}_+ \tag{D.26}$$

is in $C_{\mathrm{b}}(S)$, $\mathcal{P}_p(S) = \mathcal{P}(S)$, and, by Theorem D.6 (c) $\Rightarrow$ (a), weak convergence in $\mathcal{P}(S)$ and convergence w.r.t. $\mathcal{W}_p$ coincide. Hence it suffices to verify that

$$\pi_n \xrightarrow[n\to\infty]{} \pi \qquad \text{weakly}, \tag{D.27}$$

while $\mathbb{E}_{\pi_n}[Y|X=0]$ does not converge to $\mathbb{E}_\pi[Y|X=0]$ as $n \to \infty$. By (D.25), for every[9] $f \in C_{\mathrm{b}}(S)$,

$$\begin{aligned} \int_S f\,\mathrm{d}\pi_n &= \int_{[0,1]} \mu(\mathrm{d}x) \int_{[0,1]} \kappa_n(x,\mathrm{d}y) f(x,y) \\ &= \int_{[0,\frac{1}{n}]} \mu(\mathrm{d}x) f(x,0) + \int_{(\frac{1}{n},1]} \mu(\mathrm{d}x) \int_{[0,1]} \mu(\mathrm{d}y) f(x,y) \xrightarrow[n\to\infty]{} \int_S f\,\mathrm{d}\pi\,, \end{aligned}$$

where the convergence is by boundedness of f and the dominated convergence theorem, thus establishing (D.27). Finally, we have (see [36, Theorem 8.38])

$$\mathbb{E}_{\pi_n}[Y|X=0] = \mathbb{E}_{\kappa_n(0,\,\cdot\,)}[Y] = 0, \qquad n \in \mathbb{N}.$$

On the other hand, since X and Y are independent w.r.t. π and $\mathcal{L}_\pi[Y] = \mu$,

$$\mathbb{E}_\pi[Y|X=0] = \mathbb{E}_\mu[Y] = \frac{1}{2}.$$

[8] The conditional expectation $x \mapsto \mathbb{E}[Y|X=x]$ can be defined as a factorizing map in the sense of Lemma B.8, see also Footnote 21 on Page 59.

[9] When integrating against stochastic kernels, we use the notation $\kappa(x,\mathrm{d}y)$ instead of $\mathrm{d}\kappa(x,y)$ to highlight the variable of integration.

D.2. Measurability and Adaptedness of Random Measures

We conclude this appendix with some measurability results relating to random and in particular empirical measures, which appear in the particle approximation method discussed in Subsection 2.4.2. Fix $p \in [1, \infty)$ and recall the spaces $W_\ell = C(\mathbb{R}_+, \mathbb{K}^\ell)$ for $\ell \in \mathbb{N}$ and their canonical σ-algebras and filtrations introduced at the beginning of Section 2.1 as well as the filtration $\mathbb{H} = (\mathcal{H}_t)_{t \in \mathbb{R}_+}$ in $\widehat{\mathcal{P}}_p(W_n)$ from Definition 2.8.

Lemma D.15 (The weak and Wasserstein topologies induce the same Borel σ-algebra on $\mathcal{P}_p(S)$)**.** *Let (S, d) be a Polish space. Then the Borel σ-algebras of the weak and Wasserstein topologies on $\mathcal{P}_p(S)$ agree.*

Proof [45, Lemma 2.4.2]. We give two proofs, one based on abstract measure theoretic results, and one elementary.

(a) [45, Lemma 2.4.2] Since S is Polish, the weak topology $\mathcal{O}^{\text{weak}}_{\mathcal{P}(S)}$ on $\mathcal{P}(S)$ is metrizable by [4, Theorem 7.20]. Metrizability of the weak topology is inherited by the subspace topology $\mathcal{O}^{\text{weak}}_{\mathcal{P}_p(S)} := \mathcal{P}_p(S) \cap \mathcal{O}^{\text{weak}}_{\mathcal{P}(S)}$. In particular, $\mathcal{O}^{\text{weak}}_{\mathcal{P}_p(S)}$ is determined by its convergent sequences. Denote by $\mathcal{O}_{\mathcal{P}_p(S)}$ the topology induced on $\mathcal{P}_p(S)$ by the Wasserstein metric $\mathcal{W}_p$, which is Polish by Theorem D.8. By the implication (a) ⇒ (b) of Theorem D.6, convergence w.r.t. $\mathcal{W}_p$ implies weak convergence, hence $\mathcal{O}^{\text{weak}}_{\mathcal{P}_p(S)} \subseteq \mathcal{O}_{\mathcal{P}_p(S)}$. Since $\mathcal{O}_{\mathcal{P}_p(S)}$ is Polish and $\mathcal{O}^{\text{weak}}_{\mathcal{P}_p(S)}$ is Hausdorff, the equality $\sigma(\mathcal{O}^{\text{weak}}_{\mathcal{P}_p(S)}) = \sigma(\mathcal{O}_{\mathcal{P}_p(S)})$ of the generated σ-algebras follows from [24, 423B(a) and 423F(b)].

(b) [46, p. 11][10] The inclusion $\sigma(\mathcal{O}^{\text{weak}}_{\mathcal{P}_p(S)}) \subseteq \sigma(\mathcal{O}_{\mathcal{P}_p(S)})$ follows as in proof (a) above. For the reverse inclusion, note first that $\mathcal{O}_{\mathcal{P}_p(S)}$ is Polish by Theorem D.8 and thus generated by a countable family of $\mathcal{W}_p$-open balls. Since every open ball in a metric space is a countable union of closed balls, it suffices to show that every $\mathcal{W}_p$-closed ball is weakly closed[11] and thus contained in the Borel σ-algebra $\sigma(\mathcal{O}^{\text{weak}}_{\mathcal{P}_p(S)})$ of the weak topology. To this end, fix $\varepsilon > 0$ and $\mu \in \mathcal{P}_p(S)$, let $K_\varepsilon(\mu) := \{\nu \in \mathcal{P}_p(S) : \mathcal{W}_p(\mu, \nu) \le \varepsilon\}$ be the closed ball of radius ε and center μ w.r.t. $\mathcal{W}_p$, and let $(\nu_n)_{n \in \mathbb{N}}$ be a sequence[12] in $K_\varepsilon(\mu)$ converging weakly to a limit $\nu \in \mathcal{P}(S)$. From Lemma D.13, it follows that

$$\mathcal{W}_p(\mu, \nu) \le \liminf_{n \to \infty} \mathcal{W}_p(\mu, \nu_n) \le \varepsilon,$$

hence $\nu \in K_\varepsilon(\mu)$. Therefore, $K_\varepsilon(\mu)$ is weakly closed, (i.e. closed in $\mathcal{O}^{\text{weak}}_{\mathcal{P}_p(S)}$), proving the claim. □

Lemma D.16 (Measurability of $\mathcal{P}_p(S)$-valued random measures)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space, (S, d) a Polish space and $\mathcal{E} \subseteq \mathcal{B}_S$ an intersection stable generator of $\mathcal{B}_S$. Consider a map $\mu\colon \Omega \to \mathcal{P}_p(S)$, $\omega \mapsto \mu_\omega$. Then the following are equivalent:*

[10] In [46, p. 11], $\mathcal{P}_p(S)$ is claimed to be closed in the topology of weak convergence. This is false, since the set of finite convex combinations of Dirac measures is dense in the weak topology, see the proof of [53, Theorem 8.61], and every such convex combination has finite moments of all orders. However, as shown in proof (b) above, weak closedness of $\mathcal{P}_p(S)$ is not needed for the proof to go through.

