Well-posedness of stochastic Volterra equations with non-Lipschitz coefficients

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften der Universität Mannheim

vorgelegt von
David Scheffels
aus Frankfurt am Main

Mannheim, 2023

Dekan:	Professor Dr. Claus Hertling, Universität Mannheim
Referent:	Professor Dr. David J. Prömel, Universität Mannheim
Korreferent:	Professor Dr. Christa Cuchiero, University of Vienna
Korreferent:	Professor Dr. Christoph Czichowsky, London School of Economics and Political Science

Tag der mündlichen Prüfung: 17. November 2023

Abstract

The main focus of this doctoral thesis is to consider stochastic Volterra equations (SVEs), where the diffusion coefficients are only Hölder continuous of order $1/2$, and to prove well-posedness results for these equations, i.e. that they possess a pathwise unique strong solution.

We start with the well-posedness for SVEs with sufficiently regular kernels by proving strong existence and pathwise uniqueness directly by adapting techniques from the well-known theory for stochastic differential equations (SDEs).

Afterwards, we consider more general kernels including singular kernels and introduce a general Volterra local martingale problem. Using this, we are able to prove the weak existence of solutions to SVEs with continuous coefficients and with regular or convolutional, possibly singular, diffusion kernels.

Next, and constituting the major part of this thesis, we consider explicitly SVEs with the fractional kernel $K(s,t) = (t-s)^{-\alpha}$, where $\alpha \in [0, 1/2)$, in the drift and the diffusion, and prove pathwise uniqueness for these equations under a mild condition on the relationship of the intensity of the singularity of the kernels, i.e. on α, and the Hölder regularity of the diffusion coefficient. Together with the weak existence, this implies the well-posedness of the equation by the famous Yamada–Watanabe theorem.

To round off the work, we look at two more interesting topics. First, we introduce the class of Mean-field stochastic Volterra equations which merge two generalizations of classical SDEs, both of which have received a lot of attention recently, namely SVEs and mean-field SDEs, also referred to as McKean–Vlasov SDEs. For these equations, we prove the well-posedness and a quantitative, pointwise propagation of chaos result of Volterra-type systems of interacting particles. We do that in two settings, firstly, for finite-dimensional equations with general kernels and Lipschitz continuous coefficients, and secondly, for one-dimensional equations with regular or convolutional kernels and up to $1/2$-Hölder continuous diffusion coefficients.

Last, we introduce neural stochastic Volterra equations (neural SVEs) which is a model that is able to learn the dynamics of an SVE by a deep learning structure, inspired by the recently emerged model of neural SDEs.

Zusammenfassung

In dieser Thesis beschäftigen wir uns mit stochastischen Volterra Gleichungen (SVEs), bei denen die Diffusionskoeffizienten nur Hölder stetig mit Ordnung 1/2 sind, und beweisen sogenannte Well-posedness Resultate für diese Gleichungen, d.h. untersuchen, unter welchen Umständen sie eine pfadweise eindeutige starke Lösung besitzen.

Wir beginnen mit der Well-posedness für SVEs, bei denen die Kerne regulär sind, und zeigen für diese die starke Existenz und pfadweise Eindeutigkeit von Lösungen durch Anpassen wohlbekannter Techniken für stochastische Differentialgleichungen (SDEs).

Anschließend betrachten wir allgemeinere Kerne, die insbesondere auch Singularitäten beinhalten dürfen, und führen ein lokales Volterra Martingalproblem ein. Mithilfe dieses zeigen wir die schwache Existenz von Lösungen zu SVEs mit stetigen Koeffizienten und regulären Kernen oder Faltungskernen, die singulär sein dürfen, in der Diffusion.

Danach kommen wir zum Hauptkapitel dieser Thesis, in welchem wir explizit SVEs mit dem fraktionalen Kern $K(s,t) = (t-s)^{-\alpha}$, wobei $\alpha \in [0, 1/2)$, in Drift und Diffusion betrachten, und zeigen pfadweise Eindeutigkeit für diese Gleichungen unter einer milden Annahme an die Beziehung zwischen der Intensität der Singularität der Kerne, also α, und der Hölder Regularität des Diffusionskoeffizienten. Zusammen mit der schwachen Existenz impliziert dies die Well-posedness der Gleichung nach dem berühmten Resultat von Yamada und Watanabe.

Um die Thesis abzurunden, betrachten wir zwei weitere interessante Themengebiete. Als erstes führen wir die Klasse von Mean-field stochastischen Volterra Gleichungen ein, welche eine Verallgemeinerung von SVEs und den sogenannten Mean-field SDEs darstellen. Beide Arten von Gleichungen haben in den letzten Jahren eine große Popularität erfahren. Wir zeigen die Well-posedness sowie ein quantitatives, punktweises Propagation of Chaos Resultat für Systeme vom Volterra-Typ, die die Interaktion von Partikeln modellieren. Dies tun wir in zwei verschiedenen Situationen, zunächst für endlich-dimensionale Gleichungen mit allgemeinen Kernen und Lipschitz stetigen Koeffizienten, und anschließend für ein-dimensionale Gleichungen mit regulären Kernen oder Faltungskernen und bis zu 1/2-Hölder stetigen Diffusionskoeffizienten.

Abschließend führen wir neuronale stochastische Volterra Gleichungen ein, welche die Dynamiken von SVEs durch neuronale Strukturen lernen können, und inspiriert sind durch die kürzlich eingeführten neuronalen SDE Modelle.

Acknowledgements

During large parts of my PhD, I gratefully received financial support by the Research Training Group "Statistical Modeling of Complex Systems" (RTG 1953) funded by the German Science Foundation (DFG).

First and foremost, I want to thank everyone who made my three years as a doctoral student so enjoyable and pleasant. Despite the challenges that the global pandemic brought, I had a great time, which I have enjoyed very much, and in which I have learned and experienced a lot on various levels.

Most importantly, I would like to express my deepest gratitude to my supervisor David Prömel for the incredibly valuable support that I have received from him. I have learned so much from him, especially, but not exclusively, about mathematics and about how to work successfully in academia. However, beyond that, also about other important topics in life such as football, insurance or life planning. I have benefited a lot from his great commitment and willingness to always offer assistance and ideas. When he initially asked me if I could imagine pursuing a PhD back in my times as a master student, I was not sure if this is the right thing for me, but looking back, it was the best decision to go for it.

Next, I would like to send a big thank you to Christa Cuchiero and Christoph Czichowsky for accepting to review my thesis, dedicating their time on it and providing their expert opinions. Christa Cuchiero's work on stochastic Volterra equations has moreover been a great inspiration for my own research.

A major credit for the good time I had is due to my colleagues and friends at the university. Thanks for making my lunch and coffee breaks enjoyable, especially to Alex, André, Anna, Annalena, Khadija, Paul and Thomas. I am of course also grateful for all my other good friends who have accompanied me on my journey.

That I was able to experience and master all of this would not have been possible without the indestructible support from my family, Susanne, Rolf and Jakob, but also my larger family. They always helped me in any way they could, and even though their support in mathematical matters could not be as large as they wanted to, their emotional and moral support was even more. Last but certainly not least, I want to express my thankfulness for having my better half Nina. Knowing her by my side and that she always believes in me is the most important thing for me.

Contents

Chapter 1

Introduction

In this thesis, we consider stochastic Volterra equations (SVEs) and deal with the fundamental question of existence and uniqueness of solutions to these equations. Originally, Vito Volterra dealt with (deterministic) differential equations to describe population growth models with memory in hunter-prey-models in his work [Vol90]. SVEs have been studied in probability theory starting with the works of Berger and Mizel [BM80a, BM80b]. This class of integral equations constitutes a generalization of ordinary stochastic differential equations and serves as a well suited mathematical model for numerous random phenomena appearing, e.g. in biology (see e.g. [MS15, AJ21]) and mathematical finance (see e.g. [AJEE19b, EER19]).

We fix some end time point $T \in (0, \infty)$ and let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space which satisfies the usual conditions. In our setting, an SVE is an integral equation of the form

$$X_t = x_0(t) + \int_0^t K_\mu(s,t)\mu(s,X_s)\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s,X_s)\,\mathrm{d}B_s, \quad t \in [0,T], \tag{1.1}$$

where x_0 denotes the (deterministic) initial condition, $(B_t)_{t\in[0,T]}$ is a standard Brownian motion, the so-called kernels K_μ, K_σ, the drift coefficient μ and the diffusion coefficient σ are all deterministic, Borel measurable functions. Sometimes we will allow $x_0 \equiv X_0$ to be an $\mathcal{F}_0$-measurable random initial condition which does not make anything more difficult. We will, over large parts of the work, consider one-dimensional SVEs such that all the objects in equation (1.1) simply map to $\mathbb{R}$, but we will give more formally precise definitions of these objects in the respective settings in the course of this work. For the kernels however, we already introduce the notation $\Delta_T := \{(s,t) \in [0,T]\times[0,T] \colon 0 \le s \le t \le T\}$ and define $K_\mu, K_\sigma \colon \Delta_T \to \mathbb{R}$ to make clear that we always consider $t \ge s$ and that the kernels are always real-valued. In the center of our interest will be the question of well-posedness of the SVE, i.e. the existence of a unique strong solution to equation (1.1), in various settings with differing regularity assumptions on the kernels and the diffusion coefficient σ.

1.1 Motivation

The motivation to study stochastic Volterra equations with non-Lipschitz coefficients has two perspectives. On the one hand, it is a natural question to explore to what extent the famous results of Yamada and Watanabe [YW71], ensuring pathwise uniqueness and the existence of strong solutions for ordinary stochastic differential equations, generalizes to stochastic Volterra equations. On the other hand, our motivation comes from real-world applications. Stochastic Volterra equations with only 1/2-Hölder continuous coefficients recently got a great deal of attention in mathematical finance as so-called rough volatility models, see e.g. [AJEE19b, EER19], which have demonstrated to fit remarkably well historical and implied volatilities of financial markets, see e.g. [BFG16]. Furthermore, SVEs with non-Lipschitz continuous coefficients arise as scaling limits of branching processes in population genetics, see [MS15, AJ21]. Although the rough Heston SVE (1.4) and the branching equation (1.5) are not covered by the pathwise uniqueness result of this thesis due to the intensity of the singularities in the kernels and the weak regularity of the diffusion coefficients which are only 1/2-Hölder continuous (cf. Main Theorem 3, but the weak existence in Main Theorem 2 does hold), these examples give a good indication of why SVEs with singular kernels and non-Lipschitz diffusion coefficients are interesting to look at.

Rough volatility models

In mathematical finance, an important topic is pricing of financial instruments. For this purpose, the most popular market model is the Black–Scholes model, originally introduced in [BS12] and awarded with the 1997 Nobel Memorial Prize in Economic Sciences. In the Black–Scholes model, the dynamics of some risky asset $(S_t)_{t\in[0,T]}$ are modeled by the SDE

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}B_t, \qquad t \in [0,T],$$

where B is a standard Brownian motion, $\mu \in \mathbb{R}$ describes the expected return and $\sigma > 0$ the volatility of the return of the asset or equivalently, log-volatility of the asset. Here, explicit formulas for the pricing of European put- and call- options can be derived. However, the assumption of a deterministic, constant volatility $\sigma > 0$ is quite heavy and unrealistic as, among other inconsistencies, for example the "Volatility smile" effect (see e.g. [DK94]) shows. As a consequence, so-called stochastic volatility models came up. Here, the volatility of the risky asset's return or, equivalently, the log-variance of the risky asset, is modeled by some stochastic dynamics as well. The most famous stochastic volatility model is the Heston model (see [Hes93]), where the risky asset is modeled by the SDE

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sqrt{\nu_t} S_t \,\mathrm{d}B_t^S, \qquad t \in [0,T],$$

and the log-variance process is itself modeled by the Cox–Ingersoll–Ross SDE

$$\mathrm{d}\nu_t = \kappa(\theta - \nu_t)\,\mathrm{d}t + \xi\sqrt{\nu_t}\,\mathrm{d}B_t^\nu, \qquad t \in [0,T], \tag{1.2}$$

where B^S and B^ν are correlated Brownian motions, $\theta > 0$ describes the expected variance, $\kappa > 0$ the mean-reversion rate and $\xi > 0$ the variance of ν.

It can be seen that while stochastic volatility models as the Heston model might be able to accurately model volatilities for long maturities, their behavior is still not satisfactory for shorter time horizons. To overcome that issue, it was first proposed by Comte and Renault in [CR98] and then by Gatheral et al. in [GJR18] to use the properties induced by the fractional Brownian motion (fBM) defined for $t \in [0,T]$ by

$$B_t^H = \frac{1}{\Gamma(H+1/2)}\left(\int_{-\infty}^0 \left((t-s)^{H-1/2} - (-s)^{H-1/2}\right) \mathrm{d}B_s + \int_0^t (t-s)^{H-1/2}\, \mathrm{d}B_s\right), \quad (1.3)$$

where the parameter $H \in (0,1)$ is called Hurst parameter and controls the roughness of the paths. The fBM is a generalization of the standard Brownian motion since they coincide for $H = 1/2$. For any $H \in (0,1)$, the paths of the fBM are Hölder continuous for any order strictly smaller than H, meaning that the paths are rougher the smaller H is. Moreover, for $H \neq 1/2$, the increments are not independent anymore: for $H > 1/2$, they are positively correlated, and for $H < 1/2$ negatively.

Gatheral et al. show in [GJR18] that empirically observed volatilities fit remarkably well the properties of the fBM with Hurst parameter $H = 0.1$, since they have rougher paths than the standard Brownian motion. As a consequence, the class of so-called rough volatility models came up, with the most prominent representative being the rough Heston model, see [EER19], [AJEE19b]. To imitate the properties of the fBM with $H < 1/2$, i.e. making its paths rougher, one uses the kernel function $K_\mu(s,t) = K_\sigma(s,t) = (t-s)^{-\alpha}$, for $\alpha \in (0,1/2)$, to modify the log-volatility in (1.2) to

$$V_t = V_0 + \frac{1}{\Gamma(1-\alpha)}\int_0^t (t-s)^{-\alpha}\kappa(\theta - V_s)\,\mathrm{d}s + \frac{1}{\Gamma(1-\alpha)}\int_0^t (t-s)^{-\alpha}\xi\sqrt{V_s}\,\mathrm{d}B_s, \quad t \in [0,T], \quad (1.4)$$

which is a stochastic Volterra equation in the sense of (1.1) with the singular kernels $K_\mu(s,t) = K_\sigma(s,t) = (t-s)^{-\alpha}$ and with diffusion coefficient $\sigma(t,x) = \sqrt{x}$ that is only 1/2-Hölder continuous in x.

Volterra processes in general, and the rough Heston process in specific, are neither semimartingales nor Markov processes, making on the one hand the task of pricing derivatives in the rough Heston model challenging. However, it has been shown in [EER19] that Fourier-based techniques can be used to price European options. On the other hand, to gain theoretical results as e.g. well-posedness results is also challenging. Well-posedness for the rough Heston model (1.4) is at time of writing still an open question with Main Theorem 3 being, to the best of our knowledge, the closest yet achieved result.

Branching processes

Another motivating real-world application that reveals an SVE is the scaling limit of branching processes described in [MS15]. Consider the chemical interaction between two substances where the first one called reactant diffuses randomly in one-dimensional space, and reacts with the second one called catalyst proportionally to the concentration of the catalyst at the point of contact. To describe the total behavior, one can model the reactant as a system of n particles moving randomly according to a standard Brownian motion, and the catalyst being distributed at time t according to some deterministic measure $\rho_t(\mathrm{d}x)$. If a particle of the reactant enters the catalyst region, it either dies or branching occurs, meaning that the particle splits into two particles behaving independently but with the same spatial movement as their parent.

We can now look at the reactant as a measure valued process $(\bar{X}_t^n(\mathrm{d}x))_{t\in[0,T]}$ defined by

$$\bar{X}_t^n(A) = \frac{\text{number of particles in } A \text{ at time } t}{n}, \qquad \text{for Borel set } A.$$

If we send $n \to \infty$, it can be shown that $(\bar{X}_t^n)_{t\in[0,T]}$ converges towards some measure valued catalytic super-Brownian motion $(\bar{X}_t)_{t\in[0,T]}$ that has a density $\bar{X}_t(\mathrm{d}x) = \int_{\mathbb{R}} X_t(x)\,\mathrm{d}x$ which fulfills the stochastic partial differential equation

$$\frac{\partial X_t(x)}{\partial t} = \frac{1}{2}\Delta X_t(x) + \sqrt{X_t(x)}\dot{W}^\rho(t,x),$$

where $\dot{W}^\rho(t,x)$ is a space-time white noise with covariance structure determined by ρ (see [ZÖ5]).

Suppose one now considers branching occurring only at $x = 0$ such that $\rho_t(\mathrm{d}x) = \delta_0(\mathrm{d}x)$ is the Dirac-delta measure and assumes, as it is described in [MS15], that some absolute continuity condition on the local time of X spent in $x = 0$ holds (which in fact does not hold, but for illustrative purposes it makes sense to assume this). Then, $(X_t)_{t\in[0,T]}$ can be written as a solution to the SPDE in mild form

$$X_t(x) = \int_{\mathbb{R}} p_t(x-y)X_0(\mathrm{d}y) + \int_0^t p_{t-s}(x)\sqrt{X_s(0)}\,\mathrm{d}B_s, \qquad t \in [0,T],$$

where $p_t(x)$ is the transition density of the Brownian motion. For $x = 0$, we obtain that

$$X_t(0) = \int_{\mathbb{R}} p_t(y)X_0(\mathrm{d}y) + \int_0^t \frac{1}{\sqrt{2\pi}}(t-s)^{-1/2}\sqrt{X_s(0)}\,\mathrm{d}B_s, \qquad t \in [0,T], \tag{1.5}$$

which seems like an SVE of the form of (1.1) with singular diffusion kernel and diffusion coefficient σ that is only 1/2-Hölder continuous. Note though that the kernel $K_\sigma(s,t) = (t-s)^{-1/2}$ in (1.5) is not locally square-integrable, such that the second integral is not well-defined as an Itô integral. This is due to the false assumption on the local time. However, as soon as we decrease the negative exponent by some $\varepsilon > 0$ the equation becomes a valid SVE.

1.2 Mathematical background

Since SVEs can be seen as a generalization of stochastic differential equations (SDEs), it is natural to look at well-posedness results for SDEs for comparison.

Well-posedness in the SDE case

If we consider $K_\mu = K_\sigma = 1$, equation (1.1) becomes an ordinary stochastic differential equation (SDE) written in integral form. Note that in general, we cannot write SVEs in differential form due to the t-dependence inside the integrals. For SDEs, the question of well-posedness is well investigated. The standard textbook setting in d dimensions is when μ and σ are globally Lipschitz continuous in the space variable uniformly in the time variable and fulfill a linear growth condition, i.e.

$$\begin{aligned} |\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| &\leq C|x-y|, \quad \text{and} \\ |\mu(t,x)| + |\sigma(t,x)| &\leq C(1+|x|) \end{aligned}$$

hold for all $t \in [0,T]$, $x, y \in \mathbb{R}^d$ and some constant $C > 0$. The notation $|\cdot|$ always stands for the Euclidean norm on the space $\mathbb{R}^d$. Then, a strong solution can be constructed by a Picard iteration, and pathwise uniqueness can be shown using Grönwall's lemma (see e.g. [KS91, Theorem 5.2.9], [Kle14, Theorem 26.8], [KP92, Theorem 4.5.3]). Using some localization argument, the Lipschitz assumption on μ and σ can be relaxed to requiring only local Lipschitz continuity (see e.g. [KP92, p.134-135]).

Considering μ, the Lipschitz condition can be slightly generalized to a so-called Osgood condition (see e.g. [KS91, Remark 5.2.16]). However, it is not possible to reach only Hölder continuity for μ, since already for ordinary differential equations uniqueness fails to hold in general for coefficients that are only Hölder continuous, as the counterexample equation $X_t = \int_0^t |X_s|^\alpha \,\mathrm{d}s$ which has a continuum of solutions for $\alpha < 1$ shows (see e.g. [KS91, p.287]).

For σ, the equation $X_t = \int_0^t |X_s|^\alpha \,\mathrm{d}W_s$ for $\alpha \in (0,1/2)$ which fails to fulfill uniqueness even in the weak sense shows that we cannot expect well-posedness for SDEs with σ that is only ξ-Hölder continuous for some $\xi \in (0,1/2)$. For $\xi \in [1/2,1]$ however, the famous approach of Yamada–Watanabe (see [YW71, Theorem 1]) proves pathwise uniqueness for one-dimensional SDEs with only ξ-Hölder continuous σ. Together with the weak existence of a solution, the second important Yamada–Watanabe result (see e.g. [KS91, Corollary 5.3.23], [Kur14, Theorem 1.5]) implies then the existence of a strong solution and hence well-posedness of the SDE. A weak solution consists not only of the process but also of the underlying probability space which is the major difference to a strong solution, where one requires the existence of a solution for any given probability space. Weak existence for SDEs was first proven by Skorokhod in [Sko61] (see also [SV79], [KS91]) and typically uses the formulation of a martingale problem (see e.g. [HS12, Theorem 0.1]) which we will also introduce for SVEs.