[11] By weakly closed, we mean closed w.r.t. $\mathcal{O}^{\text{weak}}_{\mathcal{P}_p(S)}$.

[12] We can consider only sequences instead of nets since the weak topology on $\mathcal{P}(S)$ is metrizable by [4, Theorem 7.20], and metrizability is inherited to subspaces.

(a) *The map μ is $\mathcal{F}$-$\mathcal{B}_{\mathcal{P}_p(S)}$-measurable.*

(b) *For every $A \in \mathcal{E}$, the map $\Omega \ni \omega \mapsto \mu_\omega(A) \in [0,1]$ is $\mathcal{F}$-$\mathcal{B}_{[0,1]}$-measurable.*

Proof. By [4, Proposition 7.26], $\mathcal{F}$-$\mathcal{B}_{[0,1]}$-measurability of $\Omega \ni \omega \mapsto \mu_\omega(A) \in [0,1]$ for every $A \in \mathcal{E}$ is equivalent to measurability of $\Omega \ni \omega \mapsto \mu_\omega \in \mathcal{P}(S)$ w.r.t. the σ-algebra $\mathcal{B}_{\mathcal{P}(S)} := \sigma(\mathcal{O}^{\text{weak}}_{\mathcal{P}(S)})$ induced by the weak topology on $\mathcal{P}(S)$. Since $\omega \mapsto \mu_\omega$ is $\mathcal{P}_p(S)$-valued by assumption, measurability w.r.t. $\mathcal{B}_{\mathcal{P}(S)}$ is equivalent to measurability w.r.t. the subspace σ-algebra $\mathcal{P}_p(S) \cap \mathcal{B}_{\mathcal{P}(S)} = \sigma(\mathcal{O}^{\text{weak}}_{\mathcal{P}_p(S)})$. From Lemma D.15, $\sigma(\mathcal{O}^{\text{weak}}_{\mathcal{P}_p(S)})$ coincides with the Borel σ-algebra $\sigma(\mathcal{O}_{\mathcal{P}_p(S)})$ on $\mathcal{P}_p(S)$ induced by the Wasserstein metric, proving the claim. □

Lemma D.17 (Adaptedness of empirical measures)**.** *Fix $N \in \mathbb{N}$ and equip $W_{nN} = C(\mathbb{R}_+, \mathbb{K}^{nN})$ with the filtration $\mathbb{G}^{nN} = (\mathcal{G}^{nN}_t)_{t \in \mathbb{R}_+}$ generated by the canonical projections (see* (2.5)*).*

(a) *For each $t \in \mathbb{R}_+$,*

$$\hat{\mu}\colon W_{nN} \to \widehat{\mathcal{P}}_p(W_n), \quad w = (w^1, \dots, w^N) \mapsto \frac{1}{N} \sum_{i=1}^{N} \delta_{w^i} \tag{D.28}$$

is $\mathcal{G}^{nN}_t$-$\mathcal{H}_t$-measurable.

(b) *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$ be a filtered probability space. Let $X^1, \dots, X^N$ be $\mathbb{K}^n$-valued $\mathbb{F}$-adapted stochastic processes with continuous sample paths. Then the empirical measure*

$$\tilde{\mu}\colon \Omega \to \widehat{\mathcal{P}}_p(W_n), \quad \omega \mapsto \frac{1}{N} \sum_{i=1}^{N} \delta_{X^i(\omega)} \tag{D.29}$$

is $\mathcal{F}_t$-$\mathcal{H}_t$-measurable for each $t \in \mathbb{R}_+$.

Proof. Fix $t \in \mathbb{R}_+$.

(a) Since $\mathcal{H}_t = \Pi^{-1}_{[0,t]}(\mathcal{B}_{\mathcal{P}_p(W^t_n)})$ by Remark 2.9 with $\Pi_{[0,t]}$ defined in (2.15), it suffices to show that $\Pi_{[0,t]} \circ \hat{\mu}$ is $\mathcal{G}^{nN}_t$-$\mathcal{B}_{\mathcal{P}_p(W^t_n)}$-measurable. We have

$$\big(\Pi_{[0,t]} \circ \hat{\mu}\big)(w) = \hat{\mu}(w) \circ \pi^{-1}_{[0,t]} = \frac{1}{N} \sum_{i=1}^{N} \delta_{w^i\restriction_{[0,t]}}, \qquad w \in W_{nN}, \tag{D.30}$$

since for all $A \in \mathcal{B}_{W^t_n}$ and all $\tilde{w} \in W_n$,

$$\delta_{\tilde{w}}(\pi^{-1}_{[0,t]}(A)) = \mathbb{1}_{\pi^{-1}_{[0,t]}(A)}(\tilde{w}) = \mathbb{1}_A(\tilde{w}\restriction_{[0,t]}) = \delta_{\tilde{w}\restriction_{[0,t]}}(A).$$

The $\mathcal{G}^{nN}_t$-$\mathcal{B}_{\mathcal{P}_p(W^t_n)}$-measurability of the right-hand side of (D.30) is by Lemma D.16 equivalent to

$$W_{nN} \ni w \mapsto \frac{1}{N} \sum_{i=1}^{N} \delta_{w^i\restriction_{[0,t]}}(A) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_A(w^i\restriction_{[0,t]}) \tag{D.31}$$

being $\mathcal{G}_t^{nN}$-measurable for all $A \in \mathcal{B}_{W_n^t}$, which we show next. Let $\hat{p}_t$ denote the seminorm on W_{nN} defined in complete analogy to the seminorm p_t on W_n from (2.1). For each $i \in \{1,\dots,N\}$, $W_{nN} \ni w \mapsto w^i \in W_n$ is $\hat{p}_t$-p_t-continuous and thus $\mathcal{G}_t^{nN}$-$\mathcal{G}_t$-measurable by Lemma 2.1 (c), and from this same result it follows further that $W_{nN} \ni w \mapsto \pi_{[0,t]}(w^i) = w^i\restriction_{[0,t]} \in W_n^t$ is $\mathcal{G}_t^{nN}$-$\mathcal{B}_{W_n^t}$-measurable. This shows the claimed measurability of the map in (D.31), from which we conclude.

(b) Define $X\colon \Omega \to W_{nN}$, $\omega \mapsto (X^1(\omega),\dots,X^N(\omega))$, and observe that

$$\tilde{\mu} = \hat{\mu} \circ X.$$

Therefore, it suffices to show that X is $\mathcal{F}_t$-$\mathcal{G}_t^{nN}$-measurable. This is indeed the case, since $\mathcal{G}_t^{nN} = \sigma(\tilde{\pi}_s : s \in [0,t])$, where $\tilde{\pi}_s(w) := w_s$ for all $w \in W_{nN}$ and $s \in [0,t]$, and $\tilde{\pi}_s \circ X = (X_s^1,\dots,X_s^N)$ is clearly a $\mathbb{K}^{nN}$-valued $\mathcal{F}_t$-measurable random variable. □

For the particle approximation (Definition 2.51) of the McKean SDE (2.36), the following result is useful.

Lemma D.18. *In the setting of Lemma D.17, let $f\colon \mathbb{R}_+ \times W_n \times \widehat{\mathcal{P}}_p(W_n) \to \mathbb{K}^m$ be $\mathbb{G}\otimes\mathbb{H}$-progressive (Definition 2.11 (a)). Then for each $i \in \{1,\dots,N\}$, the map $\tilde{f}_i\colon \mathbb{R}_+ \times W_{nN} \to \mathbb{K}^m$,*

$$\tilde{f}_i(t,w) := f(t,w^i,\hat{\mu}(w)), \qquad t \in \mathbb{R}_+, w = (w^1,\dots,w^N) \in W_{nN},$$

is $\mathbb{G}^{nN}$-progressive.