The key technique in the Yamada–Watanabe pathwise uniqueness approach is to approximate the absolute value function smoothly and then to apply Itô's formula which relies on the semimartingale property of solutions to the SDE. This property fails in general for solutions to the SVE which rises the question if and under which conditions we can still obtain analogue results.

The aim of this thesis is to determine how far the aforementioned discussed results for SDEs can be extended to the more general case of SVEs.

SVEs with Lipschitz continuous coefficients

The existence of unique strong solutions for stochastic Volterra equations, where both coefficients μ and σ are Lipschitz continuous, is well investigated. Indeed, classical existence and uniqueness results for SVEs with sufficiently regular kernels are due to [BM80a, BM80b, Pro85].

In the pioneering work of Berger and Mizel [BM80a, BM80b], the Volterra integral equation

$$X_t = x_0(t) + \int_0^t f(t,s,X_s)\,\mathrm{d}s + \int_0^t \sigma(t,s,X_s)\,\mathrm{d}B_s, \qquad t \in [0,T],$$

was introduced. Under Lipschitz conditions on f and σ in the space variable, uniformly in the time variables s and t, which corresponds to bounded kernels and Lipschitz continuous coefficients in our setting (1.1), the authors proved strong existence and pathwise uniqueness by adapting the classical tools from the well-known SDE theory. More precisely, a Picard–Lindelöf iteration was used to construct a solution, and Grönwall's inequality to obtain pathwise uniqueness.

By using similar techniques, in [Pro85], the existence and uniqueness of solutions to SVEs of the form

$$X_t = H_t + \int_0^t f(t,s,X_s)\,\mathrm{d}Z_s, \quad t \in [0,T], \tag{1.6}$$

was proven, where $(H_t)_{t\in[0,T]}$ is adapted and right-continuous, $(Z_t)_{t\in[0,T]}$ is a right-continuous semi-martingale and the function f has some suitable Lipschitz continuity in the space variable and is differentiable in t. Moreover, the important semimartingale property for solutions of (1.6) under the assumption that f is partially differentiable with respect to t was derived. This property will be used later in this work for the case of SVEs with regular kernels (see Chapter 2).

These results have been generalized in various directions such as allowing for anticipating and path-dependent coefficients [PP90, ØZ93, AN97, Kal21] or an infinite dimensional setting [Zha10]. Well-posedness was also successfully proven for the case of Lipschitz coefficients and singular kernels that fulfill some Besov regularity condition in [CD01], see also [CLP95], which in particular covers the case $K_\mu = K_\sigma = K_\alpha$, where K_α is the fractional kernel defined by

$$K_\alpha(s,t) = (t-s)^{-\alpha}, \qquad (s,t) \in \Delta_T, \tag{1.7}$$

for $\alpha \in (0,1/2)$. A slight extension beyond Lipschitz continuous coefficients, but not covering the case where σ is only Hölder continuous, can be found in [Wan08].

SVEs with non-Lipschitz continuous coefficients

The main challenge when dealing with SVEs, compared to ordinary stochastic differential equations, is that the stochastic integral and hence the Volterra process is in general not a semimartingale. An example which illustrates this is the fractional Brownian motion (fBM) which is defined for $H \in (0,1)$ in (1.3). The fBM has infinite quadratic variation for $H < 1/2$ and zero quadratic variation for $H > 1/2$ which contradicts, if it were a semimartingale, the fact that it is not of bounded variation (for details see e.g. [Rog97, Section 2]) and thus, the fBM fails to be a semimartingale for $H \neq 1/2$. As described above, the semimartingale property is the basis of the famous Yamada–Watanabe technique in [YW71] for proving pathwise uniqueness for SDEs.

The classical approach to prove the existence of strong solutions to SDEs with less regular diffusion coefficients than Lipschitz continuous ones is to first show the existence of a weak solution, since this, in combination with pathwise uniqueness, guarantees the existence of a strong solution, see [YW71]. The existence of weak solutions for stochastic Volterra equations was derived in the work of Abi Jaber, Cuchiero, Larsson and Pulido [AJCLP21] (see also [MS15, AJLP19, AJ21]), assuming that the kernels in the stochastic Volterra equations are of convolution type, i.e. in our setting $K_\mu(s,t) = K_\sigma(s,t) = K(t-s)$ for some suitable function K. Assuming additionally that the coefficients μ, σ lead to affine Volterra processes, weak uniqueness was obtained in [MS15, AJEE19a, AJ21, CT20]. However, as we do not impose a convolutional structure on the stochastic Volterra equation (3.1), we cannot rely on the known results regarding the existence of weak solutions.

The task of proving pathwise uniqueness for the SVE (1.1) with singular kernels and σ that is only Hölder continuous, is an open question. For the explicit equation with the fractional kernel K_α defined in (1.7), i.e. $K_\mu = K_\sigma = K_\alpha$, and assuming the drift coefficient μ does not depend on the solution $(X_t)_{t\in[0,T]}$, Mytnik and Salisbury [MS15] established pathwise uniqueness for the SVE (1.1) where the diffusion coefficient σ is time-homogeneous but only ξ-Hölder continuous for $\xi \in (\frac{1}{2(1-\alpha)}, 1]$ by equivalently reformulating the SVE into a stochastic partial differential equation, which then allows to accomplish a proof of pathwise uniqueness in the spirit of Yamada–Watanabe relying on the methodology developed in [MPS06, MP11]. We generalize the results and method of Mytnik and Salisbury [MS15] to derive pathwise uniqueness for the explicit SVE (1.1) with $K_\mu = K_\sigma = K_\alpha$ to the case of general time-inhomogeneous coefficients. As classical transforms allowing to remove the drift of an SDE are not applicable to the SVE (1.1), the general time-inhomogeneous coefficients μ create severe novel challenges.

Mean-field SDEs

If the coefficients of an ordinary SDE depend not only on the time variable and the solution but also on the solution's law, the equation is called mean-field SDE. Mean-field SDEs are thus an extension of ordinary SDEs and can be written in integral form as

$$X_t = X_0 + \int_0^t \mu(s, X_s, \mathcal{L}(X_s))\,\mathrm{d}s + \int_0^t \sigma(s, X_s, \mathcal{L}(X_s))\,\mathrm{d}B_s, \qquad t \in [0,T], \tag{1.8}$$

where X_0 is a random initial condition and $\mathcal{L}(X)$ denotes the law of a random variable X. This class of stochastic equations was originally studied by Kac in [Kac56], [Kac59], and then by McKean in [McK66] and Vlasov in [Vla67] for the reason that they arise from a differential equation in physics called Boltzmann's equation which models systems of single atom gas particles and their interaction. They are therefore also often referred to as McKean–Vlasov equations. Nowadays, mean-field equations are an active topic of research with various applications, see e.g. [Szn91, JW17, CD18a, CD18b, CD22a, CD22b] for comprehensive introductions to mean-field SDEs and their applications.

It is well-known that under Lipschitz conditions on the coefficients the multi-dimensional mean-field equation (1.8) is well-posed, see e.g. [CD18a, Theorem 4.21]. The typical method to approximate the solution is to set up an N-particle system, where the particles are driven by independent Brownian motions but the law inserted into the coefficients is the empirical distribution of the N particles in each of the N equations. Then, it is possible to prove a convergence result to the solution of the mean-field equation, see e.g. [CD18b, Theorem 2.12], which in the literature is called propagation of chaos result.

Propagation of chaos results have been generalized in the SDE case from the Lipschitz setting into various directions including non-Lipschitz settings, see [HW23], [BCC11], [KP21], mixed jump-diffusions with simultaneous jumps, see [Gra92], [ADPF18], or a local Lipschitz jump-diffusion setting, see [Ern22].

1.3 Main results of the thesis

The main part of this thesis is about the well-posedness of the SVE (1.1), where the diffusion coefficient σ is only Hölder continuous.

In Chapter 2, our first main contribution is to establish the well-posedness of one-dimensional SVEs of the form (1.1) provided the diffusion coefficient σ is only $1/2+\xi$-Hölder continuous for $\xi \in [0, 1/2]$ and the kernels are regular (see Assumption 2.1). Therefore, we directly prove the existence of a strong solution by showing the convergence of an Euler type approximation of the SVE (1.1) and do not use the concept of weak solutions. For ordinary stochastic differential equations such an approach was developed by Gyöngy and Rásonyi [GR11], using ideas coming from [YW71].

For the pathwise uniqueness, we generalize the classical approach of Yamada and Watanabe [YW71] to the more general setting of stochastic Volterra equations. We show therefore the semimartingale property of the solution, and moreover some additional properties that also hold for the singular kernel case. We end up with the following main result of Chapter 2. For the mathematically precise formulation see Theorem 2.3.

Main Theorem 1. *Suppose that the kernels are regular, μ is Lipschitz continuous and σ is $\frac{1}{2}+\xi$-Hölder continuous for some $\xi \in [0, \frac{1}{2}]$. Then, there exists a unique strong solution $(X_t)_{t\in[0,T]}$ to the one-dimensional stochastic Volterra equation* (1.1). *Moreover, the solution is Hölder continuous, has bounded moments and has a semimartingale property.*

In the following, we also consider singular kernels. In Chapter 3, we first introduce a local martingale problem associated to general stochastic Volterra equations, see Definition 3.4, and show that its solvability is equivalent to the existence of a weak solution to the associated SVE, see Lemma 3.7.
Using this newly formulated Volterra martingale problem, we then prove the weak existence of solutions to the one-dimensional SVE (1.1), where the time-inhomogeneous coefficients μ and σ are only required to be uniformly continuous in x, the drift kernel K_μ only has to fulfill an L^1-condition and the diffusion kernel K_σ is allowed to be either regular or singular of convolutional form, i.e. $K_\sigma(s,t) = \tilde{K}(t-s)$ for some $\tilde{K} \in L^2([0,T])$. This includes in particular the fractional kernel K_α defined in (1.7). In this setup, we derive the following main result of Chapter 3, which is Theorem 3.10.

Main Theorem 2. *Under the above described assumptions on μ, σ, K_μ and K_σ, there exists a weak solution to the stochastic Volterra equation* (1.1).

Main Theorem 2 hence implies in particular the existence of a weak solution to the one-dimensional fractional SVE

$$X_t = x_0(t) + \int_0^t (t-s)^{-\alpha}\mu(s,X_s)\,\mathrm{d}s + \int_0^t (t-s)^{-\alpha}\sigma(s,X_s)\,\mathrm{d}B_s, \quad t \in [0,T], \tag{1.9}$$

for $\alpha \in (0, 1/2)$.

For the fractional SVE (1.9), we then prove pathwise uniqueness in Chapter 4 under the assumption that μ is Lipschitz continuous and σ is ξ-Hölder continuous for $\xi > \frac{1}{2(1-\alpha)}$. This implies then using the general Yamada–Watanabe result (see [Kur14, Theorem 1.5]) the existence of a strong solution to (1.9) and hence, the well-posedness of the fractional SVE.
We prove the pathwise uniqueness of the fractional SVE (1.9) in six steps (see Section 4.2), by generalizing the well-known techniques of Yamada–Watanabe (cf. [YW71, Theorem 1]) and the work of Mytnik and Salisbury [MS15]. One of the main challenges is the missing semimartingale property of a solution $(X_t)_{t\in[0,T]}$ to (1.9). Therefore, we transform (1.9) into a random field (see (4.7)) in Step 1, for which we can derive a semimartingale decomposition (see (4.8)). Then, we implement an approach in the spirit of Yamada–Watanabe in Steps 2 to 5 and conclude the pathwise uniqueness by using a Grönwall inequality for weak singularities (see e.g. [Kru14, Lemma A.2]) in Step 6. We end up with the following main result of Chapter 4, which is Theorem 4.3.

Main Theorem 3. *Suppose μ to be Lipschitz continuous and σ to be ξ-Hölder continuous for $\xi > \frac{1}{2(1-\alpha)}$. Then, pathwise uniqueness holds for the fractional stochastic Volterra equation* (1.9)*. Consequently, equation* (1.9) *is well-posed.*

In addition to the SVE well-posedness results, we also cover two more topics in this thesis which are mean-field SVEs and neural SVEs.

Mean-field SVEs

Inspired by mean-field SDEs of the type (1.8), we introduce in Chapter 5 the class of mean-field stochastic Volterra equations (mean-field SVEs) that have the form

$$X_t = X_0 + \int_0^t K_\mu(s,t)\mu(s, X_s, \mathcal{L}(X_s))\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s, X_s, \mathcal{L}(X_s))\,\mathrm{d}B_s, \quad t \in [0,T]. \tag{1.10}$$

We consider two different settings of the mean-field SVE (1.10). First, a multi-dimensional setting under Lipschitz conditions on μ and σ where the kernels K_μ and K_σ are allowed to be singular. Then, where we adapt the techniques of Yamada–Watanabe [YW71], a one-dimensional setting with regular or differentiable convolutional kernels, where μ is Lipschitz continuous and σ is allowed to be only $1/2 + \xi$-Hölder continuous for some $\xi \in [0, 1/2]$ but independent of the law $\mathcal{L}(X_s)$.

In each of the settings, we prove the well-posedness of the mean-field SVE (1.10), see Theorem 5.3 and Theorem 5.10, and a quantitative, pointwise propagation of chaos result of Volterra-type systems of interacting particles, see Theorem 5.4 and Theorem 5.11.

Neural SVEs

In Chapter 6, we consider the supervised learning problem for SVEs, i.e. the challenge to learn the dynamics of an SVE given training data consisting of sample paths and the underlying Brownian paths. Therefore, to learn some d-dimensional SVE, we introduce for given initial value $\xi \in \mathbb{R}^d$ and Brownian path $(B_t)_{t\in[0,T]}$ the neural stochastic Volterra equation (neural SVE) defined by

$$\begin{aligned}
Z_0 &= L_\theta(\xi),\\
Z_t &= Z_0\, g_\theta(t) + \int_0^t K_{\mu,\theta}(t-s)\mu_\theta(s, Z_s)\,\mathrm{d}s + \int_0^t K_{\sigma,\theta}(t-s)\sigma_\theta(s, Z_s)\,\mathrm{d}B_s,\\
X_t &= \Pi_\theta(Z_t), \quad t \in [0,T],
\end{aligned}$$

where all the components that are sub scripted by θ are neural networks, $(Z_t)_{t\in[0,T]}$ is an $\mathbb{R}^{d_h}$-valued hidden process for some latent dimension $d_h > d$ and $(X_t)_{t\in[0,T]}$ aims to approximate the true d-dimensional Volterra path. We show that neural SVEs are able to learn the dynamics of SVEs exemplary with the disturbed pendulum equation, the general Ornstein–Uhlenbeck equation and the Rough Heston equation.

Chapter 2

SVEs with regular kernels

The content of this chapter is published in [PS23a].

Introduction

In this chapter, we investigate the strong existence and pathwise uniqueness of solutions to one-dimensional stochastic Volterra equations with locally Hölder continuous diffusion coefficients and sufficiently regular kernels. More precisely, we consider SVEs of the form

$$X_t = x_0(t) + \int_0^t K_\mu(s,t)\mu(s,X_s)\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s,X_s)\,\mathrm{d}B_s, \quad t \in [0,T], \tag{2.1}$$

where x_0 denotes the initial condition, $(B_t)_{t\in[0,T]}$ is a Brownian motion, the kernels K_μ, K_σ are sufficiently regular functions, the coefficient μ is locally Lipschitz continuous, and the diffusion coefficient σ is only locally Hölder continuous.

While the existence of unique strong solutions for stochastic Volterra equations with Lipschitz continuous coefficients is well investigated, our motivation to study stochastic Volterra equations with non-Lipschitz coefficients comes from the natural question to explore to what extent the famous results of Yamada and Watanabe [YW71], ensuring pathwise uniqueness and the existence of strong solutions for ordinary stochastic differential equations, generalizes to stochastic Volterra equations.

The classical approach to prove the existence of strong solutions to ordinary stochastic differential equations with less regular diffusion coefficients is to first show the existence of a weak solution, since this, in combination with pathwise uniqueness, guarantees the existence of a strong solution, see [YW71]. In this chapter, we will not use that common technique.

Instead, our first main contribution of the chapter is to directly establish the existence of a strong solution to the SVE (2.1) provided the diffusion coefficient σ is locally $1/2+\xi$-Hölder continuous for $\xi \in [0,1/2]$. To that end, we prove the convergence of an Euler type approximation of the SVE (2.1) and do not use the concept of weak solutions. For ordinary stochastic differential equations such an approach was developed by Gyöngy

and Rásonyi [GR11], using ideas coming from [YW71]. As a number of results used to deal with ordinary stochastic differential equations are not available in the context of SVEs, the presented proof for the existence of a strong solution to the SVE (2.1) requires various different techniques such as a transformation formula for Volterra processes à la Protter [Pro85] and a Grönwall lemma allowing weakly singular kernels.
Our second main contribution is to establish pathwise uniqueness for the SVE (2.1) provided that the diffusion coefficient σ is locally $1/2+\xi$-Hölder continuous for $\xi \in [0,1/2]$ or even, more generally, satisfies the classical Yamada–Watanabe condition [YW71]. To that end, we generalize the classical approach of Yamada and Watanabe [YW71] to the more general setting of stochastic Volterra equations. The presented proof for pathwise uniqueness is based on similar techniques as the proof of existence and is inspired by the work of Mytnik and Salisbury [MS15]. In [MS15], pathwise uniqueness is proven for one-dimensional stochastic Volterra equations with smooth kernels and without drift (i.e. $\mu = 0$). For SVEs of convolutional type with continuous differentiable kernels admitting a resolvent of the first kind, pathwise uniqueness was shown in [AJEE19b].
Let us remark, while we need to require sufficient regularity on the kernels K_μ, K_σ to obtain the existence of a unique strong solution (see Theorem 2.3 and Corollary 2.6), the imposed regularity conditions on the coefficients are essentially the classical regularity conditions of Yamada–Watanabe. Already in case of ordinary stochastic differential equations, it is well-known that these regularity conditions cannot be relaxed in the sense that pathwise uniqueness does not hold in general if, e.g., the diffusion coefficient σ is only Hölder continuous of order strictly less than $1/2$.

Organization of the chapter: Section 2.1 presents the setting and main result: an existence and uniqueness theorem for stochastic Volterra equations with Hölder continuous diffusion coefficients. The properties of solutions to SVEs are provided in Section 2.2. The existence of a strong solution is proven in Section 2.3 and that pathwise uniqueness holds in Section 2.4.

2.1 Main result and assumptions

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space, which satisfies the usual conditions, $(B_t)_{t\in[0,T]}$ be a standard Brownian motion and $T \in (0,\infty)$. We consider the one-dimensional stochastic Volterra equation (SVE)

$$X_t = x_0(t) + \int_0^t K_\mu(s,t)\mu(s,X_s)\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s,X_s)\,\mathrm{d}B_s, \quad t \in [0,T], \tag{2.2}$$

where $x_0\colon [0,T] \to \mathbb{R}$ is a continuous function, the coefficients $\mu, \sigma\colon [0,T]\times\mathbb{R} \to \mathbb{R}$ and the kernels $K_\mu, K_\sigma\colon \Delta_T \to \mathbb{R}$ are measurable functions, using the standard notation $\Delta_T := \{(s,t) \in [0,T]\times[0,T]\colon 0 \le s \le t \le T\}$. Furthermore, $\int_0^t K_\mu(s,t)\mu(s,X_s)\,\mathrm{d}s$ is defined as a Lebesque integral and $\int_0^t K_\sigma(s,t)\sigma(s,X_s)\,\mathrm{d}B_s$ as an Itô integral.

Let $K\colon \Delta_T \to \mathbb{R}$ be a measurable function. We say $K(\cdot,t)$ is absolutely continuous for every $t\in[0,T]$ if there exists an integrable function $\partial_1 K\colon \Delta_T \to \mathbb{R}$ such that $K(s,t) -$

$K(0,t)=\int_0^s \partial_1 K(u,t)\,\mathrm{d}u$ for $(s,t)\in\Delta_T$. We say $K(s,\cdot)$ is absolutely continuous for every $s\in[0,T]$ if there exists an integrable function $\partial_2 K\colon \Delta_T\to\mathbb{R}$ such that $K(s,t)-K(s,0)=\int_0^t \partial_2 K(s,u)\,\mathrm{d}u$ for $(s,t)\in\Delta_T$. Moreover, for $p\in[1,\infty)$, we denote $K\in L^p(\Delta_T)$ if $\int_0^T\int_0^t |K(s,t)|^p\,\mathrm{d}s\,\mathrm{d}t<\infty$.

For the kernels K_μ, K_σ and the initial condition x_0 we make the following assumptions.

Assumption 2.1. *Let $\gamma\in(0,\frac{1}{2}]$, and $K_\mu, K_\sigma\colon\Delta_T\to\mathbb{R}$ and $x_0\colon[0,T]\to\mathbb{R}$ be continuous functions such that:*

(i) $K_\mu(s,\cdot)$ is absolutely continuous for every $s\in[0,T]$ and $\partial_2 K_\mu$ is bounded on Δ_T.

(ii) $K_\sigma(\cdot,t)$ is absolutely continuous for every $t\in[0,T]$, $K_\sigma(s,\cdot)$ is absolutely continuous for every $s\in[0,T]$ with $\partial_2 K_\sigma\in L^2(\Delta_T)$, and $\partial_2 K_\sigma(\cdot,t)$ is absolutely continuous for every $t\in[0,T]$. Furthermore, there is a constant $C>0$ such that $|K_\sigma(t,t)|\geq C$ for any $t\in[0,T]$, and there exist $C>0$, $\alpha\in[0,\frac{1}{2})$ and $\varepsilon>0$ such that

$$\int_0^s |K_\sigma(u,t)-K_\sigma(u,s)|^{2+\varepsilon}\,\mathrm{d}u \leq C|t-s|^{\gamma(2+\varepsilon)} \quad \textit{and}$$
$$|\partial_1 K_\sigma(s,t)|+|\partial_2 K_\sigma(s,s)|+\int_s^t |\partial_{21} K_\sigma(s,u)|\,\mathrm{d}u \leq C(t-s)^{-\alpha}$$

hold for any $(s,t)\in\Delta_T$.

(iii) x_0 is β-Hölder continuous for every $\beta\in(0,\gamma)$.

The regularity properties of the coefficients μ and σ are formulated in the next assumption. We start with assuming global Lipschitz and Hölder continuity of μ and σ, respectively. An extension to local regularity conditions is treated in Corollary 2.6 below.

Assumption 2.2. *Let $\mu,\sigma\colon[0,T]\times\mathbb{R}\to\mathbb{R}$ be measurable functions such that:*

(i) μ and σ are of linear growth, i.e. there is a constant $C_{\mu,\sigma}>0$ such that

$$|\mu(t,x)|+|\sigma(t,x)|\leq C_{\mu,\sigma}(1+|x|),$$

for all $t\in[0,T]$ and $x\in\mathbb{R}$.

(ii) μ is Lipschitz continuous and σ is Hölder continuous of order $\frac{1}{2}+\xi$ for some $\xi\in[0,\frac{1}{2}]$ in the space variable uniformly in time, i.e. there are constants $C_\mu, C_\sigma>0$ such that

$$|\mu(t,x)-\mu(t,y)|\leq C_\mu|x-y| \quad \textit{and} \quad |\sigma(t,x)-\sigma(t,y)|\leq C_\sigma|x-y|^{\frac{1}{2}+\xi}$$

hold for all $t\in[0,T]$ and $x,y\in\mathbb{R}$.

To formulate our results, let us briefly recall the concepts of strong solutions and pathwise uniqueness. For this purpose, let $L^p(\Omega \times [0,T])$ be the space of all real-valued, p-integrable functions on $\Omega \times [0,T]$. We call an $(\mathcal{F}_t)_{t\in[0,T]}$-progressively measurable stochastic process $(X_t)_{t\in[0,T]}$ in $L^p(\Omega \times [0,T])$ on the given probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, a *(strong) L^p-solution* of the SVE (2.2) if $\int_0^t (|K_\mu(s,t)\mu(s,X_s)| + |K_\sigma(s,t)\sigma(s,X_s)|^2)\,\mathrm{d}s < \infty$ for all $t \in [0,T]$ and the integral equation (2.2) hold $\mathbb{P}$-almost surely. As usual, a strong L^1-solution $(X_t)_{t\in[0,T]}$ of the SVE (2.2) is often just called *solution* of the SVE (2.2). We say *pathwise uniqueness* in $L^p(\Omega \times [0,T])$ holds for the SVE (2.2) if $\mathbb{P}(X_t = \tilde{X}_t,\, \forall t \in [0,T]) = 1$ for two L^p-solutions $(X_t)_{t\in[0,T]}$ and $(\tilde{X}_t)_{t\in[0,T]}$ of the SVE (2.2) defined on the same probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$. Moreover, we say there exists a *unique strong L^p-solution* $(X_t)_{t\in[0,T]}$ to the SVE (2.2) if $(X_t)_{t\in[0,T]}$ is a strong L^p-solution to the SVE (2.2) and pathwise uniqueness in $L^p(\Omega \times [0,T])$ holds for the SVE (2.2). We say $(X_t)_{t\in[0,T]}$ is β-Hölder continuous for $\beta \in (0,1]$ if there exists a modification of $(X_t)_{t\in[0,T]}$ with sample paths that are $\mathbb{P}$-almost surely β-Hölder continuous.

The main results of the present work are summarized in the following theorem.

Theorem 2.3. *Suppose Assumptions 2.1 and 2.2, and let $p > \max\{\frac{1}{\gamma}, 1 + \frac{2}{\varepsilon}\}$, where $\gamma \in (0, \frac{1}{2}]$ and $\varepsilon > 0$ are given by Assumption 2.1. Then, there exists a unique strong L^p-solution $(X_t)_{t\in[0,T]}$ to the stochastic Volterra equation* (2.2)*. Moreover, the solution $(X_t)_{t\in[0,T]}$ is β-Hölder continuous for every $\beta \in (0,\gamma)$, $\sup_{t\in[0,T]} \mathbb{E}[|X_t|^q] < \infty$ for every $q \in [1,\infty)$ and $(X_t - x_0(t))_{t\in[0,T]}$ is a semimartingale.*

Proof. The existence of a strong solution $(X_t)_{t\in[0,T]}$ to the stochastic Volterra equation (2.2) is provided by Theorem 2.14 and its pathwise uniqueness by Theorem 2.22. The assertions that $\sup_{t\in[0,T]} \mathbb{E}[|X_t|^q] < \infty$ for every $q \in [1,\infty)$ and of the β-Hölder continuity as well as the semimartingale property of $(X_t - x_0(t))_{t\in[0,T]}$ follow by Corollary 2.13. □

Note that the regularity assumptions (Assumption 2.2), as required in Theorem 2.3, on the coefficients μ, σ are essentially optimal. Indeed, it is well-known for ordinary stochastic differential equations that pathwise uniqueness does not hold in general if μ is only Hölder continuous of order strictly less than 1 or σ is only Hölder continuous of order strictly less than 1/2, see for instance [KS91, page 287] and [KS91, Chapter 5, Example 2.15].

Remark 2.4. *Recall that Yamada and Watanabe derived pathwise uniqueness for ordinary stochastic differential equations under the slightly weaker assumption of $|\sigma(t,x) - \sigma(t,y)| \leq \rho(|x-y|)$ for a function $\rho\colon [0,\infty) \to [0,\infty)$ with $\int_0^\varepsilon \rho(s)^{-2}\,\mathrm{d}s = \infty$ for every $\varepsilon > 0$, cf. [YW71, Theorem 1]. While the proof of pathwise uniqueness presented in Section 2.4 is given under this Yamada–Watanabe condition, in the proof of the existence of a strong solution via an approximation scheme the Hölder regularity of σ is explicitly used in various estimates, see e.g.* (2.14)*, and a modification of these estimates allowing for the Yamada–Watanabe condition appears not straightforward.*

Remark 2.5. *Assumption 2.1 is satisfied, for instance, if K_μ is continuously differentiable, K_σ is twice continuously differentiable with $K_\sigma(t,t) > 0$ for $t \in [0,T]$ and x_0 is β-Hölder continuous for some $\beta \in (0,1)$.*

While the condition $|K_\sigma(t,t)| \geq C$ for $t \in [0,T]$ is crucial for implementing the present method to prove Theorem 2.3, it might appear to be of technical nature. However, assuming $K_\sigma(t,t) = 0$ for every $t \in [0,T]$ and keeping in mind the semimartingale decomposition in Lemma 2.12, any solution of the SVE (2.2) *would be a semimartingale of bounded variation without any diffusion part and, thus, some care is needed to not lose the regularization effects of a Brownian motion.*

Based on a localization argument, the assumptions of global Lipschitz and Hölder continuity on the coefficients of the SVE (2.2) can be relaxed to local regularity assumptions. In the following, $C > 0$ denotes a generic constant that might change from line to line. To emphasize the dependence of the constant C on parameters p, q or functions f, g, we write $C_{p,q,f,g}$. Moreover, for $x, y \in \mathbb{R}$ we set $x \wedge y := \min\{x, y\}$.

Corollary 2.6. *Suppose Assumptions 2.1, 2.2 (i), and that μ is locally Lipschitz continuous and σ is locally Hölder continuous of order $\frac{1}{2} + \xi$ for some $\xi \in [0, \frac{1}{2}]$ in the space variable uniformly in time, i.e. for every $n \in \mathbb{N}$ there are constants $C_{\mu,n}, C_{\sigma,n} > 0$ such that*

$$|\mu(t,x) - \mu(t,y)| \leq C_{\mu,n}|x-y| \quad \textit{and} \quad |\sigma(t,x) - \sigma(t,y)| \leq C_{\sigma,n}|x-y|^{\frac{1}{2}+\xi}$$

hold for all $t \in [0,T]$ and $x, y \in \mathbb{R}$ with $|x|, |y| \leq n$. Let $p > \max\{\frac{1}{\gamma}, 1 + \frac{2}{\varepsilon}\}$, where $\gamma \in (0, \frac{1}{2}]$ and $\varepsilon > 0$ are given by Assumption 2.1. Then, there exists a unique strong L^p-solution $(X_t)_{t\in[0,T]}$ to the stochastic Volterra equation (2.2)*. Moreover, the solution $(X_t)_{t\in[0,T]}$ is β-Hölder continuous for every $\beta \in (0,\gamma)$, $\sup_{t\in[0,T]} \mathbb{E}[|X_t|^q] < \infty$ for every $q \in [1,\infty)$ and $(X_t - x_0(t))_{t\in[0,T]}$ is a semimartingale.*

Proof. By Assumptions 2.1 and 2.2 (i), Lemma 2.10, Corollary 2.11 and Lemma 2.12 imply the integrability, β-Hölder continuity and semimartingale property of the solution. For the well-posedness, we adapt the proofs of Theorem 2.14 and 2.22 and the notation therein.
For the uniqueness, consider two L^p-solutions $(X_t^1)_{t\in[0,T]}$ and $(X_t^2)_{t\in[0,T]}$, and define $\tilde{X}_t := X_t^1 - X_t^2$ for $t \in [0,T]$ and the hitting times $\tau_k := \inf\{t \in [0,T] \colon \max\{|X_t|, |Y_t|\} \geq k\} \wedge T$ for $k \in \mathbb{N}$ which are stopping times with $\tau_k \to T$ a.s. by the same reasoning as for the hitting times defined in (2.4). By bounding $\phi_n(\tilde{X}_t \mathbb{1}_{\{t\leq\tau_k\}}) \leq \phi_n(\tilde{X}_{t\wedge\tau_k})$ and applying Itô's formula to the right-hand-side, we obtain after performing the same steps as in (2.22)-(2.27) and sending $n \to \infty$, that

$$\begin{aligned}
&\mathbb{E}[|\tilde{X}_t|\mathbb{1}_{\{t\leq\tau_k\}}] \\
&\quad\leq C\int_0^t \mathbb{E}[|\tilde{X}_s|\mathbb{1}_{\{s\leq\tau_k\}}]\,\mathrm{d}s + \int_0^t \mathbb{E}[|\tilde{Y}_s|\mathbb{1}_{\{s\leq\tau_k\}}]\left(\partial_2 K_\sigma(s,s) + \int_s^t |\partial_{21}K_\sigma(s,u)|\,\mathrm{d}u\right)\mathrm{d}s,
\end{aligned}$$

for $t \in [0,T]$. Similarly, we get a bound on $\mathbb{E}[|\tilde{Y}_t|\mathbb{1}_{\{t\leq\tau_k\}}]$ analogue to (2.30) and denoting

$$M_k(t) := \sup_{s\in[0,t]}\left(\mathbb{E}[|\tilde{X}_s|\mathbb{1}_{\{s\leq\tau_k\}}] + \mathbb{E}[|\tilde{Y}_s|\mathbb{1}_{\{s\leq\tau_k\}}]\right)$$

we obtain $M_k(t) = 0$ for all $t \in [0,T]$, and sending $k \to \infty$ yields the uniqueness.
For the existence, we adapt the standard localization argument from the SDE case. We introduce for $n \in \mathbb{N}$ the localized coefficients

$$\mu_n(t,x) := \begin{cases} \mu(t,x), & \text{if } |x| \le n, \\ \mu(t, \frac{nx}{|x|}), & \text{if } |x| > n, \end{cases}$$

and analogously σ_n, which fulfill the regularity properties globally, such that corresponding strong solutions exist by Theorem 2.14 that we denote by X^n. Moreover, let $\kappa_n := \inf\{t \in [0,T] \colon |X_t^n| > n\} \wedge T$ and define $X(t) := X^n(t)$ for $\kappa_{n-1} < t \le \kappa_n(t)$. By the pathwise uniqueness, it holds $X^{n-1}_{\tau_{n-1}} = X^n_{\tau_{n-1}}$ for all $n \in \mathbb{N}$ such that X is continuously well-defined and we must only show that it cannot explode, i.e. that $\kappa_n \to T$ a.s. By the Garsia–Rodemich–Rumsey inequality (see [GRR71, Lemma 1.1]), Markov's inequality and Lemma 2.7, we obtain for any $\alpha \in (0,\gamma)$ and $p > 2$ chosen such that $\alpha p > 1$ that

$$\begin{aligned} \mathbb{P}\Big(\sup_{t \in [0,T]} |X_t^n - X_0^n| > n \Big) &\le \mathbb{P}\bigg(\sup_{t\in[0,T]} \Big(C_{\alpha,p} t^{\alpha - \frac{1}{p}} \Big(\int_0^T \int_0^T \frac{|X_s - X_u|^p}{|s-u|^{\alpha p + 1}} \,\mathrm{d}u \,\mathrm{d}s \Big)^{\frac{1}{p}} \Big) > n \bigg) \\ &\le n^{-p} \mathbb{E}\bigg[C_{\alpha,p,T} \Big(\int_0^T \int_0^T \frac{|X_s - X_u|^p}{|s-u|^{\alpha p+1}} \,\mathrm{d}u \,\mathrm{d}s \Big) \bigg] \\ &\le C_{\alpha,p,T,\mu,\sigma,\varepsilon} n^{-p}, \end{aligned}$$

which tends to 0 sufficiently fast such that the Borel–Cantelli lemma (see [Kle14, Theorem 2.7]) implies $\kappa_n \to T$ a.s. □

The rest of the chapter is largely devoted to prove Theorem 2.3. However, we will formulate and prove the partial findings under weaker assumptions if possible without additional effort.

2.2 Properties of a solution

In this section we establish some properties of solutions to stochastic Volterra equations. We start by the regularity and integrability of L^p-solutions, which requires only the linear growth condition of the coefficients and allows for singular kernels in the SVE (2.2).

Lemma 2.7. *Suppose Assumption 2.2 (i) and let $K_\mu, K_\sigma \colon \Delta_T \to \mathbb{R}$ be measurable functions such that, for some $\varepsilon > 0$ and $L > 0$,*

$$\begin{aligned} \int_0^t |K_\mu(s,t') - K_\mu(s,t)|^{1+\varepsilon} \,\mathrm{d}s + \int_t^{t'} |K_\mu(s,t')|^{1+\varepsilon} \,\mathrm{d}s &\le L|t'-t|^{\gamma(1+\varepsilon)}, \\ \int_0^t |K_\sigma(s,t') - K_\sigma(s,t)|^{2+\varepsilon} \,\mathrm{d}s + \int_t^{t'} |K_\sigma(s,t')|^{2+\varepsilon} \,\mathrm{d}s &\le L|t'-t|^{\gamma(2+\varepsilon)}, \end{aligned} \tag{2.3}$$

for all $(t,t') \in \Delta_T$. Furthermore, let $x_0 \colon [0,T] \to \mathbb{R}$ be β-Hölder continuous for every $\beta \in (0,\gamma)$ for some $\gamma \in (0, \frac{1}{2}]$ and let $(X_t)_{t\in[0,T]}$ be a L^p-solution of the SVE (2.2) for *some* $p > \max\{\frac{1}{\gamma}, 1+\frac{2}{\varepsilon}\}$*. Then, for any* $\beta \in (0,\gamma)$*, there is a constant* $C_{x_0,p,L,T,\mu,\sigma,\varepsilon} > 0$ *such that*

$$\mathbb{E}[|X_{t'} - X_t|^p] \le C_{x_0,p,L,T,\mu,\sigma,\varepsilon}|t'-t|^{\beta p},$$

holds for all $t,t' \in [0,T]$*. Consequently,* $(X_t)_{t\in[0,T]}$ *is* β*-Hölder continuous for any* $\beta \in (0,\gamma-\frac{1}{p})$*.*

Proof. Let $p>2$ be given by the assumption. Since x_0 is β-Hölder continuous, we observe for $t,t' \in [0,T]$ that

$$\mathbb{E}[|X_{t'} - X_t|^p] \le C_{p,x_0}|t'-t|^{\beta p} + C_p\mathbb{E}[|\tilde{X}_{t'} - \tilde{X}_t|^p] \quad \text{with} \quad \tilde{X}_t := X_t - x_0(t).$$

For $(t,t') \in \Delta_T$ we note that

$$\begin{aligned}
|\tilde{X}_{t'} - \tilde{X}_t|^p =& \bigg| \int_0^{t'} K_\mu(s,t')\mu(s,X_s)\,\mathrm{d}s + \int_0^{t'} K_\sigma(s,t')\sigma(s,X_s)\,\mathrm{d}B_s \\
&- \int_0^t K_\mu(s,t)\mu(s,X_s)\,\mathrm{d}s - \int_0^t K_\sigma(s,t)\sigma(s,X_s)\,\mathrm{d}B_s\bigg|^p \\
&\le C_p\bigg(\bigg|\int_0^t \mu(s,X_s)(K_\mu(s,t')-K_\mu(s,t))\,\mathrm{d}s\bigg|^p + \bigg|\int_t^{t'} \mu(s,X_s)K_\mu(s,t')\,\mathrm{d}s\bigg|^p \\
&\quad + \bigg|\int_0^t \sigma(s,X_s)(K_\sigma(s,t')-K_\sigma(s,t))\,\mathrm{d}B_s\bigg|^p + \bigg|\int_t^{t'} \sigma(s,X_s)K_\sigma(s,t')\,\mathrm{d}B_s\bigg|^p\bigg) \\
&=: C_p(A+B+C+D).
\end{aligned}$$

We shall bound the expectation of the terms A-D in the following. For A, we use Hölder's inequality, the linear growth of μ (Assumption 2.2 (i)), (2.3) and that $X \in L^{\frac{1+\varepsilon}{\varepsilon}}(\Omega\times[0,T])$ since $\frac{1+\varepsilon}{\varepsilon} < p$ to obtain

$$\begin{aligned}
\mathbb{E}[A] &\le \mathbb{E}\bigg[\bigg|\int_0^t |\mu(s,X_s)|^{\frac{1+\varepsilon}{\varepsilon}}\,\mathrm{d}s\bigg|^{\frac{p\varepsilon}{1+\varepsilon}}\bigg]\bigg(\int_0^t \big|K_\mu(s,t')-K_\mu(s,t)\big|^{1+\varepsilon}\,\mathrm{d}s\bigg)^{\frac{p}{1+\varepsilon}} \\
&\le C_{p,L,\mu,T,\varepsilon}\bigg(\int_0^t \big|K_\mu(s,t')-K_\mu(s,t)\big|^{1+\varepsilon}\,\mathrm{d}s\bigg)^{\frac{p}{1+\varepsilon}} \\
&\le C_{x_0,p,L,T,\mu,\sigma,\varepsilon}|t'-t|^{\gamma p}.
\end{aligned}$$

Note that the second inequality follows either with Jensen's inequality, if $\frac{p\varepsilon}{1+\varepsilon} \le 1$, or else with Hölder's inequality and Fubini's theorem. Applying the analog estimates to B gives

$$\mathbb{E}[B] \le \mathbb{E}\bigg[\bigg|\int_t^{t'} |\mu(s,X_s)|^{\frac{1+\varepsilon}{\varepsilon}}\,\mathrm{d}s\bigg|^{\frac{p\varepsilon}{1+\varepsilon}}\bigg]\bigg(\int_t^{t'} \big|K_\mu(s,t')\big|^{1+\varepsilon}\,\mathrm{d}s\bigg)^{\frac{p}{1+\varepsilon}} \le C_{x_0,p,L,T,\mu,\sigma,\varepsilon}|t'-t|^{\gamma p}.$$