Proof. We verify that for all fixed $i \in \{1,\dots,N\}$ and $t \in \mathbb{R}_+$, $\tilde{f}_i\restriction_{[0,t]\times W_{nN}}$ is $\mathcal{B}_{[0,t]} \otimes \mathcal{G}_t^{nN}$-measurable. Note that

$$\tilde{f}_i\restriction_{[0,t]\times W_{nN}} = f\restriction_{[0,t]\times W_n \times \widehat{\mathcal{P}}_p(W_n)} \circ \varphi,$$

where $\varphi\colon [0,t] \times W_{nN} \to [0,t] \times W_n \times \widehat{\mathcal{P}}_p(W_n)$ is given by

$$\varphi(t,w) := (t,w^i,\hat{\mu}(w)), \qquad (t,w) \in \mathbb{R}_+ \times W_{nN}.$$

Since f is $\mathbb{G}\otimes\mathbb{H}$-progressive, the restriction $f\restriction_{[0,t]\times W_n\times\widehat{\mathcal{P}}_p(W_n)}$ is $\mathcal{B}_{[0,t]}\otimes\mathcal{G}_t\otimes\mathcal{H}_t$-measurable. Hence it suffices to verify that φ is measurable w.r.t. $(\mathcal{B}_{[0,t]}\otimes\mathcal{G}_t^{nN})$-$(\mathcal{B}_{[0,t]}\otimes\mathcal{G}_t\otimes\mathcal{H}_t)$. Clearly, the first component map $(t,w)\mapsto t$ of φ has the required measurability, and the same is true for the component map $W_{nN} \ni (t,w) \mapsto w^i \in W_n$, as shown in the proof of Lemma D.17 (a). The same lemma implies the required measurability for the third component map $(t,w)\mapsto \hat{\mu}(w)$ of φ. □

Lemma D.19. *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $\mathbb{Q},\tilde{\mathbb{Q}}\colon \Omega \to \widehat{\mathcal{P}}_p(W_n)$ be maps.*

(a) Fix $t \in \mathbb{R}_+$ and assume that $\mathbb{Q}$ and $\tilde{\mathbb{Q}}$ are $\mathcal{F}$-$\mathcal{H}_t$-measurable (see Definition 2.8). Then

$$\Omega \ni \omega \mapsto d_p(\mathbb{Q}_t(\omega),\tilde{\mathbb{Q}}_t(\omega)) \in \mathbb{R}_+$$

(see Definition 2.4) and

$$\Omega \ni \omega \mapsto \widehat{\mathcal{W}}_{t,p}(\mathbb{Q}(\omega),\tilde{\mathbb{Q}}(\omega)) \in \mathbb{R}_+$$

(Definition 2.5) are $\mathcal{F}$-measurable, i.e. random variables.

(b) Assume that $\mathbb{Q}, \tilde{\mathbb{Q}}$ are $\mathcal{F}$-$\mathcal{H}_t$-measurable for each $t \in \mathbb{R}_+$. Then

$$\begin{aligned} &\mathbb{R}_+ \ni t \mapsto \mathbb{E}[d_p^p(\mathbb{Q}_t, \tilde{\mathbb{Q}}_t)] \in \overline{\mathbb{R}}_+ \quad \textit{and} \\ &\mathbb{R}_+ \ni t \mapsto \mathbb{E}\big[\widehat{\mathcal{W}}_{t,p}^p(\mathbb{Q}, \tilde{\mathbb{Q}})\big] \in \overline{\mathbb{R}}_+ \end{aligned} \tag{D.32}$$

are $\mathcal{B}_{\mathbb{R}_+}$-$\mathcal{B}_{\overline{\mathbb{R}}_+}$-measurable.

Proof. (a) The first map, $d_p(\mathbb{Q}_t, \tilde{\mathbb{Q}}_t)$, is the composition of the p-th order Wasserstein metric d_p on $\mathcal{P}_p(\mathbb{K}^n)$ and

$$\varphi\colon \Omega \ni \omega \mapsto (\Pi_t(\mathbb{Q}(\omega)), \Pi_t(\tilde{\mathbb{Q}}(\omega))) \in \mathcal{P}_p(\mathbb{K}^n) \times \mathcal{P}_p(\mathbb{K}^n),$$

with $\Pi_t\colon \widehat{\mathcal{P}}_p(W_n) \to \mathcal{P}_p(\mathbb{K}^n)$ defined in (2.16). By Theorem D.8, $(\mathcal{P}_p(\mathbb{K}^n), d_p)$ is Polish, hence d_p is $\mathcal{B}_{\mathcal{P}_p(\mathbb{K}^n)} \otimes \mathcal{B}_{\mathcal{P}_p(\mathbb{K}^n)}$-measurable ([53, Theorem 15.53]). By Lemma 2.10, Π_t is $\mathcal{H}_t$-$\mathcal{B}_{\mathcal{P}_p(\mathbb{K}^n)}$-measurable. Thus, using that $\mathbb{Q}$ and $\tilde{\mathbb{Q}}$ are $\mathcal{F}$-$\mathcal{H}_t$-measurable by assumption, each component map of φ is $\mathcal{F}$-$\mathcal{B}_{\mathcal{P}_p(\mathbb{K}^n)}$-measurable. Therefore, φ is $\mathcal{F}$-$(\mathcal{B}_{\mathcal{P}_p(\mathbb{K}^n)} \otimes \mathcal{B}_{\mathcal{P}_p(\mathbb{K}^n)})$-measurable, which implies the claimed measurability of $d_p(\mathbb{Q}_t, \tilde{\mathbb{Q}}_t)$.

Using (2.21), the second map, $\widehat{\mathcal{W}}_{t,p}(\mathbb{Q}, \tilde{\mathbb{Q}})$, is the composition $\mathcal{W}_{t,p} \circ \psi$ of the Wasserstein metric $\mathcal{W}_{t,p}$ on $\mathcal{P}_p(W_n^t)$ and

$$\psi\colon \Omega \ni \omega \mapsto (\Pi_{[0,t]}(\mathbb{Q}(\omega)), \Pi_{[0,t]}(\tilde{\mathbb{Q}}(\omega))) \in \mathcal{P}_p(W_n^t) \times \mathcal{P}_p(W_n^t),$$

with $\Pi_{[0,t]}\colon \widehat{\mathcal{P}}_p(W_n) \to \mathcal{P}_p(W_n^t)$ defined in (2.15). By Theorem D.8, $(\mathcal{P}_p(W_n^t), \mathcal{W}_{t,p})$ is Polish, hence $\mathcal{W}_{t,p}$ is $\mathcal{B}_{\mathcal{P}_p(W_n^t)} \otimes \mathcal{B}_{\mathcal{P}_p(W_n^t)}$-measurable ([53, Theorem 15.53]). By Remark 2.9, $\mathcal{H}_t = \Pi_{[0,t]}^{-1}(\mathcal{B}_{\mathcal{P}_p(W_n^t)})$, hence $\Pi_{[0,t]}$ is $\mathcal{H}_t$-$\mathcal{B}_{\mathcal{P}_p(W_n^t)}$-measurable. From $\mathcal{H}_t$-measurability of $\mathbb{Q}$ and $\tilde{\mathbb{Q}}$, each component map of ψ is $\mathcal{F}$-$\mathcal{B}_{\mathcal{P}_p(W_n^t)}$-measurable. Therefore, ψ is $\mathcal{F}$-$(\mathcal{B}_{\mathcal{P}_p(W_n^t)} \otimes \mathcal{B}_{\mathcal{P}_p(W_n^t)})$-measurable, which implies the claim.