For term C, relying on the Burkholder–Davis–Gundy inequality, Hölder's inequality, using the linear growth of σ (Assumption 2.2 (i)), $X \in L^{\frac{2+\varepsilon}{\varepsilon}}(\Omega \times [0,T])$ and (2.3), we get

$$\begin{aligned}
\mathbb{E}[C] &\leq \mathbb{E}\bigg[\bigg(\int_0^t \big|\sigma(s,X_s)\big(K_\sigma(s,t') - K_\sigma(s,t)\big)\big|^2 \,\mathrm{d}s\bigg)^{\frac{p}{2}}\bigg] \\
&\leq \mathbb{E}\bigg[\Big|\int_0^t |\sigma(s,X_s)|^{\frac{2+\varepsilon}{\varepsilon}} \,\mathrm{d}s\Big|^{\frac{p\varepsilon}{4+2\varepsilon}}\bigg]\bigg(\int_0^t \big|K_\sigma(s,t') - K_\sigma(s,t)\big|^{2+\varepsilon} \,\mathrm{d}s\bigg)^{\frac{p}{2+\varepsilon}} \\
&\leq C_{p,L,\sigma,T,\varepsilon}\bigg(\int_0^t \big|K_\sigma(s,t') - K_\sigma(s,t)\big|^{2+\varepsilon} \,\mathrm{d}s\bigg)^{\frac{p}{2+\varepsilon}} \\
&\leq C_{x_0,p,L,T,\mu,\sigma,\varepsilon}|t'-t|^{\gamma p}.
\end{aligned}$$

Applying (2.3) and analog estimates to term D reveals

$$\mathbb{E}[D] \leq C_{x_0,p,L,T,\mu,\sigma,\varepsilon}\bigg(\int_t^{t'} K_\sigma(s,t')^{2+\varepsilon} \,\mathrm{d}s\bigg)^{\frac{p}{2+\varepsilon}} \leq C_{x_0,p,L,T,\mu,\sigma,\varepsilon}|t'-t|^{\gamma p}.$$

Hence, with the above estimates we arrive at

$$\mathbb{E}[|X_{t'} - X_t|^p] \leq C_{p,x_0}|t'-t|^{\beta p} + C_{x_0,p,L,T,\mu,\sigma}|t'-t|^{\gamma p} \leq C_{x_0,p,L,T,\mu,\sigma,\varepsilon}|t'-t|^{\beta p},$$

as $\beta < \gamma$. Hence, by Kolmogorov–Chentsov's theorem (see e.g. [Kle14, Theorem 21.6]) and sending $\beta \to \gamma$, there exists a modification of $(X_t)_{t\in[0,T]}$ which is δ'-Hölder continuous for $\delta' \in (0, \gamma - 1/p)$. □

Remark 2.8. *Suppose that the kernels K_μ and K_σ fulfill Assumption 2.1. In this case it follows from Kolmogorov's continuity criterion and the estimates in the proof of Lemma 2.7, that, for every progressively measurable stochastic process $u \in L^p([0,T]\times\Omega)$ for some $p > \max\{\frac{1}{\gamma}, 1+\frac{2}{\varepsilon}\}$, the process $(M^u_t)_{t\in[0,T]}$, defined by $M^u_t := \int_0^t K_\mu(s,t)u_s \,\mathrm{d}s + \int_0^t K_\sigma(s,t)u_s \,\mathrm{d}B_s$, has $\mathbb{P}$-a.s. β-Hölder-continuous paths for every $\beta \in (0, \gamma - \frac{1}{p})$.*

Remark 2.9. *Note that the constant $C_{x_0,p,L,T,\mu,\sigma,\varepsilon}$ in Lemma 2.7 depends on μ and σ only through the constant appearing in the linear growth condition (Assumption 2.2 (i)).*

The integrability of solutions to the SVE (2.2) is the content of the next lemma.

Lemma 2.10. *Suppose that the assumptions of Lemma 2.7 hold. Then,*

$$\sup_{t\in[0,T]} \mathbb{E}[|X_t|^q] \leq C_{q,L,T,\mu,\sigma}\bigg(1 + \sup_{t\in[0,T]} |x_0(t)|^q\bigg),$$

holds for any $q \geq 1$, where the constant $C_{q,L,T,\mu,\sigma}$ depends only on q, L, T and the growth constants of μ and σ.

Proof. Let us introduce the hitting times

$$\tau_k := \inf\{t \in [0,T] \colon |X_t| \geq k\} \wedge T, \quad \text{for } k \in \mathbb{N}. \tag{2.4}$$

Note that $\tau_k \to T$ a.s. as $k \to \infty$, since the paths of the solution X are $\mathbb{P}$-a.s. continuous by Lemma 2.7. Since the underlying filtered probability space satisfies the usual conditions, by the Début theorem (see [RY99, Chapter I, (4.15) Theorem]), the hitting times $(\tau_k)_{k\in\mathbb{N}}$ are stopping times.
First, let $q > 2$ be big enough such that $q' := \frac{q}{q-1} \leq 1+\varepsilon$ and $\tilde{q} := \frac{q}{q-2} \leq 1+\varepsilon/2$. Using Hölder's inequality, the Burkholder–Davis–Gundy inequality, and the linear growth condition (Assumption 2.2 (i)), we get

$$
\begin{aligned}
&\mathbb{E}[|X_t|^q \mathbf{1}_{\{t\leq\tau_k\}}] \\
&\quad = \mathbb{E}\bigg[\bigg|x_0(t) + \int_0^t K_\mu(s,t)\mu(s,X_s)\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s,X_s)\,\mathrm{d}B_s\bigg|^q \mathbf{1}_{\{t\leq\tau_k\}}\bigg] \\
&\quad = \mathbb{E}\bigg[\bigg|x_0(t)\,\mathbf{1}_{\{t\leq\tau_k\}} + \int_0^t K_\mu(s,t)\mu(s,X_s)\,\mathrm{d}s\,\mathbf{1}_{\{t\leq\tau_k\}} + \int_0^t K_\sigma(s,t)\sigma(s,X_s)\,\mathrm{d}B_s\,\mathbf{1}_{\{t\leq\tau_k\}}\bigg|^q\bigg] \\
&\quad \leq C_q\mathbb{E}\bigg[|x_0(t)|^q + \bigg|\int_0^t K_\mu(s,t)\mu(s,X_s)\,\mathbf{1}_{\{s\leq\tau_k\}}\,\mathrm{d}s\bigg|^q + \bigg|\int_0^t K_\sigma(s,t)\sigma(s,X_s)\,\mathbf{1}_{\{s\leq\tau_k\}}\,\mathrm{d}B_s\bigg|^q\bigg] \\
&\quad \leq C_q\bigg(|x_0(t)|^q + \bigg(\int_0^t |K_\mu(s,t)|^{q'}\,\mathrm{d}s\bigg)^{\frac{q}{q'}} \int_0^t \mathbb{E}[|\mu(s,X_s)|^q\mathbf{1}_{\{s\leq\tau_k\}}]\,\mathrm{d}s \\
&\qquad\qquad + \mathbb{E}\bigg[\bigg(\int_0^t |K_\sigma(s,t)\sigma(s,X_s)|^2\mathbf{1}_{\{s\leq\tau_k\}}\,\mathrm{d}s\bigg)^{\frac{q}{2}}\bigg]\bigg) \\
&\quad \leq C_q\bigg(|x_0(t)|^q + C_{q,\mu}\bigg(\int_0^t |K_\mu(s,t)|^{q'}\,\mathrm{d}s\bigg)^{\frac{q}{q'}} \int_0^t \mathbb{E}[1+|X_s|^q\mathbf{1}_{\{s\leq\tau_k\}}]\,\mathrm{d}s \\
&\qquad\qquad + C_{q,\sigma}\bigg(\int_0^t |K_\sigma(s,t)|^{2\tilde{q}}\,\mathrm{d}s\bigg)^{\frac{q}{2\tilde{q}}} \int_0^t \mathbb{E}[1+|X_s|^q\mathbf{1}_{\{s\leq\tau_k\}}]\,\mathrm{d}s\bigg)
\end{aligned}
\tag{2.5}
$$

for $t \in [0,T]$. Due to (2.3) we arrive at

$$
\mathbb{E}[|X_t|^q\mathbf{1}_{\{t\leq\tau_k\}}] \leq C_{q,L,T,\mu,\sigma}\bigg(1 + |x_0(t)|^q + \int_0^t \mathbb{E}[|X_s|^q\mathbf{1}_{\{s\leq\tau_k\}}]\,\mathrm{d}s\bigg)
$$

and, thus, as $t \mapsto \mathbb{E}[|X_t|^q\mathbf{1}_{\{t\leq\tau_k\}}]$ is bounded by k^q on $[0,T]$, we can apply Grönwall's lemma (see e.g. [Kle14, Lemma 26.9]) to get

$$
\mathbb{E}[|X_t|^q\mathbf{1}_{\{t\leq\tau_k\}}] \leq C_{q,L,T,\mu,\sigma}\bigg(1 + \sup_{t\in[0,T]} |x_0(t)|^q\bigg), \quad t \in [0,T].
$$

Sending $k \to \infty$ and taking the supremum over $[0,T]$ reveals the assertion. The orderedness of the L^p-spaces implies the statement also for $q_2 \in [1,q)$. □

We conclude that the regularity of a solution can be improved.

Corollary 2.11. *Under the assumptions of Lemma 2.7, any L^p-solution to the SVE* (2.2) *for some $p > \max\{\frac{1}{\gamma}, 1+\frac{2}{\varepsilon}\}$ is β-Hölder continuous for any $\beta \in (0,\gamma)$.*

Proof. The statement follows by applying Lemma 2.10 and Lemma 2.7 with $q > 2$ and then sending $q \to \infty$. $\square$

Assuming sufficient regularity of the kernels K_μ, K_σ, every solution to the stochastic Volterra equation (2.2) is essentially a semimartingale as first observed in [Pro85, Theorem 3.3].

Lemma 2.12. *Let $K_\mu, K_\sigma \colon \Delta_T \to \mathbb{R}$ be measurable functions. Suppose $K_\mu(s,\cdot)$ is absolutely continuous for every $s \in [0,T]$ with $\partial_2 K_\mu \in L^1(\Delta_T)$, $K_\sigma(s,\cdot)$ is absolutely continuous for every $s \in [0,T]$ with $\partial_2 K_\sigma \in L^2(\Delta_t)$, and Assumption 2.2 (i) holds. Let $(X_t)_{t\in[0,T]}$ be a solution to the SVE (2.2) such that $\mathbb{E}[|X_t|^2] \le C$ for all $t \in [0,T]$ and some constant C. Then, $(X_t - x_0(t))_{t\in[0,T]}$ is a semimartingale with decomposition $X_t - x_0(t) = M_t + A_t$ where*

$$\begin{aligned}
M_t &:= \int_0^t K_\sigma(s,s)\sigma(s,X_s)\,\mathrm{d}B_s \quad \textit{and} \\
A_t &:= \int_0^t K_\mu(s,s)\mu(s,X_s)\,\mathrm{d}s \\
&\quad + \int_0^t \Big(\int_0^s \partial_2 K_\mu(u,s)\mu(u,X_u)\,\mathrm{d}u + \int_0^s \partial_2 K_\sigma(u,s)\sigma(u,X_u)\,\mathrm{d}B_u \Big)\,\mathrm{d}s
\end{aligned}$$

for $t \in [0,T]$.

Proof. Setting

$$Y_t := \int_0^t \sigma(s,X_s)\,\mathrm{d}B_s \quad \text{and} \quad Z_t := \int_0^t \mu(s,X_s)\,\mathrm{d}s, \quad \text{for } t \in [0,T],$$

and using the absolute continuity of K_μ, K_σ, we get

$$\begin{aligned}
X_t &= \int_0^t K_\mu(s,s)\,\mathrm{d}Z_s + \int_0^t \Big(\int_s^t \partial_2 K_\mu(s,u)\,\mathrm{d}u \Big)\,\mathrm{d}Z_s \\
&\quad + \int_0^t \left(\int_s^t \partial_2 K_\sigma(s,u)\,\mathrm{d}u \right)\,\mathrm{d}Y_s + \int_0^t K_\sigma(s,s)\,\mathrm{d}Y_s.
\end{aligned}$$

Since

$$\mathbb{E}\Big[\int_{\Delta_T} |\partial_2 K_\mu(s,u)\mu(s,X_s)|\,\mathrm{d}s\,\mathrm{d}u \Big] + \mathbb{E}\Big[\int_{\Delta_T} (\partial_2 K_\sigma(s,u)\sigma(s,X_s))^2\,\mathrm{d}s\,\mathrm{d}u \Big] < \infty$$

due to $\mathbb{E}[|X_t|^2] \le C$ for all $t \in [0,T]$, $\partial_2 K_\mu \in L^1(\Delta_T)$ and $\partial_2 K_\sigma \in L^2(\Delta_T)$, we can apply the classical and the stochastic Fubini theorem (see e.g. [Ver12, Theorem 2.2]) to get

$$\begin{aligned}
X_t &= \int_0^t K_\mu(s,s)\,\mathrm{d}Z_s + \int_0^t \Big(\int_0^u \partial_2 K_\mu(s,u)\,\mathrm{d}Z_s \Big)\,\mathrm{d}u \\
&\quad + \int_0^t \left(\int_0^u \partial_2 K_\sigma(s,u)\,\mathrm{d}Y_s \right)\,\mathrm{d}u + \int_0^t K_\sigma(s,s)\,\mathrm{d}Y_s,
\end{aligned}$$

which completes the proof. $\square$

Applying the previous lemmas to the setting of Theorem 2.3 leads to the following corollary.

Corollary 2.13. *Suppose Assumptions 2.1 and 2.2. Let $(X_t)_{t\in[0,T]}$ be a L^p-solution to the SVE* (2.2) *for some $p > \max\{\frac{1}{\gamma}, 1+\frac{2}{\varepsilon}\}$. Then, $(X_t)_{t\in[0,T]}$ satisfies $\sup_{t\in[0,T]}\mathbb{E}[|X_t|^q] < \infty$ for every $q \in [1,\infty)$, $(X_t)_{t\in[0,T]}$ is β-Hölder continuous for every $\beta \in (0,\gamma)$ for $\gamma \in (0,1/2]$ given in Assumption 2.1, and $(X_t - x_0(t))_{t\in[0,T]}$ is a semimartingale.*

Proof. Note that the existence and boundedness of $\partial_2 K_\mu$ from Assumption 2.1 (i) imply that

$$\begin{aligned}\int_0^s |K_\mu(u,t) - K_\mu(u,s)|^{1+\varepsilon}\,\mathrm{d}u &= \int_0^s \Big|\int_s^t \partial_2 K_\mu(u,r)\,\mathrm{d}r\Big|^{1+\varepsilon}\,\mathrm{d}u\\ &\le C|t-s|^\gamma\end{aligned}$$

holds for some $C > 0$ and any $(s,t) \in \Delta_T$, using $\varepsilon > 0$ and $\gamma \in (0,1/2]$ from Assumption 2.1. Hence, $\sup_{t\in[0,T]}\mathbb{E}[|X_t|^q] < \infty$ for every $q \in [1,\infty)$ by Lemma 2.10. Moreover, since Assumption 2.1 implies (2.3), Corollary 2.11 states the claimed β-Hölder continuity. The semimartingale property follows by Lemma 2.12. □

2.3 Existence of a strong solution

This section is devoted to establish the existence of a strong solution to the SVE (2.2):

Theorem 2.14. *Suppose Assumptions 2.1 and 2.2, and let $p > \max\{\frac{1}{\gamma}, 1+\frac{2}{\varepsilon}\}$. Then, there exists a strong L^p-solution $(X_t)_{t\in[0,T]}$ to the SVE* (2.2).

The construction of a strong solution relies on an Euler type approximation. To set up the approximation, we use the sequence $(\pi_m)_{m\in\mathbb{N}}$ of partitions defined by

$$\pi_m := \{t_0^m, \dots, t_{2^{m^5}}^m\} \quad \text{with} \quad t_i^m := \frac{iT}{2^{m^5}} \quad \text{for } i = 0,\dots,2^{m^5}$$

and introduce, for every $m \in \mathbb{N}$, the function $\kappa_m\colon [0,T] \to [0,T]$ by

$$\kappa_m(T) := T \quad \text{and} \quad \kappa_m(t) := t_i^m \quad \text{for } t_i^m \le t < t_{i+1}^m, \quad \text{for } i = 0,1,\dots,2^{m^5}-1.$$

For every $m \in \mathbb{N}$, we iteratively define the process $(X^m(t))_{t\in[0,T]}$ by $X^m(0) := x_0(0)$ and for $t \in (t_i^m, t_{i+1}^m]$ by

$$\begin{aligned}X^m(t) :=& x_0(t) + \int_0^{t_i^m} K_\mu(s,t)\mu(s, X^m(\kappa_m(s)))\,\mathrm{d}s + \int_{t_i^m}^t K_\mu(s,t)\mu(s, X^m(t_i^m))\,\mathrm{d}s\\ &+ \int_0^{t_i^m} K_\sigma(s,t)\sigma(s, X^m(\kappa_m(s)))\,\mathrm{d}B_s + \int_{t_i^m}^t K_\sigma(s,t)\sigma(s, X^m(t_i^m))\,\mathrm{d}B_s,\end{aligned}$$

for $i = 0, \dots, 2^{m^5} - 1$.
Note that we neither discretize the kernels K_μ, K_σ nor the time-component in the coefficients μ, σ. While these additional discretizations might be desirable to derive an implementable numerical scheme, for our purpose of proving the existence of a strong solution, it is sufficient to avoid this additional approximation.

Lemma 2.15. *Suppose Assumptions 2.1 and 2.2. $X^m \in L^q(\Omega \times [0,T])$ for every $m \in \mathbb{N}$ and any $q \in [1, \infty)$. In particular, $X^m \in L^p(\Omega \times [0,T])$ for every $m \in \mathbb{N}$ and $p > \max\{\frac{1}{\gamma}, 1 + \frac{2}{\varepsilon}\}$.*

Proof. For $m \in \mathbb{N}$ and $q \in (2, \infty)$ we define

$$g_m(t) := \mathbb{E}\big[|X^m(t)|^q\big] \quad \text{for } t \in [0,T].$$

To prove that $X^m \in L^q(\Omega \times [0,T])$, it is sufficient to show that the function g_m is bounded on $[0,T]$ since

$$\mathbb{E}\bigg[\int_0^T |X^m(t)|^q \,\mathrm{d}t\bigg] = \int_0^T g_m(t)\,\mathrm{d}t \le T \sup_{t \in [0,T]} g_m(t).$$

For $t = 0$ we have $\mathbb{E}[|X^m(0)|^q] = |x_0(0)|^q < \infty$ and, thus, g_m is bounded on $[0, t_1^m]$. For $t \in (t_i^m, t_{i+1}^m]$ with $i = 1, \dots, 2^{m^5} - 1$, using similar estimates as in (2.5), we iteratively get that

$$\begin{aligned}
&\mathbb{E}[|X^m(t)|^q] \\
&\quad \le C\Bigg(|x_0(t)|^q \\
&\qquad + \mathbb{E}\bigg[\bigg|\int_0^{t_i^m} K_\mu(s,t)\mu(s, X^m(\kappa_m(s)))\,\mathrm{d}s\bigg|^q\bigg] + \mathbb{E}\bigg[\bigg|\int_{t_i^m}^t K_\mu(s,t)\mu(s, X^m(t_i^m))\,\mathrm{d}s\bigg|^q\bigg] \\
&\qquad + \mathbb{E}\bigg[\bigg|\int_0^{t_i^m} K_\sigma(s,t)\sigma(s, X^m(\kappa(s)))\,\mathrm{d}B_s\bigg|^q\bigg] + \mathbb{E}\bigg[\bigg|\int_{t_i^m}^t K_\sigma(s,t)\sigma(s, X^m(t_i^m))\,\mathrm{d}B_s\bigg|^q\bigg]\Bigg) \\
&\quad \le C\bigg(|x_0(t)|^q + \int_0^{t_i^m} \mathbb{E}\big[|\mu(s, X^m(\kappa_m(s)))|^q\big]\,\mathrm{d}s + \int_{t_i^m}^t \mathbb{E}\big[|\mu(s, X^m(t_i^m))|^q\big]\,\mathrm{d}s \\
&\qquad + \int_0^{t_i^m} \mathbb{E}\big[|\sigma(s, X^m(\kappa(s)))|^q\big]\,\mathrm{d}s + \int_{t_i^m}^t \mathbb{E}\big[|\sigma(s, X^m(t_i^m))|^q\big]\,\mathrm{d}s\bigg) \\
&\quad \le C\bigg(1 + \int_0^{t_i^m} \mathbb{E}[|X^m(\kappa(s))|^q]\,\mathrm{d}s + \int_{t_i^m}^t \mathbb{E}[|X^m(t_i^m)|^q]\,\mathrm{d}s\bigg) < \infty.
\end{aligned}$$

Therefore, $\sup_{t \in [0,T]} g_m(t) < \infty$. □

It can be quickly seen that the integrability and regularity results from Section 2.2 also hold for the process $(X^m(t))_{t \in [0,T]}$.