(b) The expectations in (D.32) exist (in $\overline{\mathbb{R}}_+$) due to part (a). By Lemma 2.10, $\mathbb{R}_+ \ni t \mapsto d_p(\mathbb{Q}_t(\omega), \tilde{\mathbb{Q}}_t(\omega)) \in \mathbb{R}_+$ is continuous for all $\omega \in \Omega$. For each $K \in \mathbb{R}_+$, dominated convergence implies continuity of

$$\mathbb{R}_+ \ni t \mapsto \mathbb{E}[d_p^p(\mathbb{Q}_t, \tilde{\mathbb{Q}}_t) \wedge K] \in \mathbb{R}_+,$$

which by taking $K \to \infty$ and using monotone convergence yields Borel-measurability of the first map in (D.32). The second map, $\mathbb{R}_+ \ni t \mapsto \mathbb{E}\big[\widehat{\mathcal{W}}_{t,p}^p(\mathbb{Q}, \tilde{\mathbb{Q}})\big]$, is clearly measurable as it is increasing by Remark 2.6 (e). □

E. Main Findings of the Thesis

This section lists the major and minor findings of this thesis, grouped into those results which are mainly literature-based and those which were developed independently of or extending the literature. The division is complicated by the fact that many of the findings appearing in this text are improvements or generalizations of existing results. In such cases, we list the corresponding item in the more appropriate section below, but indicate the nature of the improvement as well as the primary source(s) in a footnote.

Main Findings Based on the Literature

(a) Section 2.1 up to Definition 2.2 is based on the literature cited therein, in particular [53, §8.1].

(b) The examples of McKean SDEs in Section 2.3 are from [10, §1.3.2].[1]

(c) Subsection 2.4.1, which introduces chaoticity of probability measures, is based on [26, §10.3.2].[2]

(d) Section 3.1 (on the calibration of local stochastic volatility models) is largely based on the literature, mainly [26, §11].[3] The central result here is Proposition 3.5, whose heuristic proof along the lines of [26, Proposition 11.1] we describe in some detail.

(e) Subsections 3.2.1 and 3.2.3 (on the classical and mean-field LIBOR models) is based mostly on the references cited therein, in particular [17].

(f) Appendix A (on pseudometric spaces) is with the exceptions of Lemma A.11, Theorem A.16 and Example A.17, based on the cited references.

(g) Appendix B (background from measure and probability theory) is based on the references cited therein.[4] Appendix B.2 is based mainly on [53, §15.6.4], with the exception of the proof (but not the content) of Proposition B.20 as well as Corollary B.21.

[1] Lemma 2.39 and Lemma 2.41 verify that under Lipschitz continuity and light additional measurability assumptions, the coefficients associated with the examples in Section 2.3 have the progressive measurability properties required for coefficients of McKean SDEs. These lemmas have been developed independently of the literature.

[2] Compared to [26, §10.3.2], the definitions, results and proofs in Subsection 2.4.1 have been extended from $\mathbb{R}^n$ to a general metric space. Furthermore, the proof of Lemma 2.48 has been clarified.

[3] In the next section below, we list those results in Section 3.1 developed in this thesis.

[4] A subset of elementary measure-theoretic results were developed independently of the literature in Appendix B for later use in Appendix C.4, e.g. Lemma B.5, Corollary B.6 and Lemma B.7. Results of this type are well-known, so we do not list them in the next section.

(h) Appendices C.1 and C.2 contain material on classical SDEs, based mainly on [53, §8] and the other references cited.[5]

(i) Appendix C.4 is devoted to the topic of flows of classical SDEs, based mainly on [38] and [51].[6]

(j) Appendix D.1 contains background on Wasserstein spaces, based mainly on [60], [61] and [45].[7]

Main Findings Developed Independently of or Extending the Literature

(a) The space $\widehat{\mathcal{P}}_p(W_n)$ (Definition 2.2), its (pseudo-)metrization (Definition 2.5), the resulting properties (Remark 2.6), in particular $\widehat{\mathcal{P}}_p(W_n)$ being a Polish space w.r.t. the topology induced by the family $(\widehat{\mathcal{W}}_{k,p})_{k\in\mathbb{N}}$ of pseudometrics (Proposition 2.7).

(b) The σ-algebra and filtration on $\widehat{\mathcal{P}}_p(W_n)$, their relation to weak topologies and further properties (Definition 2.8, Remark 2.9 and Lemma 2.10). The resulting filtrations (Definition 2.11) on $W_n \times \widehat{\mathcal{P}}_p(W_n)$ and $W_n \times \widehat{\mathcal{P}}_p(W_n) \times \Omega$, used to define the measurability properties for coefficients which later ensure the well-posedness of McKean SDEs.

(c) The variants (2.34) and (2.35) of McKean SDEs considered are to our knowledge more general than those appearing in the literature,[8] as we consider simultaneously

   (i) the infinite time horizon,

   (ii) the cases of real- and complex-valued driving Brownian motions and coefficients,

   (iii) dependence of coefficients on time, the past trajectory of the solution, the law of the past trajectory of the solution and additional randomness subject to an adaptedness condition,

   (iv) the $\mathcal{L}^p$-case for $p \in [2,\infty)$ (rather than the more common $\mathcal{L}^2$-case), i.e. p-integrable initial conditions and SDE coefficients depending on distributions with finite p-th moments satisfying a Lipschitz continuity condition involving the p-th order Wasserstein distance.

(d) A number of elementary measurability results in Section 2.2, in particular Lemma 2.15[9] and Remarks 2.16, 2.17, 2.18 and 2.24, which ensure that the stochastic integrals associated with McKean SDEs are well-defined and establish connections to classical SDEs.

[5] We extended the treatment of classical SDEs in [53] to allow for random dependence. The relevant results are listed in the next section.

[6] Compared to our sources, we establish the main result, Theorem C.43, in a more general form (e.g., we consider both $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$, progressive instead of predictable SDE coefficients and a relaxed completeness condition on the underlying filtration). See Remark C.44 for details.

[7] The exceptions are the proof of Lemma D.13 and Example D.14, which were developed in this thesis.

[8] Individually, several of the above extensions of the originally introduced McKean SDEs have of course already appeared in the literature, see also Remark 2.14.

[9] Lemma 2.15 extends Lemma C.7, which itself is based on [53, Lemma 8.2].

(e) The Lipschitz property (Definition 2.19) and linearly bounded growth condition (Definition 2.20) as well as the instantaneous Lipschitz property (Definition 2.22) are generalized from [10, §1.3][10] to the infinite time horizon and dependence on the law of the entire solution process instead of merely the one-dimensional marginals.[11] Associated are Lemma 2.21 and Remark 2.23, which show that the Lipschitz property allows to relax the measurability required from the coefficients.

(f) Lemma 2.31, used to prove the main existence and uniqueness result for McKean SDEs (Theorem 2.32) as well as propagation of chaos, based on ideas from the proofs of [53, Theorem 8.27] and [33, Proposition 2].

(g) Theorem 2.32, a main result in this thesis, extends (see Point (c) above) the classical strong existence and pathwise uniqueness result for McKean SDEs appearing in [10, §1.3.1] and [33, Proposition 2].[12] Corollary 2.34, whereby a solution of the McKean SDE can be obtained as a limit of solutions of classical SDEs.[13] Remark 2.36, which provides an explicit bound for the p-th moment of the supremum norm on finite time horizons of (increments of) the solution of the McKean SDE in terms of the L^p-norm of the initial condition and the bounding functions from the Lipschitz and bounded growth properties.

(h) Subsection 2.4.2 (the particle approximation) extends known results regarding the convergence of the particle approximation ([10, §1.3.4] and [33, §1.1]) to coefficients with pathwise dependence. Remark 2.53 establishes that the particle approximation (Definition 2.51) can, even in the pathwise setting, be viewed as a (well-defined) classical SDE. Theorem 2.56, the second main result in this thesis, establishes the convergence (see (2.107)) of the particle approximation in the pathwise setting.[14] Theorem 2.58 gives an explicit bound of the rate of convergence of the particle approximation in the setting of instantaneous dependence.[15] Theorem 2.60, generalizing [26, Theorem 10.3], establishes the pathwise propagation of chaos for McKean SDEs with pathwise, nonrandom dependence.[16]

(i) Remark 3.11, which shows that a literal application of the particle method to the McKean SDE (3.30) for the LSVM produces under mild additional assumptions, counterintuitively, particles following independently local volatility dynamics.

(j) Subsection 3.2.2 generalizes [17, Theorem 5.2], an existence and uniqueness result for McKean SDEs with coefficients of a more general form than in Theorem 2.32, allowing

[10] The definitions in [10, §1.3] are for coefficients with instantaneous dependence (Definition 2.13(a)).

[11] The inspiration for the penalization was [53, §8]. Roughly similar conditions appear in [2, Assumption (I)] and [13, Assumption ($\boldsymbol{A}_{A,b,\sigma}$)], although differences arise due to the different settings.

[12] The fixed point argument employed in the proof of Theorem 2.32 ultimately traces back to [56, Theorem 1.1]. The treatment of the infinite time horizon using Proposition 2.7 and Theorem A.16 was developed in this thesis.

[13] Corollary 2.34 can be seen as generalizing a result in McKean's original paper [40], see Remark 2.35.

[14] The proof of Theorem 2.56 extends ideas from the proofs of [10, Theorem 1.10] and [33, Theorem 3], which apply to the setting of instantaneous dependence.

[15] Theorem 2.58 extends [33, Theorem 3], a similar result in the setting $\mathbb{K} = \mathbb{R}$, a finite time horizon, $p = 2$ and with a less explicit bound.

[16] The proof of Theorem 2.60 is different from and simplified compared to that of [26, Theorem 10.3].

for a measure dependence which is of a local rather than global Lipschitz type (see (3.55) and (3.58)). The result is obtained in [17] in the setting of instantaneous dependence, on a finite time horizon, with $\mathbb{K} = \mathbb{R}$ and the Wasserstein order $p = 2$, and extended in Theorem 3.19 to more general coefficients and the setting described in part (c) above.

The remainder of Subsection 3.2.2 is, beginning with Example 3.21, devoted to the study of a particular family of SDEs made accessible by Theorem 3.19, which is characterized by variance-based volatility dampening. This example, which is inspired by [17, §2.1], is discussed in detail regarding both theoretical (Lemmas 3.22 and 3.23) and practical (Examples 3.24 and 3.26) aspects. In particular, we obtain in Example 3.26 the probability distribution of the forward LIBORs (under their respective forward measures) in the mean-field LMM.

(k) Lemma A.11, extending a classical fixed point result from [8] to pseudometric spaces, the fixed point theorem Theorem A.16 for topologies induced by families of pseudometrics,[17] and the associated counterexample Example A.17.

(l) The proof of Proposition B.20 (an extension result for $\mathcal{P}(W_n)$) using compact classes[18] as well as Corollary B.21 were developed in this thesis.[19]

(m) Appendices C.1 and C.2, which assemble the required background on classical SDEs, extend the theory from [53, §8] to classical SDEs with random dependence. In particular Lemma C.7(b), Example C.24, Lemma C.29, Lemma C.31, Lemma C.32 and Theorem C.34.

(n) Appendix C.3, which illustrates how to switch from an SDE driven by correlated Brownian motions to an SDE driven by a multidimensional standard Brownian motion, plugging a gap between the classical SDEs presented in Appendix C.1 and some applications in mathematical finance, see also Remark 3.10.

(o) Appendix D.2, in particular Lemmas D.16 to D.19 therein, which establish measurability and adaptedness results for random measures, ensuring e.g. well-definedness of the particle approximation (Definition 2.51).

[17] I thank my advisor for suggesting some of the main features of Theorem A.16. An analogous result in the setting of uniformizable spaces is [57, Theorem 1.1].

[18] The result in Proposition B.20 is well-known, see the discussion preceding the proposition on Page 113.

[19] We suspect that the proof of Proposition B.20 using compact classes is known, but have not found a reference.

F. Open Questions and Ideas

(a) An intriguing open question is the rate of convergence of the particle approximation in the setting of Theorem 2.56. That is, the goal is to establish an analogue of Theorem 2.58 for coefficients with the Lipschitz property but without the instantaneous Lipschitz property (Definitions 2.19 and 2.22). By Remark 2.57, the missing link here is a result on the rate of convergence in Theorem D.9 when the underlying space is $S = W_n^T$, equipped with its associated Wasserstein metric, $\mathcal{W}_{T,p}$ (cf. Theorems D.10 and D.11).[1]

(b) The rate of convergence of the particle approximation given in Theorem 2.58(b) is explicit in the sense that it contains no unknown constants. However, even for small times, say $t = 10^{-3}$, the factor L_t multiplying the N-dependent term in (2.125) is large and furthermore grows rapidly in t, see Figure 2.1. The source of this blowup is the rapid growth in the time argument of the right-hand side in the moment bound (2.82) for solutions of McKean (but ultimately, see Remark C.35, classical) SDEs, as well as the appearance of the exponential function in (2.120) due to the application of Grönwall's inequality. The question here is to which degree these moment bounds can be improved, which would yield corresponding improvements in the rate of convergence of the particle approximation.

(c) We established in Theorem 2.56 the convergence of the particle approximation in the setting of nonrandom dependence, i.e. for coefficients as in (2.34). We restricted to nonrandom dependence since Theorem 2.32 guarantees weak uniqueness in this setting, which was used in turn to obtain exchangeability of the family $((X_i, X^{i,N}))_{i\in\{1,\dots,N\}}$ in Lemma 2.55.[2] For coefficients with random dependence for which this exchangeability property can still be established, the convergence of the particle approximation should hold as well. Thus, the question is which (useful) conditions on the form of the random dependence ensure the required exchangeability property.

(d) We have considered in this thesis only SDEs for $\mathbb{K}^n$-valued solutions driven by standard Brownian motion, with coefficients satisfying Lipschitz continuity and bounded growth conditions. Possible extensions involve matrix-valued solutions or solutions taking values in Hilbert or Banach spaces, with locally-Lipschitz coefficients (and, correspondingly, solutions only defined up to predictable stopping times) and SDEs driven by general continuous or càdlàg semimartingales. See also [53, Remark 8.5] and [54, §3].

(e) According to Theorem 3.19, there is, under Assumption 3.14, strong existence and pathwise uniqueness for McKean SDEs with coefficients as in (3.55), which includes in

[1] If the coefficients in (2.36) have the instantaneous Lipschitz property and depend only on a finite number of past marginal distributions, the arguments in Theorem 2.58 carry through with minor modifications and yield an explicit rate of convergence for this case.

[2] See also Footnote 32.

particular the class of McKean SDEs with variance-based volatility dampening (see Example 3.21). This form of the coefficients causes (see Remark 3.20) obstructions in the proof of Theorem 2.56 (convergence of the particle approximation). Nevertheless, we conjecture that the particle approximation converges also for SDEs of this type, and the open question is how this can be shown.