Proposition 2.16. *Suppose Assumptions 2.1 and 2.2. Let $\gamma \in (0, 1/2]$ be as given in Assumption 2.1. Then, for any $m \in \mathbb{N}$, there is a constant $C > 0$ such that*

$$\sup_{t\in[0,T]} \mathbb{E}[|X^m(t)|^q] \le C\bigg(1 + \sup_{t\in[0,T]} |x_0(t)|^q\bigg).$$

holds for any $q \ge 1$. Moreover, for any $\beta \in (0, \gamma)$, there is a constant $C > 0$ such that

$$\mathbb{E}[|X^m(t') - X^m(t)|^q] \le C|t' - t|^{\beta q}$$

holds for all $t', t \in [0, T]$. Consequently, $(X^m(t))_{t\in[0,T]}$ is β-Hölder continuous for any $\beta \in (0, \gamma)$.

Proof. The L^q-bound of $(X^m(t))_{t\in[0,T]}$ follows by similar arguments as used in the proof of Lemma 2.10.
For $t \in (t_i^m, t_{i+1}^m]$ and fixed $m \in \mathbb{N}$ and $q \ge 2$, we get

$$\mathbb{E}[|X^m(t)|^q] \le C\bigg(|x_0(t)|^q + \int_0^{t_i^m} \mathbb{E}[|X^m(\kappa_m(s))|^q] \,\mathrm{d}s + \int_{t_i^m}^t \mathbb{E}[|X^m(t_i^m)|^q] \,\mathrm{d}s\bigg),$$

where we used Hölder's inequality, Burkholder–Davis–Gundy's inequality, and the linear growth condition (Assumption 2.2 (i)). Hence, we arrive at

$$\sup_{u\in[0,t]} \mathbb{E}[|X^m(u)|^q] \le C\bigg(\sup_{u\in[0,T]} |x_0(u)|^q + \int_0^t \sup_{u\in[0,s]} \mathbb{E}[|X^m(u)|^q] \,\mathrm{d}s\bigg).$$

Since $t \mapsto \sup_{u\in[0,t]} \mathbb{E}[|X^m(u)|^q]$ is bounded by the proof of Lemma 2.15, we can apply Grönwall's lemma (see e.g. [Kle14, Lemma 26.9]) to get

$$\sup_{t\in[0,T]} \mathbb{E}[|X^m(t)|^q] \le C\bigg(1 + \sup_{t\in[0,T]} |x_0(t)|^q\bigg), \quad t \in [0, T],$$

which reveals the assertion.
The regularity statement follows by adapting the proof of Lemma 2.7. Indeed, the regularity assumption on the kernels (Assumption 2.1) yields that condition (2.3) is fulfilled. Thus, performing similar estimations as in the proof of Lemma 2.7 and using the just established L^q-bound of X^m, we obtain

$$\mathbb{E}[|X^m(t') - X^m(t)|^q] \le C|t' - t|^{\beta q},$$

for $\beta \in (0, \gamma)$. Hence, by Kolmogorov–Chentsov's theorem (see e.g. [Kle14, Theorem 21.6]), there exists a modification of $(X^m(t))_{t\in[0,T]}$ which is δ'-Hölder continuous for $\delta' \in (0, \beta - 1/q)$. Sending $\beta \to \gamma$ and $q \to \infty$ leads to the claimed Hölder regularity. □

Due to Proposition 2.16, for every $m \in \mathbb{N}$ the process $(X^m(t))_{t\in[0,T]}$ has a continuous modification. Hence, keeping the definition of $(X^m(t))_{t\in[0,T]}$ in mind, we see that $(X^m(t))_{t\in[0,T]}$ fulfills the integral equation

$$X^m(t) = x_0(t)+\int_0^t K_\mu(s,t)\mu(s, X^m(\kappa_m(s)))\,\mathrm{d}s+\int_0^t K_\sigma(s,t)\sigma(s, X^m(\kappa_m(s)))\,\mathrm{d}B_s, \quad (2.6)$$

for $t \in [0,T]$. Moreover, using the just derived regularity estimates of $(X^m(t))_{t\in[0,T]}$, we obtain the following bound.

Corollary 2.17. *Suppose Assumptions 2.1 and 2.2. Then, for any $q, \delta \in (0,\infty)$, there is a constant $C > 0$ such that*

$$\mathbb{E}\left[\left(\int_0^T |X^m(s) - X^m(\kappa_m(s))|^\delta \,\mathrm{d}s\right)^q\right] \le C2^{-\delta q\beta m^5},$$

holds for all $\beta \in (0,\gamma)$ and $m \in \mathbb{N}$.

Proof. Let $\delta > 0$ be fixed. First, assume $q \ge 1$ is sufficiently large such that $q\delta > 2$. For $\beta \in (0,\gamma)$ and $m \in \mathbb{N}$, we use Hölder's inequality, Fubini's theorem and Proposition 2.16 to get

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^T |X^m(s) - X^m(\kappa_m(s))|^\delta \,\mathrm{d}s\right)^q\right] &\le C\mathbb{E}\left[\int_0^T |X^m(s) - X^m(\kappa_m(s))|^{\delta q} \,\mathrm{d}s\right] \\
&= C\int_0^T \mathbb{E}\left[|X^m(s) - X^m(\kappa_m(s))|^{\delta q}\right] \mathrm{d}s \\
&\le C\int_0^T |s-\kappa_m(s)|^{\delta q\beta}\,\mathrm{d}s \\
&\le C2^{-\delta q\beta m^5}. \qquad (2.7)
\end{aligned}$$

For $0 < q \le \frac{2}{\delta}$, we choose $q' > q$ is sufficiently large such that $q'\delta > 2$. Applying Jensen's inequality and (2.7), we obtain

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^T |X^m(s) - X^m(\kappa_m(s))|^\delta \,\mathrm{d}s\right)^q\right] &\le C\mathbb{E}\left[\left(\int_0^T |X^m(s) - X^m(\kappa_m(s))|^\delta \,\mathrm{d}s\right)^{q'}\right]^{\frac{q}{q'}} \\
&\le C2^{-\delta q\beta m^5}.
\end{aligned}$$

□

Lemma 2.18. *Suppose Assumptions 2.1 and 2.2. Then, there is a sequence $(C_m)_{m\in\mathbb{N}}$ of constants such that*

$$\mathbb{E}[|X^{m+1}(t) - X^m(t)|] \le C_m$$

holds for every $t \in [0,T]$, and $\sum_{m=1}^\infty C_m^{1/4} < \infty$.

Proof. Following Gyöngy–Rásonyi [GR11] and Yamada–Watanabe [YW71], we approximate the function $\phi(x) := |x|$ by smooth functions $\phi_{\delta\varepsilon}(x)$ for $\delta > 1$ and $\varepsilon > 0$. To that end, note that

$$\int_{\frac{\varepsilon}{\delta}}^{\varepsilon} \frac{1}{x}\,\mathrm{d}x = \ln(\delta),$$

and, thus, there is a continuous, non-negative function $\psi_{\delta\varepsilon}\colon \mathbb{R}_+ \to \mathbb{R}_+$, that is zero outside the interval $[\frac{\varepsilon}{\delta}, \varepsilon]$, $\int_0^\infty \psi_{\delta\varepsilon}(x)\,\mathrm{d}x = 1$ and satisfies

$$\psi_{\delta\varepsilon}(x) \leq \frac{2}{x\ln(\delta)}.$$

We define

$$\phi_{\delta\varepsilon}(x) := \int_0^{|x|}\int_0^y \psi_{\delta\varepsilon}(z)\,\mathrm{d}z\,\mathrm{d}y \quad \text{for } x \in \mathbb{R},$$

such that the inequalities

$$|x| \leq \phi_{\delta\varepsilon}(x) + \varepsilon, \quad 0 \leq |\phi'_{\delta\varepsilon}(x)| \leq 1 \quad \text{and} \quad \phi''_{\delta\varepsilon}(x) = \psi_{\delta\varepsilon}(|x|) \leq \frac{2}{|x|\ln(\delta)}\mathbf{1}_{[\frac{\varepsilon}{\delta},\varepsilon]}(|x|) \quad (2.8)$$

hold for all $x \in \mathbb{R}$, where $\mathbf{1}_{[\frac{\varepsilon}{\delta},\varepsilon]}$ denotes the indicator function of the interval $[\frac{\varepsilon}{\delta}, \varepsilon]$.
To apply Itô's formula to $\phi_{\delta\varepsilon}(\tilde{X}_t^m)$, where

$$\tilde{X}_t^m := X^{m+1}(t) - X^m(t), \quad t \in [0,T],$$

we need to find the semimartingale decomposition of $(\tilde{X}_t^m)_{t\in[0,T]}$. For this purpose, we introduce the local martingale

$$\tilde{Y}_t^m := Y_t^{m+1} - Y_t^m \quad \text{with} \quad Y_t^m := \int_0^t \sigma(s, X^m(\kappa_m(s)))\,\mathrm{d}B_s$$

and the process of finite variation

$$\tilde{Z}_t^m := \int_0^t \mu(s, X^{m+1}(\kappa_{m+1}(s)))\,\mathrm{d}s - \int_0^t \mu(s, X^m(\kappa_m(s)))\,\mathrm{d}s, \quad \text{for } t \in [0,T].$$

Since $\partial_2 K_\mu \in L^1(\Delta_T)$, $\partial_2 K_\sigma \in L^2(\Delta_T)$ (see Assumption 2.1) and the integrability property of $(X^m(t))_{\in[0,T]}$ as presented in Proposition 2.16, we obtain, as in the proof of Lemma 2.12, the following semimartingale decomposition

$$\begin{aligned}
\tilde{X}_t^m &= \int_0^t K_\mu(s,t)\,\mathrm{d}\tilde{Z}_s^m + \int_0^t K_\sigma(s,t)\,\mathrm{d}\tilde{Y}_s^m \\
&= \int_0^t K_\mu(s,s)\,\mathrm{d}\tilde{Z}_s^m + \int_0^t \Big(\int_0^s \partial_2 K_\mu(u,s)\,\mathrm{d}\tilde{Z}_u^m\Big)\,\mathrm{d}s \\
&\quad + \int_0^t \tilde{H}_s^m\,\mathrm{d}s + \int_0^t K_\sigma(s,s)\,\mathrm{d}\tilde{Y}_s^m,
\end{aligned}$$

where $\tilde{H}_t^m := H_t^{m+1} - H_t^m$ with $H_t^m := \int_0^t \partial_2 K_\sigma(s,t)\,\mathrm{d}Y_s^m$. Note that the quadratic variation of $(\tilde{X}_t^m)_{t\in[0,T]}$ is given by

$$\begin{aligned}\langle \tilde{X}^m\rangle_t &= \left\langle \int_0^\cdot K_\sigma(s,s)\Big(\sigma\big(s, X^{m+1}(\kappa_{m+1}(s))\big) - \sigma\big(s, X^m(\kappa_m(s))\big)\Big)\,\mathrm{d}B_s\right\rangle_t \\ &= \int_0^t K_\sigma(s,s)^2\Big(\sigma\big(s, X^{m+1}(\kappa_{m+1}(s))\big) - \sigma\big(s, X^m(\kappa_m(s))\big)\Big)^2\,\mathrm{d}s, \quad t\in[0,T].\end{aligned}$$

Hence, using (2.8) and applying Itô's formula for fixed $\varepsilon > 0$ and $\delta > 1$ yields

$$\begin{aligned}|\tilde{X}_t^m| &\le \varepsilon + \phi_{\delta\varepsilon}(\tilde{X}_t^m) \\ &= \varepsilon + \int_0^t \phi'_{\delta\varepsilon}(\tilde{X}_s^m)\,\mathrm{d}\tilde{X}_s^m + \frac{1}{2}\int_0^t \phi''_{\delta\varepsilon}(\tilde{X}_s^m)\,\mathrm{d}\langle\tilde{X}^m\rangle_s \\ &= \varepsilon + \int_0^t \phi'_{\delta\varepsilon}(\tilde{X}_s^m)K_\mu(s,s)\,\mathrm{d}\tilde{Z}_s^m + \int_0^t \phi'_{\delta\varepsilon}(\tilde{X}_s^m)\bigg(\int_0^s \partial_2 K_\mu(u,s)\,\mathrm{d}\tilde{Z}_u^m\bigg)\,\mathrm{d}s \\ &\quad + \int_0^t \phi'_{\delta\varepsilon}(\tilde{X}_s^m)\tilde{H}_s^m\,\mathrm{d}s + \int_0^t \phi'_{\delta\varepsilon}(\tilde{X}_s^m)K_\sigma(s,s)\,\mathrm{d}\tilde{Y}_s^m \\ &\quad + \frac{1}{2}\int_0^t \phi''_{\delta\varepsilon}(\tilde{X}_s^m)K_\sigma(s,s)^2\Big(\sigma\big(s, X^{m+1}(\kappa_{m+1}(s))\big) - \sigma\big(s, X^m(\kappa_m(s))\big)\Big)^2\,\mathrm{d}s \\ &=: \varepsilon + I_{1,t}^{\delta\varepsilon} + I_{2,t}^{\delta\varepsilon} + I_{3,t}^{\delta\varepsilon} + I_{4,t}^{\delta\varepsilon} + I_{5,t}^{\delta\varepsilon},\end{aligned} \tag{2.9}$$

for $t\in[0,T]$.
In order to bound $\mathbb{E}[|\tilde{X}_t^m|]$, we shall estimate the five terms appearing in (2.9) separately. We set

$$U_t^m := |X^m(t) - X^m(\kappa_m(t))|, \quad t\in[0,T].$$

For $I_{1,t}^{\delta\varepsilon}$, we use the boundedness of K_μ (Assumption 2.1), the Lipschitz continuity of μ (Assumption 2.2 (ii)) and the bound $\|\phi'_{\delta\varepsilon}\|_\infty \le 1$ to estimate

$$\begin{aligned}\mathbb{E}[I_{1,t}^{\delta\varepsilon}] &= \mathbb{E}\bigg[\int_0^t \phi'_{\delta\varepsilon}(\tilde{X}_s^m)K_\mu(s,s)\Big(\mu\big(s, X^{m+1}(\kappa_{m+1}(s))\big) - \mu\big(s, X^m(\kappa_m(s))\big)\Big)\,\mathrm{d}s\bigg] \\ &\le C\mathbb{E}\bigg[\int_0^t \big(|\tilde{X}_s^m| + U_s^m + U_s^{m+1}\big)\,\mathrm{d}s\bigg].\end{aligned}$$

Since, by Corollary 2.17,

$$\mathbb{E}\bigg[\int_0^t (U_s^m + U_s^{m+1})\,\mathrm{d}s\bigg] \le C2^{-\beta m^5}$$

for any $\beta\in(0,\gamma)$, we get

$$\mathbb{E}[I_{1,t}^{\delta\varepsilon}] \le C\bigg(2^{-\beta m^5} + \int_0^t \mathbb{E}\big[|\tilde{X}_s^m|\big]\,\mathrm{d}s\bigg). \tag{2.10}$$

For $I_{2,t}^{\delta\varepsilon}$, using the boundedness of $\partial_2 K_\mu(u,s)$ on Δ_T (Assumption 2.1), the Lipschitz continuity of μ (Assumption 2.2 (ii)) and the bound $\|\phi'_{\delta\varepsilon}\|_\infty \le 1$, we obtain

$$\begin{aligned}
&\mathbb{E}[I_{2,t}^{\delta\varepsilon}] \\
&= \mathbb{E}\bigg[\int_0^t \phi'_{\delta\varepsilon}(\tilde{X}_s^m)\bigg(\int_0^s \partial_2 K_\mu(u,s)\Big(\mu\big(u, X^{m+1}(\kappa_{m+1}(u))\big) - \mu\big(u, X^m(\kappa_m(u))\big)\Big)\,\mathrm{d}u\bigg)\,\mathrm{d}s\bigg] \\
&\le C\mathbb{E}\bigg[\int_0^t \big(|\tilde{X}_s^m| + U_s^m + U_s^{m+1}\big)\,\mathrm{d}s\bigg].
\end{aligned}$$

Hence, as for $I_{1,t}^{\delta\varepsilon}$, we arrive at

$$\mathbb{E}[I_{2,t}^{\delta\varepsilon}] \le C\bigg(2^{-\beta m^5} + \int_0^t \mathbb{E}\big[|\tilde{X}_s^m|\big]\,\mathrm{d}s\bigg). \tag{2.11}$$

For $I_{3,t}^{\delta\varepsilon}$, we have

$$\mathbb{E}[I_{3,t}^{\delta\varepsilon}] = \mathbb{E}\bigg[\int_0^t \phi'_{\delta\varepsilon}(\tilde{X}_s^m)\tilde{H}_s^m\,\mathrm{d}s\bigg].$$

Noting that an application of the integration by parts formula for semimartingales (cf. [RW00, Theorem (VI).38.3]) gives

$$\tilde{H}_s^m = \int_0^s \partial_2 K_\sigma(u,s)\,\mathrm{d}\tilde{Y}_u^m = \partial_2 K_\sigma(s,s)\tilde{Y}_s^m - \int_0^s \tilde{Y}_u^m \partial_{21} K_\sigma(u,s)\,\mathrm{d}u,$$

we use $\|\phi'_{\delta\varepsilon}\|_\infty \le 1$ and the classical Fubini theorem to get

$$\begin{aligned}
\mathbb{E}[I_{3,t}^{\delta\varepsilon}] &\le \int_0^t \mathbb{E}[|\tilde{H}_s^m|]\,\mathrm{d}s \\
&\le \int_0^t |\partial_2 K_\sigma(s,s)|\mathbb{E}[|\tilde{Y}_s^m|]\,\mathrm{d}s + \int_0^t\int_0^s |\partial_{21}K_\sigma(u,s)|\mathbb{E}[|\tilde{Y}_u^m|]\,\mathrm{d}u\,\mathrm{d}s \\
&\le \int_0^t \mathbb{E}[|\tilde{Y}_s^m|]\bigg(|\partial_2 K_\sigma(s,s)| + \int_s^t |\partial_{21}K_\sigma(s,u)|\,\mathrm{d}u\bigg)\,\mathrm{d}s.
\end{aligned} \tag{2.12}$$

For $I_{4,t}^{\delta\varepsilon}$, we get

$$\begin{aligned}
&\mathbb{E}[I_{4,t}^{\delta\varepsilon}] \\
&\quad = \mathbb{E}\left[\int_0^t \phi'_{\delta\varepsilon}(\tilde{X}_s^m)K_\sigma(s,s)\Big(\sigma\big(s, X^{m+1}(\kappa_{m+1}(s))\big) - \sigma\big(s, X^m(\kappa_m(s))\big)\Big)\,\mathrm{d}B_s\right] \\
&\quad = 0,
\end{aligned} \tag{2.13}$$

since $I_{4,t}^{\delta\varepsilon}$ is a martingale by [Pro92, p.73, Corollary 3], since $\mathbb{E}[\langle I_{4,t}^{\delta\varepsilon}\rangle_t] < \infty$ for all $t \in [0,T]$ due to the boundedness of K_σ (Assumption 2.1), the growth bound on σ and Proposition 2.16.