Conventions, Abbreviations, Symbols and Notation

Conventions

- We use *greater, less, increasing and decreasing* in the weak sense and use *strictly greater, strictly less, strictly increasing and strictly decreasing*, when excluding equality.
- We use *positive* synonymously with *non-negative*, *negative* synonymously with *non-positive* and write *strictly positive* and *strictly negative* to exclude zero. This convention does not apply to *positive definite* matrices, where we used *positive semidefinite* to admit a zero eigenvalue.
- When writing displayed formulas, we follow in most cases the custom to omit the words *for all* or the symbol $\forall$ when quantifying variables.
- When referring to displayed formulas, we usually omit the word *Equation* or its abbreviation *Eq.* for brevity.
- The union over an empty index set is the empty set: $\bigcup_{i\in\varnothing} A_i \coloneqq \varnothing$.
- The intersection over an empty index set is the full set (which depends on the context).
- The sum over an empty index set equals the neutral element of addition.
- The product over an empty index set equals the neutral element of multiplication.
- The conventions for the infimum and supremum over the empty set are $\inf\varnothing = \infty$ and $\sup\varnothing = -\infty$, respectively, unless otherwise noted.
- For every topological space (E,τ), we denote by $\mathcal{B}_E$ its Borel σ-algebra. When E is considered as a measurable space without specification of the σ-algebra, we mean $\mathcal{B}_E$.

Abbreviations

- a.e., almost everywhere (with respect to a measure)
- a.s., almost surely (with respect to a probability measure)
- e.g., Latin ‘exempli gratia’ (translation: for example)
- i.e., Latin ‘id est’ (translations: that means, in other words, in this case)
- i.i.d., independent and identically distributed
- w.l.o.g., without loss of generality
- w.r.t., with respect to
- SDE, stochastic differential equation

Symbols

- $:=$, definition of the object on the left-hand side as the object on the right-hand side
- $=:$, definition of the object on the right-hand side as the object on the left-hand side
- $\subset$, strict subset
- $\subseteq$, subset (equality is possible)
- $\supset$, strict superset (i.e. contains as a strict subset)
- $\supseteq$, superset (i.e. contains as a subset, equality is possible)
- $\cup$, $\bigcup$, union of sets
- $\sqcup$, $\bigsqcup$, disjoint union of sets
- $\cap$, $\bigcap$, intersection of sets
- $\times$, $\bigtimes$, Cartesian product of sets
- $\wedge$, (pointwise) minimum of extended real numbers or stopping times
- $\vee$, (pointwise) maximum of extended real numbers or stopping times
- $(\,\cdot\,)^+$, positive part, $x^+ := \max\{x, 0\}$ for $x \in \mathbb{R}$
- $(\,\cdot\,)^-$, negative part, $x^- := \max\{-x, 0\}$ for $x \in \mathbb{R}$
- $\lceil\,\cdot\,\rceil$, ceiling function, $\lceil t\rceil := \min\{n \in \mathbb{Z} \mid n \geq t\}$ for $t \in \mathbb{R}$
- $\lfloor\,\cdot\,\rfloor$, floor function, $\lfloor t\rfloor := \max\{n \in \mathbb{Z} \mid n \leq t\}$ for $t \in \mathbb{R}$
- $|\cdot|$, absolute value on $\mathbb{R}$ and $\mathbb{C}$, Euclidean norm on $\mathbb{R}^d$ and $\mathbb{C}^d$, Frobenius matrix norm on $\mathbb{R}^{n\times d}$ and $\mathbb{C}^{n\times d}$
- $\|\cdot\|_\infty$, the (extended) sup norm on the space Y^X of functions from some set X to a normed space Y
- $\square$, end of a proof

Notation

Greek letters are ordered alphabetically according to their spelling in the Latin alphabet.

- $\overset{\text{a.s.}}{=}$, almost sure equality
- $\xrightarrow{\text{a.s.}}$, almost sure convergence
- $B_r(x)$, open ball of radius $r > 0$ with center $x \in S$ in a (pseudo)metric space (S, d)
- $\mathcal{B}_S$, Borel σ-algebra of the topological space S
- $c_{\mathbb{K}} := \dim_{\mathbb{R}}(\mathbb{K})$, equals 1 or 2 depending on whether $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$
- $c_n \to c$, convergence of a sequence $(c_n)_{n\in\mathbb{N}}$ in a topological space to a limit c as $n \to \infty$.
- $C(T, S)$, space of continuous functions from a topological space T to a topological space S
- $C_{\mathrm{b}}(T)$, space of bounded continuous functions from a topological space T to $\mathbb{R}$.

- $\xrightarrow{\mathrm{d}}$, convergence in distribution
- $\stackrel{\mathrm{d}}{=}$, equality in distribution
- δ_x, Dirac measure at x, i.e. probability measure concentrated in x
- det, determinant (of a square matrix)
- $\operatorname{diag}(v)$, diagonal matrix with the entries of the vector $v \in \mathbb{K}^d$ on the diagonal.
- e, Euler's number 2.71828..., also used for the exponential function
- $\mathbb{E}[X\,\mathbb{1}_A] = \int_A X\,\mathrm{d}\mathbb{P} = \int_A X(\omega)\,\mathrm{d}\mathbb{P}(\omega)$, expectation, Lebesgue integral
- $\mathbb{E}[X|\mathcal{G}]$, conditional expectation of X given the sub-σ-algebra $\mathcal{G}$
- $\int f\,\mathrm{d}\mu$, integral of the function f w.r.t. the measure μ
- $\operatorname{ess\,inf}$, essential infimum
- $\operatorname{ess\,sup}$, essential supremum
- exp, exponential function
- f', derivative of the function f
- $\mathcal{F}$, σ-algebra of the sample space Ω
- $\mathbb{F}$, filtration of a σ-algebra $\mathcal{F}$
- $\mathbb{F}^+$, right-continuous enlargement of the filtration $\mathbb{F}$
- $T(\mu)$, the image or pushforward measure $\mu \circ T^{-1}$ of the measure μ under a measurable map T
- i, imaginary unit in the field $\mathbb{C}$ of complex numbers
- $\mathbb{1}_A$, indicator function of the set A
- $\mathbb{1}_s$, the Iverson bracket of the statement s
- I_d, the $(d \times d)$-identity matrix
- $\mathbb{K}$, either the field $\mathbb{R}$ of real or the field $\mathbb{C}$ of complex numbers
- λ, Lebesgue measure on $\mathbb{K}^d$
- $L^p(\Omega, \mathcal{F}, \mu)$ for $p \in [1, \infty)$, Banach space of equivalences classes of real-valued p-integrable functions on the measure space $(\Omega, \mathcal{F}, \mu)$
- $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ for $p \in [1, \infty)$, seminormed vector space of real-valued p-integrable functions on the measure space $(\Omega, \mathcal{F}, \mu)$
- $L_n(X)$, the vector space of $\mathbb{K}^{n\times d}$-valued progressive processes stochastically integrable against a $\mathbb{K}^d$-valued continuous semimartingale X, see [53, Definition 5.110].
- $L(X) := L_1(X)$, the vector space of $\mathbb{K}^d$-valued progressive processes stochastically integrable against a $\mathbb{K}^d$-valued continuous semimartingale where X.
- $\mathcal{L}_{\mathbb{P}}[X]$ or $\mathcal{L}[X]$ when this is unambiguous, the law (or probability distribution) of the random variable $X\colon \Omega \to S$, i.e. the pushforward measure $\mathbb{P} \circ X^{-1}$ of $\mathbb{P}$ on the measurable space $(S, \mathcal{S})$
- $\underline{\lim}$ and $\liminf$ denote the (possibly set theoretic) *limes inferior*.

- $\overline{\lim}$ and $\limsup$ denote the (possibly set theoretic) *limes superior*.
- $\mu(f)$ and $\langle f, \mu\rangle$, synonymous with $\int f \,\mathrm{d}\mu$
- $\mathbb{N} = \{1, 2, 3, \dots\}$, natural numbers (without zero)
- $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, the natural numbers including zero
- $\mathcal{N}(\mu, \sigma^2)$, normal distribution with expectation $\mu \in \mathbb{R}$ and variance $\sigma^2 \geq 0$
- $\mathcal{N}(\mu, C)$, multivariate normal distribution with expectation vector $\mu \in \mathbb{R}^d$ and covariance matrix $C \in \mathbb{R}^{d\times d}$
- $\mathcal{O}_X$, topology on a set X.
- $\mathcal{O}_d$, topology on a set X induced by a (pseudo)metric d on X
- Ω, sample space
- p_t, seminorm of uniform convergence of functions $\mathbb{R}_+ \to \mathbb{K}^n$ on $[0, t]$, see also (2.1).
- $\mathbb{P}$, probability measure on the measurable space $(\Omega, \mathcal{F})$
- $\mathbb{P}_X$, synonymous with $\mathcal{L}[X]$, the law $\mathbb{P} \circ X^{-1}$ of a random variable X
- $\mathbb{P}^*$, outer measure on $(\Omega, \mathcal{F})$ arising from $\mathbb{P}$
- $\mathfrak{P}(\Omega)$, power set, i.e., the set of all subsets of Ω
- $\mathcal{P}(S, \mathcal{S})$ and $\mathcal{P}(S)$, the set of probability measures on a measurable space $(S, \mathcal{S})$
- $\mathcal{P}_p(S)$, the Wasserstein space of order p over a Polish space S
- $\mathcal{P}(\mu, \nu)$, the set of probability measures $\tau \in \mathcal{P}(S \times S, \mathcal{S} \otimes \mathcal{S})$ with marginals $\mu, \nu \in \mathcal{P}(S, \mathcal{S})$
- $\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n$, the product σ-algebra of $\mathcal{F}_1, \dots, \mathcal{F}_n$
- $\mu_1 \otimes \cdots \otimes \mu_n$, the product measure of the σ-finite measures $\mu_1, \dots, \mu_n$
- $\mu^{\otimes n}$, the n-fold product measure of a σ-finite measure μ with itself
- $Q = f \odot P$ or $\mathrm{d}Q = f\,\mathrm{d}P$ or $\frac{\mathrm{d}Q}{\mathrm{d}P} = f$, the measure Q has density (Radon–Nikodym derivative) f w.r.t. the measure P
- $\mathbb{Q}$, the field of rational numbers or a probability measure, depending on the context
- $\mathbb{R}$, the field of real numbers
- $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\} = [-\infty, \infty]$, set of affinely extended real numbers
- $\mathbb{R}_+ = [0, \infty)$, set of positive real numbers
- $\mathbb{R}_+^* = (0, \infty)$, set of strictly positive real numbers
- $\overline{\mathbb{R}}_+ = [0, \infty]$, set of extended positive numbers
- $f\restriction_Y$, the restriction of a map $f\colon X \to Z$ to a subset Y of its domain X
- $\sigma(\mathcal{E})$, the σ-algebra generated by a family $\mathcal{E}$ of subsets of a set Ω.
- $\sigma(f_i : i \in I) = \sigma(\bigcup_{i\in I} f_i^{-1}(\mathcal{F}_i))$, the initial sigma algebra on a set Ω generated by a family of maps $\{f_i\}_{i\in I}$ of maps $f_i\colon \Omega \to \Omega_i$, where $(\Omega_i, \mathcal{F}_i)$ is a measurable space for every $i \in I$.

- $\xrightarrow{\text{ucc}}$, uniform convergence on compact time intervals
- $\xrightarrow{\text{ucp}}$, uniform convergence on compact time intervals in probability
- $\stackrel{\text{uti}}{=}$, equality up to indistinguishability
- $\mathcal{U}_x$, the neighbourhood system of a point x in a topological space
- W_n, the Wiener space $C(\mathbb{R}_+, \mathbb{K}^n)$, where $n \in \mathbb{N}$ and $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, depending on the context
- W_n^t, the space $C([0,t], \mathbb{K}^n)$
- $\mathbb{W}$, the Wiener measure, i.e. the law of $\mathbb{K}^n$-valued Brownian motion, on W_n
- $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, the integers

List of Figures

Bibliography

[1] P. Austing. *Smile Pricing Explained*. Palgrave Macmillan UK, London, 2014. 51, 54

[2] A. Bernou and Y. Liu. Path-Dependent McKean–Vlasov Equation: Strong Well-Posedness, Propagation of Chaos and Convergence of an Interpolated Euler Scheme, Mar. 2023. 16, 185

[3] A. Bernou and Y. Liu. Particle Method for the Numerical Simulation of the Path-Dependent McKean-Vlasov Equation, June 2024. 16, 46

[4] D. Bertsekas and S. E. Shreve. *Stochastic Optimal Control: the Discrete-Time Case*, volume 5. Athena Scientific, 1996. 37, 174, 179, 180

[5] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Mathematical Statistics. Wiley, 1. edition, 1968. 177

[6] N. Bourbaki. *General Topology: Chapters 5–10*. Number v. 4 in Actualités scientifiques et industrielles. Springer, Berlin, Heidelberg, 1998. 97

[7] D. T. Breeden and R. H. Litzenberger. Prices of State-Contingent Claims Implicit in Option Prices. *Journal of Business*, pages 621–651, 1978. 58

[8] V. W. Bryant. A Remark on a Fixed-Point Theorem for Iterated Mappings. *The American Mathematical Monthly*, 75(4):399–400, 1968. 96, 186

[9] G. Cain. *Introduction to General Topology*. Addison-Wesley, 1994. 93, 94, 95, 96

[10] R. Carmona. *Lectures on BSDEs, Stochastic Control, and Stochastic Differential Games with Financial Applications*. SIAM, 2016. 15, 18, 20, 22, 27, 30, 32, 34, 36, 40, 43, 44, 183, 185

[11] L.-P. Chaintron and A. Diez. Propagation of Chaos: A Review of Models, Methods and Applications. I. Models and Methods. *Kinetic and Related Models*, 15(6):895, 2022. 15, 36, 49

[12] R. Cont and J. da Fonseca. Dynamics of Implied Volatility Surfaces. *Quantitative Finance*, 2(1):45, Feb. 2002. 53, 195

[13] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham, and M. Rosestolato. Optimal Control of Path-Dependent McKean–Vlasov SDEs in Infinite Dimension. *The Annals of Applied Probability*, 33(4):2863–2918, 2023. 15, 18, 185

[14] J. C. Cox, J. E. Ingersoll, S. A. Ross, et al. A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2):385–407, 1985. 56

[15] C. Cuchiero and J. Möller. Polynomial McKean-Vlasov SDEs. *arXiv:2502.19203*, 2025. 35

[16] C. Dellacherie and P.-A. Meyer. *Probabilities and Potential*, volume A. North-Holland Publishing Co., Amsterdam, 1978. 101, 104, 105, 106

[17] S. Desmettre, S. Hochgerner, S. Omerovic, and S. Thonhauser. A Mean-Field Extension of the LIBOR Market Model. *International Journal of Theoretical and Applied Finance*, 25(01), 2022. 1, 22, 66, 67, 68, 69, 70, 79, 80, 81, 88, 89, 90, 183, 185, 186