For $I^{\delta\varepsilon}_{5,t}$, using the boundedness of K_σ (Assumption 2.1), the Hölder continuity of σ (Assumption 2.2 (ii)) and the inequality (2.8), we get that

$$\begin{aligned}\mathbb{E}[I^{\delta\varepsilon}_{5,t}] &= \mathbb{E}\bigg[\frac{1}{2}\int_0^t \phi''_{\delta\varepsilon}(\tilde{X}^m_s)K_\sigma(s,s)^2\Big(\sigma\big(s,X^{m+1}(\kappa_{m+1}(s))\big)-\sigma\big(s,X^m(\kappa_m(s))\big)\Big)^2\,\mathrm{d}s\bigg]\\ &\le C\mathbb{E}\bigg[\int_0^t \phi''_{\delta\varepsilon}(\tilde{X}^m_s)\big(|\tilde{X}^m(s)|+U^m_s+U^{m+1}_s\big)^{1+2\xi}\,\mathrm{d}s\bigg]\\ &\le C\mathbb{E}\bigg[\int_0^t \mathbf{1}_{[\frac{\varepsilon}{\delta},\varepsilon]}(|\tilde{X}^m(s)|)\frac{\big(|\tilde{X}^m(s)|+U^m_s+U^{m+1}_s\big)^{1+2\xi}}{|\tilde{X}^m(s)|\ln(\delta)}\,\mathrm{d}s\bigg]\\ &\le C\bigg(\frac{\varepsilon^{2\xi}}{\ln(\delta)}+\frac{\delta}{\varepsilon\ln(\delta)}\mathbb{E}\bigg[\int_0^t\big(U^m_s+U^{m+1}_s\big)^{1+2\xi}\,\mathrm{d}s\bigg]\bigg).\end{aligned} \tag{2.14}$$

Moreover, by Corollary 2.17, we derive that

$$\mathbb{E}\bigg[\int_0^t\big(U^m_s+U^{m+1}_s\big)^{1+2\xi}\,\mathrm{d}s\bigg]\le C2^{-(1+2\xi)\beta m^5}$$

for any $\beta\in(0,\gamma)$ and, hence, we conclude

$$\mathbb{E}[I^{\delta\varepsilon}_{5,t}]\le C\bigg(\frac{\varepsilon^{2\xi}}{\ln(\delta)}+\frac{\delta}{\varepsilon\ln(\delta)}2^{-(1+2\xi)\beta m^5}\bigg). \tag{2.15}$$

Combining (2.9) with the five estimates (2.10), (2.11), (2.12), (2.13) and (2.15), we end up with

$$\begin{aligned}\mathbb{E}[|\tilde{X}^m_t|]\le C\bigg(&2^{-\beta m^5}+\frac{\varepsilon^{2\xi}}{\ln(\delta)}+\frac{\delta}{\varepsilon\ln(\delta)}2^{-(1+2\xi)\beta m^5}+\int_0^t\mathbb{E}[|\tilde{X}^m_s|]\,\mathrm{d}s\\ &+\int_0^t\mathbb{E}[|\tilde{Y}^m_s|]\bigg(|\partial_2K_\sigma(s,s)|+\int_s^t|\partial_{21}K_\sigma(s,u)|\,\mathrm{d}u\bigg)\,\mathrm{d}s\bigg).\end{aligned}$$

Therefore, choosing $\delta:=2^{\rho m^5}$ for $\rho\in(0,((1+2\xi)\beta)/2]$ and $\varepsilon:=2^{-\frac{(1+2\xi)\beta}{2}m^5}$, we get

$$\begin{aligned}\mathbb{E}[|\tilde{X}^m_t|]\le C\bigg(&C_m+\int_0^t\mathbb{E}[|\tilde{X}^m_s|]\,\mathrm{d}s\\ &+\int_0^t\mathbb{E}[|\tilde{Y}^m_s|]\bigg(|\partial_2K_\sigma(s,s)|+\int_s^t|\partial_{21}K_\sigma(s,u)|\,\mathrm{d}u\bigg)\,\mathrm{d}s\bigg),\end{aligned} \tag{2.16}$$

with

$$C_m:=2^{-\beta m^5}+m^{-5}2^{-(1+2\xi)\beta\xi m^5}+m^{-5}2^{-(\frac{(1+2\xi)\beta}{2}-\rho)m^5}. \tag{2.17}$$

To apply a Grönwall lemma, we set

$$M_m(t):=\sup_{s\in[0,t]}\Big(\mathbb{E}[|\tilde{X}^m_s|]+\mathbb{E}[|\tilde{Y}^m_s|]\Big),\quad t\in[0,T],$$

and derive in the following an inequality of the form $M_m(t) \leq C_m + \int_0^t f(t-s)M_m(s)\,\mathrm{d}s$ for a suitable function f.
To get a bound for $\mathbb{E}[|\tilde{Y}_t^m|]$, we first apply the integration by part formula to obtain

$$\begin{aligned}\tilde{X}_t^m &= \int_0^t K_\mu(s,t)\Big(\mu\big(s, X^{m+1}(\kappa_{m+1}(s))\big) - \mu\big(s, X_s^m(\kappa_m(s))\big)\Big)\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\,\mathrm{d}\tilde{Y}_s^m \\ &= \int_0^t K_\mu(s,t)\Big(\mu\big(s, X^{m+1}(\kappa_{m+1}(s))\big) - \mu\big(s, X_s^m(\kappa_m(s))\big)\Big)\,\mathrm{d}s \\ &\quad + K_\sigma(t,t)\tilde{Y}_t^m - \int_0^t \partial_1 K_\sigma(s,t)\tilde{Y}_s^m\,\mathrm{d}s,\end{aligned}$$

where we used that $K_\sigma(\cdot,t)$ is absolutely continuous for every $t \in [0,T]$. Since $K_\sigma(t,t) > C$ for some constant $C > 0$, K_μ is bounded (both by Assumption 2.1) and μ is Lipschitz continuous (Assumption 2.2), we get

$$\begin{aligned}\mathbb{E}[|\tilde{Y}_t^m|] \leq & C\mathbb{E}\Bigg[|\tilde{X}_t^m| + \int_0^t |K_\mu(s,t)|\Big|\mu\big(s, X^{m+1}(\kappa_{m+1}(s))\big) - \mu\big(s, X_s^m(\kappa_m(s))\big)\Big|\,\mathrm{d}s \\ & \quad + \int_0^t |\partial_1 K_\sigma(s,t)||\tilde{Y}_s^m|\,\mathrm{d}s\Bigg] \\ \leq & C\Bigg(\mathbb{E}[|\tilde{X}_t^m|] + \int_0^t \mathbb{E}\big[|\tilde{X}_s^m|\big]\,\mathrm{d}s + \mathbb{E}\Bigg[\int_0^t (U_s^m + U_s^{m+1})\,\mathrm{d}s\Bigg] \\ & \quad + \int_0^t |\partial_1 K_\sigma(s,t)|\mathbb{E}[|\tilde{Y}_s^m|]\,\mathrm{d}s\Bigg) \\ \leq & C\Bigg(2^{-\beta m^5} + \mathbb{E}[|\tilde{X}_t^m|] + \int_0^t \mathbb{E}\big[|\tilde{X}_s^m|\big]\,\mathrm{d}s + \int_0^t |\partial_1 K_\sigma(s,t)|\mathbb{E}[|\tilde{Y}_s^m|]\,\mathrm{d}s\Bigg),\end{aligned}$$

where we used Corollary 2.17 for the last estimate. Hence, by (2.16) we obtain

$$\begin{aligned}\mathbb{E}[|\tilde{Y}_t^m|] \leq C\Bigg(& C_m + \int_0^t \mathbb{E}[|\tilde{X}_s^m|]\,\mathrm{d}s \\ & + \int_0^t \mathbb{E}[|\tilde{Y}_s^m|]\Bigg(|\partial_1 K_\sigma(s,t)| + |\partial_2 K_\sigma(s,s)| + \int_s^t |\partial_{21} K_\sigma(s,u)|\,\mathrm{d}u\Bigg)\,\mathrm{d}s\Bigg).\end{aligned} \tag{2.18}$$

By the bound on the partial derivatives of K_σ made in Assumption 2.1, (2.16) and (2.18) can be further estimated to

$$\begin{aligned}\mathbb{E}[|\tilde{X}_t^m|] &\leq C\Bigg(C_m + \int_0^t \mathbb{E}[|\tilde{X}_s^m|]\,\mathrm{d}s + \int_0^t (t-s)^{-\alpha}\,\mathbb{E}[|\tilde{Y}_s^m|]\,\mathrm{d}s\Bigg), \\ \mathbb{E}[|\tilde{Y}_t^m|] &\leq C\Bigg(C_m + \int_0^t \mathbb{E}[|\tilde{X}_s^m|]\,\mathrm{d}s + \int_0^t (t-s)^{-\alpha}\,\mathbb{E}[|\tilde{Y}_s^m|]\,\mathrm{d}s\Bigg),\end{aligned}$$

for $\alpha \in [0, \frac{1}{2})$ as given in Assumption 2.1. Hence, we arrive at

$$\begin{aligned} M_m(t) &\leq \sup_{s\in[0,t]} \mathbb{E}[|\tilde{X}_t^m|] + \sup_{s\in[0,t]} \mathbb{E}[|\tilde{Y}_t^m|] \\ &\leq C\bigg(C_m + \int_0^t \big(1 + (t-s)^{-\alpha}\big) M_m(s)\,\mathrm{d}s\bigg). \end{aligned}$$

Note that Proposition 2.16 secures the integrability of M_m. An application of the Grönwall's lemma for weak singularities (see e.g. [Kru14, Lemma A.2]) reveals that $M_m(t) \leq CC_m$. The claimed summability of the sequence $(C_m)_{m\in\mathbb{N}}$ follows immediately by (2.17).
□

Remark 2.19. *The approximation $\phi_{\delta\varepsilon}$ of the absolute value, as used in the proof of Theorem 2.14, was introduced by Gyöngy and Rásonyi [GR11]. It is a modification of the approximation originally used by Yamada and Watanabe [YW71] and appears to be more involved. While the original approximation of Yamada and Watanabe is sufficient to prove pathwise uniqueness, as we will also see in Section 2.4, to prove the existence of a solution the approximation $\phi_{\delta\varepsilon}$ seems necessary. Indeed, one needs $\varepsilon \to 0$ to ensure that $\phi_{\delta\varepsilon} \to |\cdot|$ but the second parameter δ is essential to obtain the convergence of the Euler type approximation $(X^m)_{m\in\mathbb{N}}$ in the case $\xi = 0$ (i.e. σ is $1/2$-Hölder continuous), as one can see from* (2.16) *and* (2.17)*,*

With these preparation at hand we are ready to prove Theorem 2.14.

Proof of Theorem 2.14. Step 1: The sequence $(X^m)_{m\in\mathbb{N}}$ is a Cauchy sequence in $L^p(\Omega \times [0,T])$ for p given in the statement of Theorem 2.14.
By Fubini's theorem and Lemma 2.18, there exists a sequence $(C_m)_{m\in\mathbb{N}}$ such that

$$\mathbb{E}\bigg[\int_0^T \big|X^{m+1}(s) - X^m(s)\big|\,\mathrm{d}s\bigg] \leq C \sup_{s\in[0,T]} \mathbb{E}\left[|X^{m+1}(s) - X^m(s)|\right] \leq C_m$$

for $m \in \mathbb{N}$. Hence, using Hölder's inequality and the moment bound for $(X^m(t))_{t\in[0,T]}$ from Proposition 2.16, we get

$$\begin{aligned} &\mathbb{E}\bigg[\int_0^T |X^{m+1}(t) - X^m(t)|^p\,\mathrm{d}t\bigg] \\ &\quad\leq \mathbb{E}\bigg[\int_0^T |X^{m+1}(t) - X^m(t)|^{2p-1}\,\mathrm{d}t\bigg]^{\frac{1}{2}} \mathbb{E}\bigg[\int_0^T |X^{m+1}(t) - X^m(t)|\,\mathrm{d}t\bigg]^{\frac{1}{2}} \\ &\quad\leq 2^{p-1}\bigg(1 + \sup_{t\in[0,T]} |x_0(t)|^{2p-1}\bigg)^{\frac{1}{2}} C_m^{\frac{1}{2}}. \end{aligned}$$

Due to the summability property of $(C_m)_{m\in\mathbb{N}}$, the sequence $(X^m)_{m\in\mathbb{N}}$ is a Cauchy sequence in $L^p(\Omega\times[0,T])$. Hence, there exists a process $X = (X_t)_{t\in[0,T]} \in L^p(\Omega\times[0,T])$, such that

$$\lim_{m\to\infty} \mathbb{E}\left[\int_0^T |X^m(s) - X_s|^p\,\mathrm{d}s\right] = 0. \tag{2.19}$$

Step 2: $(X_t)_{t\in[0,T]}$ yields a strong solution to the SVE (2.2)
By construction, the processes $(X^m(t))_{t\in[0,T]}$ are $(\mathcal{F}_t)_{t\in[0,T]}$-progressively measurable on the given probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in[0,T]},\mathbb{P})$. Since (2.19) also shows the $L^p([0,t]\times\Omega)$-convergence of $(X_s^m)_{s\in[0,t]}$ to $(X_s)_{s\in[0,t]}$ for every $t\in[0,T]$, the completeness of the L^p spaces (see e.g. [Kle14, Theorem 7.3]) yields $\mathcal{B}([0,t])\otimes\mathcal{F}_t$-measurability of $(s,\omega)\mapsto X_s(\omega)$, $(s,\omega)\in[0,t]\times\Omega$ for every $t\in[0,T]$. Hence, the process $(X_t)_{t\in[0,T]}$ is also $(\mathcal{F}_t)_{t\in[0,T]}$-progressively measurable on $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in[0,T]},\mathbb{P})$. Moreover, by the growth conditions on μ and σ (see Assumption 2.2 (i)) and the integrability properties of K_μ and K_σ, we get that

$$\int_0^t (|K_\mu(s,t)\mu(s,X_s)|+|K_\sigma(s,t)\sigma(s,X_s)|^2)\,\mathrm{d}s<\infty \quad \text{for all } t\in[0,T].$$

It remains to show that the process $(X_t)_{t\in[0,T]}$ fulfills the SVE (2.2). To that end, we show that the two integrals in (2.6) preserve the $L^p(\Omega\times[0,T])$-convergence. For the Riemann–Stieltjes integral, we use the boundedness of K_μ, the Lipschitz continuity of μ, Hölder's inequality and Fubini's theorem to obtain

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T \Big|\int_0^t K_\mu(s,t)\,(\mu(s,X^m(\kappa_m(s)))-\mu(s,X_s))\,\mathrm{d}s\Big|^p\,\mathrm{d}t\right]\\
&\quad\le C\int_0^T\int_0^T \mathbb{E}\left[|X^m(\kappa_m(s))-X_s|^p\right]\mathrm{d}s\,\mathrm{d}t\\
&\quad\le C\bigg(\mathbb{E}\bigg[\int_0^T |X^m(\kappa_m(s))-X^m(s)|^p\,\mathrm{d}s\bigg]+\mathbb{E}\bigg[\int_0^T |X^m(s)-X_s|^p\,\mathrm{d}s\bigg]\bigg)\to 0,
\end{aligned}$$

as $m\to\infty$ by Corollary 2.17 and (2.19). For the stochastic integral, we use Fubini's theorem, Burkholder–Davis–Gundy's inequality, Hölder's inequality, the boundedness of K_σ, and the Hölder regularity of σ to get that

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T \left|\int_0^t K_\sigma(s,t)\,(\sigma(s,X^m(\kappa_m(s)))-\sigma(s,X_s))\,\mathrm{d}B_s\right|^p\,\mathrm{d}t\right]\\
&\quad=\int_0^T \mathbb{E}\bigg[\Big|\int_0^t K_\sigma(s,t)\,(\sigma(s,X^m(\kappa_m(s)))-\sigma(s,X_s))\,\mathrm{d}B_s\Big|^p\bigg]\,\mathrm{d}t\\
&\quad\le\int_0^T \mathbb{E}\bigg[\int_0^t K_\sigma(s,t)^2\,(\sigma(s,X^m(\kappa_m(s)))-\sigma(s,X_s))^2\,\mathrm{d}s\bigg]^{\frac{p}{2}}\,\mathrm{d}t\\
&\quad\le C\left(\int_0^T\int_0^T \mathbb{E}[|X^m(\kappa_m(s))-X_s|^{\frac{p}{2}+p\xi}]\,\mathrm{d}s\,\mathrm{d}t\right)\\
&\quad\le C\bigg(\mathbb{E}\bigg[\int_0^T |X^m(\kappa_m(s))-X^m(s)|^{\frac{p}{2}+p\xi}\,\mathrm{d}s\bigg]+\mathbb{E}\bigg[\int_0^T |X^m(s)-X_s|^{\frac{p}{2}+p\xi}\,\mathrm{d}s\bigg]\bigg).
\end{aligned}$$

Thus, by Corollary 2.17 and the convergence $X^m\to X$ in $L^{\frac{p}{2}+p\xi}(\Omega\times[0,T])$ as $m\to\infty$, for $\xi\in[0,\frac{1}{2}]$, which is implied by the one in $L^p(\Omega\times[0,T])$, we see that the stochastic integral does preserve the $L^p(\Omega\times[0,T])$-convergence. Thus, we have proven that the limiting process $(X_t)_{t\in[0,T]}$ fulfills the SVE (2.2) for almost all $(t,\omega)\in[0,T]\times\Omega$. By Remark 2.8,

$(X_t)_{t\in[0,T]}$ has an $\mathbb{P}$-a.s. continuous version, which fulfills the SVE (2.2) for all $t \in [0,T]$ for almost all $\omega \in \Omega$, and hence, is a strong solution of (2.2). $\square$

2.4 Pathwise uniqueness

In this section we establish the pathwise uniqueness for the stochastic Volterra equation (2.2) under Assumptions 2.1, 2.2 (i), and under slightly weaker regularity assumptions on μ and σ than Assumption 2.2 (ii), namely an Osgood-type condition on μ and the Yamada–Watanabe condition on σ, as formulated in the next assumption.

Assumption 2.20. *Let $\mu, \sigma \colon [0,T] \times \mathbb{R} \to \mathbb{R}$ be measurable functions such that:*

(i) there is some continuous, non-decreasing and concave function $\kappa \colon [0,\infty) \to [0,\infty)$ with $\kappa(0) = 0$ and $\kappa(x) > 0$ for $x > 0$, such that, with the notation $\tilde{\kappa}(x) := \kappa(x) + |x|$,

$$\int_0^\varepsilon \frac{\mathrm{d}x}{\left(\tilde{\kappa}(\sqrt[q]{x})\right)^q} = \infty,$$

holds for all $\varepsilon > 0$ and $q \in (\frac{1}{1-\alpha}, \frac{1}{1-\alpha} + \tilde{\varepsilon})$ for some $\tilde{\varepsilon} > 0$, where $\alpha \in [0, \frac{1}{2})$ is given by Assumption 2.1 (ii), and

$$|\mu(t,x) - \mu(t,y)| \leq \kappa(|x-y|),$$

for all $t \in [0,T]$, $x, y \in \mathbb{R}$,

(ii) there is some continuous strictly increasing function $\rho \colon [0,\infty) \to [0,\infty)$ with $\rho(0) = 0$ and $\rho(x) > 0$ for $x > 0$, such that

$$\int_0^\varepsilon \frac{\mathrm{d}x}{\rho(x)^2} = \infty,$$

holds for all $\varepsilon > 0$, and

$$|\sigma(t,x) - \sigma(t,y)| \leq \rho(|x-y|),$$

for all $t \in [0,T]$, $x, y \in \mathbb{R}$.

Remark 2.21. *Choosing $\kappa(x) = C_\mu |x|$ and $\rho(x) = C_\sigma |x|^{\frac{1}{2}+\xi}$ shows that Assumption 2.2 (ii) implies Assumption 2.20. We note that if μ is assumed to be Lipschitz continuous and σ to fulfill the Yamada–Watanabe condition, it is sufficient to use a fractional Grönwall lemma like the one in [Kru14, Lemma A.2] instead of the fractional Bihari inequality in* (2.31). *Moreover, if one considers $K_\sigma = 1$, the Osgood-type condition in Assumption 2.20 (i) can be replaced by the classical Osgood condition for SDEs (see e.g. [KS91, Chapter 5, Remark 2.16]) since one can then use the classical instead of the fractional Bihari inequality and the application of integration by parts to the stochastic integral is not required.*

The main result of this section reads as follows.