[18] M. F. Djete. Non–Regular McKean–Vlasov Equations and Calibration Problem in Local Stochastic Volatility Models. *arXiv:2208.09986*, 2024. 61

[19] M. F. Djete, D. Possamaï, and X. Tan. McKean–Vlasov Optimal Control: The Dynamic Programming Principle. *The Annals of Probability*, 50(2):791–833, Mar. 2022. 15, 18

[20] B. Dupire. Pricing with a Smile. *Derivatives Pricing: The Classic Collection*, 1994. 53

[21] G. B. Folland. *Real Analysis: Modern Techniques and their Applications*. John Wiley & Sons, 1999. 59, 139, 145, 147

[22] N. Fournier. Convergence of the Empirical Measure in Expected Wasserstein Distance: Non-Asymptotic Explicit Bounds in $\mathbb{R}^d$. *ESAIM: Probability and Statistics*, 27:749–775, 2023. 47, 175

[23] N. Fournier and A. Guillin. On the Rate of Convergence in Wasserstein Distance of the Empirical Measure. *Probability Theory and Related Fields*, 162(3):707–738, 2015. 47, 174

[24] D. H. Fremlin. *Measure Theory*, volume 4. Torres Fremlin, 2000. 179

[25] J. Gatheral. *The Volatility Surface: A Practitioner's Guide*. John Wiley & Sons, 2011. 51, 52, 195

[26] J. Guyon and P. Henry-Labordère. *Nonlinear Option Pricing*. CRC Press, 2013. 27, 36, 38, 40, 49, 51, 55, 57, 58, 60, 61, 63, 64, 65, 66, 183, 185

[27] P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward. Managing Smile Risk. *The Best of Wilmott*, 1:249–296, 2002. 54

[28] S. I. Heston. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6(2):327–343, 1993. 56

[29] C. Homescu. Local Stochastic Volatility Models: Calibration and Pricing, July 2014. 54

[30] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2. edition, 2012. 139

[31] J. Hull and A. White. Numerical Procedures for Implementing Term Structure Models I: Single-Factor Models. *Journal of Derivatives*, 2(1):7–16, 1994. 56

[32] M. Jeanblanc, M. Yor, and M. Chesney. *Mathematical Methods for Financial Markets*. Springer, London, 2. edition, 2009. 22, 56, 136

[33] B. Jourdain, S. Méléard, and W. Woyczynski. Nonlinear SDEs Driven by Lévy Processes and Related PDEs. *arXiv:0707.2723*, 2007. 15, 24, 27, 30, 36, 43, 44, 46, 47, 171, 173, 174, 185

[34] O. Kallenberg. *Foundations of Modern Probability*. Springer, 3. edition, 2021. 106

[35] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113. Springer, 2014. 58, 119

[36] A. Klenke. *Probability Theory: A Comprehensive Course*. Springer Nature, 3. edition, 2020. 6, 7, 8, 37, 50, 59, 63, 109, 114, 166, 167, 174, 178

[37] D. Lacker. On a Strong Form of Propagation of Chaos for McKean–Vlasov Equations. *Electronic Communications in Probability*, 23:1–11, Jan. 2018. 15

[38] J.-F. Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus*. Springer, 2016. 58, 118, 119, 143, 144, 145, 146, 147, 151, 158, 161, 162, 163, 167, 168, 184

[39] L. Mattner. *Complex Differentiation under the Integral*. Institut für Mathematische Stochastik, Universität Hamburg, 1999. 107

[40] H. P. McKean. Propagation of Chaos for a Class of Non-Linear Parabolic Equations. In *Lectures in Differential Equations*, volume 2 of *Van Nostrand Mathematical Studies*, pages 177–194. Van Nostrand, New York, 1969. 1, 14, 15, 31, 185

[41] A. Medvedev and O. Scaillet. Pricing American Options under Stochastic Volatility and Stochastic Interest Rates. *Journal of Financial Economics*, 98:145–159, 2010. 141

[42] J. Munkres. *Topology*. Prentice Hall Inc., 2000. 6

[43] M. Musiela and M. Rutkowski. Continuous-Time Term Structure Models: Forward Measure Approach. *Finance and Stochastics*, 1(4), 1997. 67, 68

[44] B. Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer Science & Business Media, 6. edition, 2010. 22

[45] V. M. Panaretos and Y. Zemel. *An Invitation to Statistics in Wasserstein Space*. Springer Nature, 2020. 176, 179, 184

[46] V. M. Panaretos and Y. Zemel. *An Invitation to Statistics in Wasserstein Space, Supplement*. Springer Nature, 2020. 179

[47] K. R. Parthasarathy. *Probability Measures on Metric Spaces*. Academic Press Inc., 1967. 113

[48] H. Pham. *Continuous-Time Stochastic Control and Optimization with Financial Applications*, volume 61. Springer Science & Business Media, 2009. 121

[49] L. Rasmussen. *Computational Finance: On the Search for Performance*. PhD thesis, Department of Mathematical Sciences, Faculty of Science, University of Copenhagen, 2016. 60

[50] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales, Foundations*, volume 1. Cambridge University Press, 1994. 119

[51] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales, Itô Calculus*, volume 2. Cambridge University Press, 2000. 143, 144, 149, 150, 151, 156, 158, 184

[52] M. Roper. Arbitrage Free Implied Volatility Surfaces. *preprint*, 2010. 60

[53] U. Schmock. *Stochastic Analysis for Financial and Actuarial Mathematics, Lecture Notes*. Unpublished, May 2024. 6, 9, 10, 18, 20, 24, 25, 26, 27, 29, 37, 38, 39, 40, 46, 48, 50, 56, 68, 69, 77, 78, 83, 86, 88, 90, 93, 101, 102, 105, 107, 108, 110, 111, 112, 117, 118, 119, 120, 122, 123, 124, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 144, 145, 146, 147, 148, 149, 152, 154, 155, 158, 159, 162, 163, 164, 166, 167, 174, 177, 179, 182, 183, 184, 185, 186, 187, 191

[54] M. Schranzhofer. Selected Generalizations in Stochastic Analysis. Master's Thesis, Technische Universität Wien, 2024. 138, 139, 187

[55] S. E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*, volume 11. Springer, 2. edition, 2004. 54, 56, 67, 68

[56] A.-S. Sznitman. Topics in Propagation of Chaos. *Ecole d'Été de Probabilités de Saint-Flour XIX—1989*, 1464:165–251, 1991. 30, 185

[57] E. Tarafdar. An Approach to Fixed-Point Theorems on Uniform Spaces. *Transactions of the American Mathematical Society*, 191:209–225, 1974. 98, 186

[58] A. W. Van der Vaart. *Asymptotic Statistics*, volume 3. Cambridge University Press, 2000. 39

[59] V. S. Varadarajan. On the Convergence of Sample Probability Distributions. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 19(1/2):23–26, 1958. 173

[60] C. Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics. American Mathematical Society, 2003. 172, 184

[61] C. Villani. *Optimal Transport: Old and New*, volume 338. Springer, 2009. 171, 172, 173, 176, 184

[62] H. von Weizsäcker. *Stochastic Integrals: An Introduction*. Springer, 2013. 158

[63] D. Williams. *Probability with Martingales*. Cambridge University Press, 1991. 174

[64] T. Yamada and S. Watanabe. On the Uniqueness of Solutions of Stochastic Differential Equations. *Journal of Mathematics of Kyoto University*, 11(1):155–167, 1971. 158

Index

A C D E F H I K L M P R S T U V W