Theorem 2.22. *Suppose Assumptions 2.1, 2.2 (i) and 2.20. Then, pathwise uniqueness holds for the stochastic Volterra equation* (2.2)*.*

Proof. Since the proof relies partly on similar techniques as the proof of Lemma 2.18, we try to give a condense presentation and refer to the analogue calculation in Section 2.3. Let $(X_t^1)_{t\in[0,T]}$ and $(X_t^2)_{t\in[0,T]}$ be solutions to the SVE (2.2). Analogously to Section 2.3, we define $Y_t^i := \int_0^t \sigma(s, X_s^i)\,\mathrm{d}B_s$ and $H_t^i := \int_0^t \partial_2 K_\sigma(s,t)\,\mathrm{d}Y_s^i$, for $i = 1, 2$, as well as $\tilde{Y}_t := Y_t^1 - Y_t^2$, $\tilde{X}_t := X_t^1 - X_t^2$, $\tilde{H}_t := H_t^1 - H_t^2$, and $\tilde{Z}_t := \int_0^t \big(\mu(s, X_s^1) - \mu(s, X_s^2)\big)\,\mathrm{d}s$, for $t \in [0,T]$. By Lemma 2.12, we obtain the semimartingale decomposition

$$\begin{aligned}\tilde{X}_t = &\int_0^t K_\mu(s,s)(\mu(s,X_s^1) - \mu(s,X_s^2))\,\mathrm{d}s + \int_0^t \int_0^s \partial_2 K_\mu(u,s)\,\mathrm{d}\tilde{Z}_u\,\mathrm{d}s \\ &+ \int_0^t \tilde{H}_s\,\mathrm{d}s + \int_0^t K_\sigma(s,s)\,\mathrm{d}\tilde{Y}_s, \quad t \in [0,T]. \end{aligned} \tag{2.20}$$

To construct an approximation of the absolute value by smooth functions allowing us to apply Itô's formula, we use the classical approximation of Yamada–Watanabe [YW71] for simplicity, cf. Remark 2.19. Based on the strictly increasing function ρ from Assumption 2.20 (ii), we define a sequence $(\phi_n)_{n\in\mathbb{N}}$ of functions mapping from $\mathbb{R}$ to $\mathbb{R}$ that approximates the absolute value in the following way: Let $(a_n)_{n\in\mathbb{N}}$ be a strictly decreasing sequence with $a_0 = 1$ such that $a_n \to 0$ as $n \to \infty$ and

$$\int_{a_n}^{a_{n-1}} \frac{1}{\rho(x)^2}\,\mathrm{d}x = n.$$

Furthermore, we define a sequence of mollifiers: let $(\psi_n)_{n\in\mathbb{N}} \in C_0^\infty(\mathbb{R})$ be smooth functions with compact support such that $\operatorname{supp}(\psi_n) \subset (a_n, a_{n-1})$, and with the properties

$$0 \le \psi_n(x) \le \frac{2}{n\rho(x)^2}, \quad \forall x \in \mathbb{R}, \quad \text{and} \quad \int_{a_n}^{a_{n-1}} \psi_n(x)\,\mathrm{d}x = 1. \tag{2.21}$$

We set

$$\phi_n(x) := \int_0^{|x|} \left(\int_0^y \psi_n(z)\,\mathrm{d}z \right) \mathrm{d}y, \quad x \in \mathbb{R}.$$

By (2.21) and the compact support of ψ_n, it follows that $\phi_n(\cdot) \to |\cdot|$ uniformly as $n \to \infty$. Since every ψ_n and, thus, every ϕ_n is zero in a neighborhood around zero, the functions ϕ_n are smooth with

$$\|\phi_n'\|_\infty \le 1, \quad \phi_n'(x) = \operatorname{sgn}(x) \int_0^{|x|} \psi_n(y)\,\mathrm{d}y, \quad \text{and} \quad \phi_n''(x) = \psi_n(|x|) \quad \text{for } x \in \mathbb{R}.$$

Since the quadratic variation of the semimartingale $(\tilde{X}_t)_{t\in[0,T]}$ is given by

$$\langle \tilde{X} \rangle_t = \int_0^t K_\sigma(s,s)^2 \big(\sigma(s,X_s^1) - \sigma(s,X_s^2)\big)^2\,\mathrm{d}s, \quad t \in [0,T],$$

we get, by applying Itô's formula and using the semimartingale decomposition (2.20), that

$$
\begin{aligned}
\phi_n(\tilde{X}_t) &= \int_0^t \phi_n'(\tilde{X}_s)\,\mathrm{d}\tilde{X}_s + \frac{1}{2}\int_0^t \phi_n''(\tilde{X}_s)\,\mathrm{d}\langle \tilde{X}\rangle_s \\
&= \int_0^t \phi_n'(\tilde{X}_s)K_\mu(s,s)(\mu(s,X_s^1)-\mu(s,X_s^2))\,\mathrm{d}s + \int_0^t \phi_n'(\tilde{X}_s)\bigg(\int_0^s \partial_2 K_\mu(u,s)\,\mathrm{d}\tilde{Z}_u\bigg)\,\mathrm{d}s \\
&\quad + \int_0^t \phi_n'(\tilde{X}_s)\tilde{H}_s\,\mathrm{d}s + \int_0^t \phi_n'(\tilde{X}_s)K_\sigma(s,s)\,\mathrm{d}\tilde{Y}_s \\
&\quad + \frac{1}{2}\int_0^t \phi_n''(\tilde{X}_s)K_\sigma(s,s)^2\big(\sigma(s,X_s^1)-\sigma(s,X_s^2)\big)^2\,\mathrm{d}s \\
&=: I_{1,t}^n + I_{2,t}^n + I_{3,t}^n + I_{4,t}^n + I_{5,t}^n
\end{aligned}
\tag{2.22}
$$

for $t \in [0,T]$.
For $I_{1,t}^n$, we use Assumption 2.20 (i), the boundedness of K_μ (Assumption 2.1), the bound $\|\phi_n'\|_\infty \leq 1$ and Jensen's inequality to estimate

$$
\mathbb{E}[I_{1,t}^n] \leq C\int_0^t \mathbb{E}[\kappa(|\tilde{X}_s|)]\,\mathrm{d}s \leq C\int_0^t \kappa(\mathbb{E}[|\tilde{X}_s|])\,\mathrm{d}s. \tag{2.23}
$$

For $I_{2,t}^n$, we additionally use the boundedness of $\partial_2 K_\mu(u,s)$ on Δ_T to obtain

$$
\mathbb{E}[I_{2,t}^n] \leq C\int_0^t \kappa(\mathbb{E}[|\tilde{X}_s|])\,\mathrm{d}s. \tag{2.24}
$$

For $I_{3,t}^n$, similarly to (2.12), we use the integration by parts formula to estimate

$$
\begin{aligned}
\mathbb{E}[I_{3,t}^n] &\leq \int_0^t \mathbb{E}[|\tilde{H}_s|]\,\mathrm{d}s \\
&\leq \int_0^t |\partial_2 K_\sigma(s,s)|\mathbb{E}[|\tilde{Y}_s|]\,\mathrm{d}s + \int_0^t\int_0^s |\partial_{21}K_\sigma(u,s)|\mathbb{E}[|\tilde{Y}_u|]\,\mathrm{d}u\,\mathrm{d}s \\
&\leq \int_0^t \mathbb{E}[|\tilde{Y}_s|]\bigg(\partial_2 K_\sigma(s,s) + \int_s^t |\partial_{21}K_\sigma(s,u)|\,\mathrm{d}u\bigg)\,\mathrm{d}s.
\end{aligned}
\tag{2.25}
$$

For $I_{4,t}^n$, since $I_{4,t}^n$ is a martingale by [Pro92, p.73, Corollary 3] due to the boundedness of K_σ, the growth bound on σ and Lemma 2.10, we get

$$
\mathbb{E}[I_{4,t}^n] = \mathbb{E}\left[\int_0^t \phi_n'(\tilde{X}_s)K_\sigma(s,s)(\sigma(s,X_s^1)-\sigma(s,X_s^2))\,\mathrm{d}B_s\right] = 0, \tag{2.26}
$$

For $I_{5,t}^n$, we get by using the boundedness of K_σ (Assumption 2.1), the regularity of σ

from Assumption 2.20 (ii), and the inequality (2.21) that

$$\begin{aligned}\mathbb{E}[I^n_{5,t}] &\leq C\mathbb{E}\bigg[\int_0^t \phi_n''(\tilde{X}_s)\rho(|\tilde{X}_s|)^2\,\mathrm{d}s\bigg]\\ &\leq C\mathbb{E}\bigg[\int_0^t \frac{2}{n\rho(|\tilde{X}_s|)^2}\rho(|\tilde{X}_s|)^2\,\mathrm{d}s\bigg]\\ &\leq \frac{C}{n},\end{aligned}\tag{2.27}$$

for some $C > 0$.
Finally, sending $n \to \infty$ and combining the five previous estimates (2.23), (2.24), (2.25), (2.26) and (2.27) with (2.22) implies

$$\mathbb{E}[|\tilde{X}_t|] \leq C\int_0^t \kappa(\mathbb{E}[|\tilde{X}_s|])\,\mathrm{d}s + \int_0^t \mathbb{E}[|\tilde{Y}_s|]\bigg(\partial_2 K_\sigma(s,s) + \int_s^t |\partial_{21}K_\sigma(s,u)|\,\mathrm{d}u\bigg)\,\mathrm{d}s. \tag{2.28}$$

To apply a Grönwall lemma, we set

$$M(t) := \sup_{s\in[0,t]}\Big(\mathbb{E}[|\tilde{X}_s|] + \mathbb{E}[|\tilde{Y}_s|]\Big), \quad t \in [0,T],$$

and derive in the following an inequality of the form $M(t) \leq \int_0^t f(t-s)\tilde{\kappa}(M(s))\,\mathrm{d}s$ for suitable functions f and $\tilde{\kappa}$. To find a bound for $\mathbb{E}[|\tilde{Y}_t|]$, we apply the integration by part formula to obtain

$$\begin{aligned}\tilde{X}_t &= \int_0^t K_\mu(s,t)(\mu(s,X_s^1) - \mu(s,X_s^2))\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\,\mathrm{d}\tilde{Y}_s\\ &= \int_0^t K_\mu(s,t)(\mu(s,X_s^1) - \mu(s,X_s^2))\,\mathrm{d}s + K_\sigma(t,t)\tilde{Y}_t - \int_0^t \partial_1 K_\sigma(s,t)\tilde{Y}_s\,\mathrm{d}s\end{aligned}\tag{2.29}$$

keeping in mind that that $K_\sigma(\cdot,t)$ is absolutely continuous for every $t \in [0,T]$. Due to $|K_\sigma(t,t)| > C$ for some constant $C > 0$, we can rearrange (2.29) and use (2.28) to get

$$\begin{aligned}\mathbb{E}\Big[|\tilde{Y}_t|\Big] \leq & C\bigg(\int_0^t \mathbb{E}\big[|\mu(s,X_s^1) - \mu(s,X_s^2)|\big]\,\mathrm{d}s\\ &+ \mathbb{E}\big[|\tilde{X}_t|\big] + \int_0^t |\partial_1 K_\sigma(s,t)|\mathbb{E}\big[|\tilde{Y}_s|\big]\,\mathrm{d}s\bigg)\\ \leq & C\bigg(\int_0^t \Big(\mathbb{E}\big[|\tilde{X}_s|\big] + \kappa(\mathbb{E}\big[|\tilde{X}_s|\big])\Big)\,\mathrm{d}s\\ &+ \int_0^t \mathbb{E}\big[|\tilde{Y}_s|\big]\bigg(|\partial_1 K_\sigma(s,t)| + |\partial_2 K_\sigma(s,s)| + \int_s^t |\partial_{21}K_\sigma(s,u)|\,\mathrm{d}u\bigg)\,\mathrm{d}s\bigg).\end{aligned}\tag{2.30}$$

Using Assumption 2.1 to bound the partial derivative terms in (2.28) and (2.30), we end up with

$$\begin{aligned} M(t) &\leq \sup_{s\in[0,t]} \mathbb{E}[|\tilde{X}_t|] + \sup_{s\in[0,t]} \mathbb{E}[|\tilde{Y}_t|] \\ &\leq C\bigg(\int_0^t \Big(\sup_{u\in[0,s]} \mathbb{E}[|\tilde{X}_u|] + \kappa\Big(\sup_{u\in[0,s]} \mathbb{E}[|\tilde{X}_u|]\Big)\Big) \,\mathrm{d}s + \int_0^t (t-s)^{-\alpha} \sup_{u\in[0,s]} \mathbb{E}[|\tilde{Y}_u|] \,\mathrm{d}s \bigg) \\ &\leq C \int_0^t (t-s)^{-\alpha} \tilde{\kappa}(M(s)) \,\mathrm{d}s, \end{aligned} \tag{2.31}$$

where $\tilde{\kappa}(x) := \kappa(x) + |x|$. An application of the fractional Bihari inequality, [OHNO21, Theorem 2.3], with sending $q \to \frac{1}{1-\alpha}$ like in [OHNO21, proof of Theorem 3.1, Step 1] with the condition on $\tilde{\kappa}$ in Assumption 2.20 (i) that $M(t) = 0$ holds. Hence, $\tilde{X}_t = 0$ almost surely, and, thus, by the continuity of the solutions, the processes $(X_t^1)_{t\in[0,T]}$ and $(X_t^2)_{t\in[0,T]}$ are indistinguishable. □

Chapter 3

Weak existence of solutions

The content of this chapter is published in [PS23c].

Introduction

We investigate the existence of weak solutions to stochastic Volterra equation (SVEs)

$$X_t = x_0(t) + \int_0^t K_\mu(s,t)\mu(s,X_s)\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s,X_s)\,\mathrm{d}B_s, \quad t \in [0,T], \tag{3.1}$$

where x_0 is a continuous function, B is a Brownian motion, and the kernels K_μ, K_σ are measurable functions. The time-inhomogeneous coefficients μ, σ are only supposed to be continuous in space uniformly in time. In case of ordinary stochastic differential equations (SDEs), i.e. $K_\sigma = K_\mu = 1$, the existence of weak solutions was first proven by Skorokhod [Sko61] and can, nowadays, be found in different generality in standard textbooks like [SV79, KS91].

A comprehensive study of weak solutions to stochastic Volterra equations which additionally allow for jumps was recently initiated by Abi Jaber, Cuchiero, Larsson and Pulido [AJCLP21], see also [MS15]. The extension of the theory of weak solutions from ordinary stochastic differential equations to SVEs constitutes a natural generalization of the classical theory and is motivated by successful applications of SVEs with non-Lipschitz coefficients as volatility models in mathematical finance, see e.g. [EER19, AJEE19b]. Assuming that the kernels in the SVE (3.1) are of convolution type, i.e. $K_\mu(s,t) = K_\sigma(s,t) = K(t-s)$ for some function $K\colon \mathbb{R} \to \mathbb{R}$, and that the coefficients μ, σ are continuous jointly in space-time, the existence of weak solutions was derived in [AJCLP21], see also [MS15, AJLP19, AJ21]. To that end, Abi Jaber et al. [AJCLP21] introduces a local martingale problem associated to SVEs of convolutional type.

In this chapter, we establish a local martingale problem associated to general stochastic Volterra equations, see Definition 3.4, and show that its solvability is equivalent to the existence of a weak solution to the associated SVE, see Lemma 3.7. Using this newly formulated Volterra local martingale problem, we obtain the existence of weak solutions to stochastic Volterra equations with time-inhomogeneous coefficients, that are not necessarily continuous in t, and allowing for general kernels in the drift and convolutional kernels as well as bounded general kernels in the diffusion term, see Theorem 3.10. The presented approach can be considered, roughly speaking, as a generalization of Skorokhod's original construction to the more general case of SVEs, and is developed in a one-dimensional setting to keep the presentation fairly short without cumbersome notation. However, as for ordinary SDEs and for SVEs of convolutional type, all concepts and results are expected to extend to a multi-dimensional setting in a straightforward manner.

Organization of the chapter: In Section 3.1 we introduce a local martingale problem associated to SVEs. The existence of weak solutions to SVEs is provided in Section 3.2.

3.1 Weak solutions and the Volterra local martingale problem

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space, which satisfies the usual conditions. For $T \in (0, \infty)$ we consider the one-dimensional stochastic Volterra equation

$$X_t = x_0(t) + \int_0^t K_\mu(s,t)\mu(s, X_s)\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s, X_s)\,\mathrm{d}B_s, \quad t \in [0,T], \tag{3.2}$$

where $x_0 \colon [0,T] \to \mathbb{R}$ is a continuous function, $(B_t)_{t\in[0,T]}$ is a Brownian motion, and the coefficients $\mu, \sigma \colon [0,T] \times \mathbb{R} \to \mathbb{R}$ and the kernels $K_\mu, K_\sigma \colon \Delta_T \to \mathbb{R}$ are measurable functions, using the notation $\Delta_T := \{(s,t) \in [0,T] \times [0,T] \colon\ 0 \le s \le t \le T\}$. The integral $\int_0^t K_\mu(s,t)\mu(s,X_s)\,\mathrm{d}s$ is defined as a Lebesque integral and $\int_0^t K_\sigma(s,t)\sigma(s,X_s)\,\mathrm{d}B_s$ as an Itô integral. Moreover, for $p \in [1,\infty)$ we write $L^p(\Omega \times [0,T])$ and $L^p([0,T])$ for the space of p-integrable functions on $\Omega \times [0,T]$ and on $[0,T]$, respectively.

Analogous to the notion of weak solutions to ordinary stochastic differential equations (see e.g. [KS91, Chapter 5.3, Definition 3.1], we make the following definition.

Definition 3.1. *A* weak solution *to* (3.2) *is a triple* (X, B), $(\Omega, \mathcal{F}, \mathbb{P})$, $(\mathcal{F}_t)_{t\in[0,T]}$ *such that*

(i) $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space,* $(\mathcal{F}_t)_{t\in[0,T]}$ *is a filtration of sub-σ-algebras of $\mathcal{F}$ satisfying the usual conditions,*

(ii) $X = (X_t)_{t\in[0,T]} \in L^1(\Omega \times [0,T])$ *is an $(\mathcal{F}_t)$-progressively measurable process,* $B = (B_t)_{t\in[0,T]}$ *is a Brownian motion w.r.t.* $(\mathcal{F}_t)_{t\in[0,T]}$*,*

(iii) $\int_0^t \left(|K_\mu(s,t)\mu(s,X_s)| + |K_\sigma(s,t)\sigma(s,X_s)|^2\right)\mathrm{d}s < \infty$ $\mathbb{P}$*-a.s. for any* $t \in [0,T]$*, and*

(iv) (3.2) *holds for* (X, B) *on* $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{P}$*-a.s.*

Under suitable assumptions on the coefficients and kernels, the existence of weak solutions to the stochastic Volterra equation (3.2) can be equivalently formulated in terms of solutions to an associated local martingale problem, see Definition 3.4 below. To that end, we make the following assumption.

Assumption 3.2. *Let $K_\mu, K_\sigma\colon \Delta_T \to \mathbb{R}$ be measurable functions with $K_\mu(\cdot,t) \in L^1([0,T])$ and $K_\sigma(\cdot,t) \in L^2([0,T])$ for every $t \in [0,T]$, and let $\mu,\sigma\colon [0,T]\times\mathbb{R}\to\mathbb{R}$ be measurable functions fulfilling the linear growth condition*

$$|\mu(t,x)| + |\sigma(t,x)| \leq C_{\mu,\sigma}(1+|x|), \quad t\in[0,T],\, x\in\mathbb{R},$$

for some constant $C_{\mu,\sigma} > 0$.

Let $C^2(\mathbb{R})$ be the space of twice continuously differentiable functions $f\colon \mathbb{R}\to\mathbb{R}$ and $C_0^2(\mathbb{R})$ be the space of all $f\in C^2(\mathbb{R})$ with compact support. For two stochastic processes $X = (X_t)_{t\in[0,T]}$ and $Z = (Z_t)_{t\in[0,T]}$ such that $X \in L^1(\Omega\times[0,T])$ is $(\mathcal{F}_t)$-progressively measurable and Z is $(\mathcal{F}_t)$-adapted and continuous, we introduce the process $(\mathcal{M}_t^f)_{t\in[0,T]}$ by

$$\mathcal{M}_t^f := f(Z_t) - \int_0^t \mathcal{A}^f(s,X_s,Z_s)\,\mathrm{d}s, \quad t\in[0,T], \tag{3.3}$$

for $f\in C^2(\mathbb{R})$, where

$$\mathcal{A}^f\colon [0,T]\times\mathbb{R}\times\mathbb{R}\to\mathbb{R} \quad \text{with} \quad \mathcal{A}^f(t,x,z) := \mu(t,x)f'(z) + \frac{1}{2}\sigma(t,x)^2 f''(z). \tag{3.4}$$

As we shall see in the next proposition, assuming that $(\mathcal{M}_t^f)_{t\in[0,T]}$ is a local martingale for all $f\in C_0^2(\mathbb{R})$ implies that the stochastic process Z is a semimartingale.

Proposition 3.3. *Suppose Assumption 3.2. Let $(X_t)_{t\in[0,T]}$ be an $(\mathcal{F}_t)$-progressively measurable process in $L^1(\Omega\times[0,T])$ and $(Z_t)_{t\in[0,T]}$ be an $(\mathcal{F}_t)$-adapted and continuous process on a filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in[0,T]},\mathbb{P})$ satisfying the usual conditions. If $(\mathcal{M}_t^f)_{t\in[0,T]}$ is a local martingale for every $f\in C_0^2(\mathbb{R})$, then we have:*

(i) $(Z_t)_{t\in[0,T]}$ is a semimartingale with characteristics $\big(\int_0^\cdot \mu(s,X_s)\,\mathrm{d}s, \int_0^\cdot \sigma(s,X_s)^2\,\mathrm{d}s, 0\big)$.

(ii) There exists a filtered probability space $(\tilde\Omega,\tilde{\mathcal{F}},(\tilde{\mathcal{F}}_t)_{t\in[0,T]},\tilde{\mathbb{P}})$ satisfying the usual conditions such that $(Z_t)_{t\in[0,T]}$ is a semimartingale on $(\tilde\Omega,\tilde{\mathcal{F}},(\tilde{\mathcal{F}}_t)_{t\in[0,T]},\tilde{\mathbb{P}})$ and

$$Z_t = \int_0^t \mu(s,X_s)\,\mathrm{d}s + \int_0^t \sigma(s,X_s)\,\mathrm{d}B_s, \quad t\in[0,T],$$

holds $\tilde{\mathbb{P}}$-a.s., for some Brownian motion $(B_t)_{t\in[0,T]}$ on $(\tilde\Omega,\tilde{\mathcal{F}},(\tilde{\mathcal{F}}_t)_{t\in[0,T]},\tilde{\mathbb{P}})$.

Proof. (i) By [JS03, Theorem II.2.42], in order to prove the assertion, it is sufficient to show that $(\mathcal{M}_t^f)_{t\in[0,T]}$, defined in (3.3), is a local martingale for every bounded function $f\in C^2(\mathbb{R})$.
Let $f\in C^2(\mathbb{R})$ be bounded and define the hitting times

$$\tau_n := \inf_{t\in[0,T]}\{\max(|X_t|,|Z_t|)\geq n\}, \quad n\in\mathbb{N}.$$

Note that $\tau_n \to T$ a.s. as $n \to \infty$ since $X \in L^1(\Omega \times [0,T])$ and Z is continuous. Since the underlying filtered probability space satisfies the usual conditions, by the Début theorem (see [RY99, Chapter I, (4.15) Theorem]), the hitting times $(\tau_n)_{n\in\mathbb{N}}$ are stopping times. It remains to show that $(\tau_n)_{n\in\mathbb{N}}$ is a localizing sequence for $(\mathcal{M}_t^f)_{t\in[0,T]}$. To that end, we approximate f by the functions $(f_n)_{n\in\mathbb{N}} \subset C_0^2(\mathbb{R})$ given by $f_n := \phi_n f$ for some $\phi_n \in C_0^2(\mathbb{R})$ taking values in $[0,1]$ and being identical to 1 on $[-n,n]$. Hence, $(\mathcal{M}_t^{f_n})_{t\in[0,T]}$ is a local martingale for every $n \in \mathbb{N}$ and, thus, the stopped process $(\mathcal{M}_{t\wedge\tau_n}^{f_n})_{t\in[0,T]}$, given by

$$\mathcal{M}_{t\wedge\tau_n}^{f_n} = (f_n)(Z_{t\wedge\tau_n}) - \int_0^{t\wedge\tau_n} \mathcal{A}^{f_n}(s, X_s, Z_s)\,\mathrm{d}s, \quad t \in [0,T],$$

is a martingale as

$$|\mathcal{M}_{t\wedge\tau_n}^{f_n}| \leq \sup_{x\in\mathbb{R}} |f(x)| + C_{\sigma,\mu,n} n^2 < \infty,$$

for some constant $C_{\sigma,\mu,n} > 0$, using the definition of τ_n and the linear growth condition on μ and σ. Since $\mathcal{M}_{t\wedge\tau_n}^{f_n} = \mathcal{M}_{t\wedge\tau_n}^{f}$ for $t \in [0,T]$, $(\mathcal{M}_{t\wedge\tau_n}^{f})_{t\in[0,T]}$ is a martingale for every $n \in \mathbb{N}$ and, hence, $(\tau_n)_{n\in\mathbb{N}}$ a localizing sequence for $(\mathcal{M}_t^f)_{t\in[0,T]}$.
(ii) Since the process $(Z_t)_{t\in[0,T]}$ is a semimartingale with absolutely continuous characteristics $\big(\int_0^\cdot \mu(s,X_s)\,\mathrm{d}s, \int_0^\cdot \sigma^2(s,X_s)\,\mathrm{d}s, 0\big)$, the assertion follow by [JP12, Theorem 2.1.2]. □

Keeping these preliminary considerations and the classical martingale problem (see e.g. [KS14, Definition 7.1.1]) in mind, we formulate a local martingale problem associated to the stochastic Volterra equation (3.2).

Definition 3.4. *A solution to the Volterra local martingale problem given $(x_0, \mu, \sigma, K_\mu, K_\sigma)$ is a triple (X, Z), $(\Omega, \mathcal{F}, \mathbb{P})$, $(\mathcal{F}_t)_{t\in[0,T]}$ such that*

(i) $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $(\mathcal{F}_t)_{t\in[0,T]}$ is a filtration of sub-σ-algebras of $\mathcal{F}$ satisfying the usual conditions,

(ii) $X = (X_t)_{t\in[0,T]} \in L^1(\Omega \times [0,T])$ is an $(\mathcal{F}_t)$-progressively measurable process,

(iii) $(Z_t)_{t\in[0,T]}$ is a continuous semimartingale with $Z_0 = 0$ and decomposition $Z = A + M$ for some process $(A_t)_{t\in[0,T]}$ of bounded variation and some local martingale $(M_t)_{t\in[0,T]}$,

(iv) the process $(\mathcal{M}_t^f)_{t\in[0,T]}$, given by

$$\mathcal{M}_t^f := f(Z_t) - \int_0^t \mathcal{A}^f(s, X_s, Z_s)\,\mathrm{d}s, \quad t \in [0,T], \tag{3.5}$$

is a local martingale for every $f \in C_0^2(\mathbb{R})$, where $\mathcal{A}^f$ is defined as in (3.4), and

(v) the following equality holds:

$$X_t = x_0(t) + \int_0^t K_\mu(s,t)\,\mathrm{d}A_s + \int_0^t K_\sigma(s,t)\,\mathrm{d}M_s, \quad t \in [0,T], \quad \mathbb{P}\text{-a.s.} \tag{3.6}$$

Remark 3.5. *The first Volterra local martingale problem was formulated in [AJCLP21] for stochastic Volterra equations of convolution type, that is, the kernels K_μ, K_σ are supposed to be of the form $K(t-s)$ for a deterministic function $K\colon [0,T] \to \mathbb{R}$, see [AJCLP21, Definition 3.1]. However, [AJCLP21, Definition 3.1] fundamentally relies on the convolutional structure to ensure that a weak solution to the SVE leads to a solution of the Volterra local martingale problem. The latter conclusion is based on a substitution and stochastic Fubini argument, which is not applicable for general kernels. Compared to [AJCLP21, Definition 3.1], the essential difference is that we reformulated [AJCLP21, (3.3)] to the condition* (3.6). *While both conditions are equivalent for kernels of convolutional type, the advantage of* (3.6) *is that it allows for general kernels.*
Moreover, notice that the Volterra local martingale problem as presented in Definition 3.4 reduces to the local martingale problem for ordinary stochastic differential equations in the case $K_\mu = K_\sigma = 1$. Indeed, in this case conditions (i) and (iv) imply conditions (iii) and (v) on a possibly extended probability space, see Proposition 3.3.

Remark 3.6. *Condition (iii) of Definition 3.4 can be relaxed to the condition "$(Z_t)_{t\in[0,T]}$ is an $(\mathcal{F}_t)$-adapted and continuous process" since this together with (iv) of Definition 3.4 already implies the semimartingale property of $(Z_t)_{t\in[0,T]}$, see Proposition 3.3. However, we decided to directly postulate the semimartingale property of $(Z_t)_{t\in[0,T]}$ in the formulation of the Volterra local martingale problem to ensure that condition (v) is obviously well-defined.*

As for ordinary stochastic differential equations, the existence of weak solutions to SVEs is equivalent to the solvability of the associated Volterra local martingale problem.

Lemma 3.7. *Suppose Assumption 3.2. There exists a weak solution to the SVE* (3.2) *if and only if there exists a solution to the Volterra local martingale problem given* $(x_0, \mu, \sigma, K_\mu, K_\sigma)$.

Proof. Let (X, B) be a (weak) solution to (3.2) on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Setting

$$Z_t := A_t + M_t := \int_0^t \mu(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s, \quad t \in [0,T],$$

Itô's formula applied to $f(Z_t)$ for $f \in C_0^2(\mathbb{R})$ yields that

$$\begin{aligned}\mathcal{M}_t^f &= f(Z_t) - \int_0^t f'(Z_s)\mu(s, X_s)\,\mathrm{d}s - \frac{1}{2}\int_0^t f''(Z_s)\sigma(s, X_s)^2\,\mathrm{d}s \\ &= f(Z_0) + \int_0^t f'(Z_s)\sigma(s, X_s)\,\mathrm{d}B_s,\end{aligned}$$

which is a local martingale and, by its definition, Z is a semimartingale satisfying (3.6). Conversely, if there exists a solution to the Volterra local martingale problem, we obtain a weak solution to the SVE (3.2) by using (3.6) and Proposition 3.3, which yields that $A_t = \int_0^t \mu(s, X_s)\,\mathrm{d}s$ and $M_t = \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s$ for some Brownian motion $(B_t)_{t\in[0,T]}$. ☐

3.2 Existence of weak solutions

In this section we establish the existence of a weak solution to the SVE (3.2) and, equivalently, of a solution to the associated Volterra local martingale problem, under suitable assumptions on the initial condition, coefficients and kernels, which we state in the following.

Assumption 3.8. *There is some $p \in (4, \infty)$ and some $\gamma \in (\frac{2}{p}, \frac{1}{2})$ such that:*

(i) There is a constant $C_p > 0$ such that, for all $(t, t') \in \Delta_T$,

$$\begin{aligned}\int_0^t |K_\mu(s,t') - K_\mu(s,t)|^{\frac{p}{p-1}}\,\mathrm{d}s + \int_t^{t'} |K_\mu(s,t')|^{\frac{p}{p-1}}\,\mathrm{d}s \le C_p|t'-t|^{\frac{\gamma p}{p-1}},\\ \int_0^t |K_\sigma(s,t') - K_\sigma(s,t)|^{\frac{2p}{p-2}}\,\mathrm{d}s + \int_t^{t'} |K_\sigma(s,t')|^{\frac{2p}{p-2}}\,\mathrm{d}s \le C_p|t'-t|^{\frac{2\gamma p}{p-2}}.\end{aligned} \tag{3.7}$$

(ii) The coefficients $\mu, \sigma\colon [0,T]\times\mathbb{R} \to \mathbb{R}$ are measurable functions such that for every compact set $\mathcal{K} \subset \mathbb{R}$ and every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \le \varepsilon, \quad t \in [0,T],\ x, y \in \mathcal{K} \text{ with } |x-y| \le \delta,$$

and μ, σ fulfill the linear growth condition

$$|\mu(t,x)| + |\sigma(t,x)| \le C_{\mu,\sigma}(1+|x|), \quad t \in [0,T],\ x \in \mathbb{R}, \tag{3.8}$$

for a constant $C_{\mu,\sigma} > 0$

(iii) The initial condition $x_0\colon [0,T] \to \mathbb{R}$ is β-Hölder continuous for every $\beta \in (0, \gamma - 1/p)$.

Note that Assumption 3.8 directly implies (2.3) with the choice $\varepsilon = \frac{2p}{p-2} - 2$, and vice versa (2.3) implies Assumption 3.8 (i) with $p = 4/\varepsilon + 2$ (and if necessary rescaling the exponent using Hölder's inequality if $\varepsilon \ge 2$ to secure $p > 4$).
To formulate our second assumption, for a measurable function $K\colon \Delta_T \to \mathbb{R}$, we say $K(\cdot, t)$ is absolutely continuous for every $t \in [0,T]$ if there exists an integrable function $\partial_1 K\colon \Delta_T \to \mathbb{R}$ such that $K(s,t) - K(0,t) = \int_0^s \partial_1 K(u,t)\,\mathrm{d}u$ for $(s,t) \in \Delta_T$.

Assumption 3.9. *The kernel K_μ is measurable and bounded in $L^1([0,T])$ uniformly in the second variable, i.e.*

$$\sup_{t\in[0,T]} \int_0^t |K_\mu(s,t)|\,\mathrm{d}s \le C$$

for some constant $C > 0$. The kernel K_σ is measurable and satisfies at least one of the following conditions:

(i) K_σ *is a bounded function and* $K_\sigma(\cdot,t)$ *is absolutely continuous for every* $t \in [0,T]$ *such that* $\partial_1 K_\sigma$ *fulfills*

$$\sup_{t\in[0,T]} \left| \int_0^t |\partial_1 K_\sigma(s,t)|^p \,\mathrm{d}s \right|^{\frac{1}{p}} \le C$$

for some $p > 1$ *and some constant* $C > 0$.

(ii) $K_\sigma(s,t) = \tilde{K}(t-s)$ *for all* $(s,t) \in \Delta_T$ *for a function* $\tilde{K} \in L^2([0,T])$.

Note, that Assumption 3.9 is satisfied by every convolutional kernel $K_\mu(s,t) = \tilde{K}(t-s)$ for all $(s,t) \in \Delta_T$ for a function $\tilde{K} \in L^1([0,T])$, and in case of Assumption 3.9 (i), the bound on the second summand in (3.7) is trivially fulfilled. With these assumptions at hand we are ready to state our main result.

Theorem 3.10. *Suppose Assumptions 3.8 and 3.9. Then, there exists a weak solution (in the sense of Definition 3.1) with* $(X_t)_{t\in[0,T]} \in C([0,T])$ *to the stochastic Volterra equation* (3.2).

Before proving the aforementioned existence result, let us briefly discuss some properties of weak solutions to the SVE (3.2) and some exemplary kernels.

Remark 3.11. *Suppose Assumption 3.8. Due to Lemma 2.10 and Corollary 2.11, any weak solution with* $(X_t)_{t\in[0,T]} \in C([0,T])$ *to the SVE* (3.2) *satisfies* $\sup_{t\in[0,T]} \mathbb{E}[|X_t|^q] < \infty$ *for any* $q \in [1,\infty)$ *and possesses a* β*-Hölder continuous modification for any* $\beta \in (0, \gamma - 1/p)$.

Remark 3.12. *Assumptions 3.8 (i) and 3.9 are satisfied, e.g., by the following type of diffusion kernels:*

(i) $K_\sigma(s,t) := (t-s)^{-\alpha}$ *for* $\alpha \in (0, \frac{1}{2})$ *for any* $p \in (\frac{6}{1-2\alpha}, \infty)$ *with* $\gamma = \frac{1}{2} - \alpha - \frac{1}{p}$,

(ii) $K_\sigma(s,t) := \tilde{K}(t-s)$ *for a Lipschitz continuous function* $\tilde{K}\colon [0,T] \to \mathbb{R}$,

(iii) regular kernels fulfilling Assumption 2.1, and

(iv) weakly differentiable kernels such that $\partial_1 K_\sigma(s,t) \le C(t-s)^{-\alpha}$ *for* $\alpha \in (0, \frac{1}{2})$.

The following calculation shows that Assumption 3.8 (i) holds for (i), the rest is easy to see.

$$\begin{aligned}
\int_0^t \left|(t'-s)^{-\alpha} - (t-s)^{-\alpha}\right|^{\frac{2p}{p-2}} \mathrm{d}s &\lesssim \int_0^t \left((t-s)^{-\frac{2p\alpha}{p-2}} - (t'-s)^{-\frac{2p\alpha}{p-2}}\right) \mathrm{d}s \\
&\lesssim -\left[(t-s)^{1-\frac{2p\alpha}{p-2}}\right]_0^t + \left[(t'-s)^{1-\frac{2p\alpha}{p-2}}\right]_0^t \\
&\lesssim t^{1-\frac{2p\alpha}{p-2}} + (t'-t)^{1-\frac{2p\alpha}{p-2}} - t'^{1-\frac{2p\alpha}{p-2}} \\
&\lesssim |t'-t|^{1-\frac{2p\alpha}{p-2}},
\end{aligned}$$

and with $\gamma = \frac{1}{2} - \alpha - p$*,*

$$\frac{2\gamma p}{p-2} = \frac{p - 2\alpha p - 2}{p-2} = 1 - \frac{2\alpha p}{p-2}.$$

The other terms in (3.7) *follow analogue.*

The remainder of the chapter is devoted to implement the proof of Theorem 3.10 based on several auxiliary lemmas. Note that Lemma 3.16 implies Theorem 3.10 due to Lemma 3.7. Note further that the continuity of $(X_t)_{t\in[0,T]}$ in Theorem 3.10 follows by the convergence $\hat{X}^k \to X$ in $C([0,T])$ in Lemma 3.15.

Assuming the coefficients μ, σ satisfy Assumption 3.8, the next lemma provides a way to approximate μ, σ locally uniformly by Lipschitz continuous coefficients.

Lemma 3.13. *Let* $f\colon [0,T]\times\mathbb{R} \to \mathbb{R}$ *be a measurable function such that for every compact set* $\mathcal{K} \subset \mathbb{R}$ *and every* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that*

$$|f(t,x) - f(t,y)| \le \varepsilon, \quad t \in [0,T],\, x,y \in \mathcal{K} \text{ with } |x-y| \le \delta,$$

and such that f *fulfills the linear growth condition*

$$|f(t,x)| \le C_f(1+|x|), \quad t \in [0,T],\, x \in \mathbb{R}, \tag{3.9}$$

for some constant $C_f > 0$*. Then, there is a sequence* $(f_n)_{n\in\mathbb{N}}$ *of measurable functions* $f_n\colon [0,T]\times\mathbb{R}\to\mathbb{R}$*, which satisfies:*

(i) linear growth: for $C_f > 0$ *as in* (3.9)*, we have*

$$|f_n(t,x)| \le 2C_f(1+|x|), \quad t \in [0,T],\, x \in \mathbb{R};$$

(ii) Lipschitz continuity: for each $n \in \mathbb{N}$ *there is a* $C_n > 0$ *such that*

$$|f_n(t,x) - f_n(t,y)| \le C_n|x-y|, \quad t \in [0,T],\, x,y \in \mathbb{R};$$

(iii) locally uniform convergence: for all $r \in (0,\infty)$ *we have*

$$\sup_{t\in[0,T],x\in[-r,r]} |f(t,x) - f_n(t,x)| \to 0, \quad \text{as } n \to \infty.$$

Proof. We explicitly choose the sequence $(f_n)_{n\in\mathbb{N}}$ by

$$f_n(t,x) := \phi_n(x)\int_{\mathbb{R}} f(t,x-y)\delta_n(y)\,\mathrm{d}y, \quad n \in \mathbb{N},$$

for some $\phi_n \in C_0^2(\mathbb{R})$ with support in $[-(n+1), n+1]$, taking values in $[0,1]$ and being identical to 1 on $[-n,n]$, where $\delta_n(y) := \frac{1}{c_n}(1-y^2)^n\mathbf{1}_{[-1,1]}(y)$ with $c_n := \int_{[-1,1]}(1-y^2)^n\,\mathrm{d}y$.

(i) For $t \in [0,T]$ and $x \in \mathbb{R}$, using the linear growth condition on f, we get

$$|f_n(t,x)| \leq C_f \int_{[-1,1]} (1+|x-y|)\delta_n(y)\,\mathrm{d}y \leq C_f \int_{[-1,1]} (2+|x|)\delta_n(y)\,\mathrm{d}y \leq 2C_f(1+|x|).$$

(ii) Let $t \in [0,T]$, $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$. Using the compact support of f_n and the fact, that every δ_n is Lipschitz continuous as a smooth function with compact support, we get

$$\big|f_n(t,x) - f_n(t,y)\big| \leq C_f c_n |x-y| \int_{-(n+2)}^{n+2} (1+|z|)\,\mathrm{d}z \leq C_n|x-y|$$

for some constant C_n.

(iii) Due to the continuity property of f, we can find for every $r > 0$ and for every $\varepsilon > 0$ some $\delta > 0$ such that for all $x, y \in [-r,r]$ with $|x-y| \leq \delta$ and all $t \in [0,T]$ holds $|f(t,x) - f(t,y)| \leq \varepsilon$. Assuming $n \in \mathbb{N}$ to be large enough that $\phi_n \equiv 1$ on $[-r,r]$, we get for any $x \in [-r,r]$,

$$\begin{aligned}&|f(t,x) - f_n(t,x)| \\ &\quad = \int_{[-\delta,\delta]} \delta_n(y)\big|f(t,x) - f(t,x-y)\big|\,\mathrm{d}y + \int_{[-1,1]\setminus[-\delta,\delta]} \delta_n(y)\big|f(t,x) - f(t,x-y)\big|\,\mathrm{d}y.\end{aligned}$$

Let now $N(\varepsilon,r) > 0$ be big enough, such that $\int_{[-1,1]\setminus[-\delta,\delta]} \delta_n(y)\,\mathrm{d}y < \varepsilon$ and $\phi_n \equiv 1$ on $[-r,r]$ for all $n \geq N(\varepsilon,r)$. Then, setting $\tilde{r} := r+1$ for all $n \geq N(\varepsilon,r)$

$$|f(t,x) - f_n(t,x)| \leq \int_{[-\delta,\delta]} \delta_n(y)\varepsilon\,\mathrm{d}y + 2\varepsilon \sup_{\substack{s\in[0,T],\\ \tilde{x}\in[-\tilde{r},\tilde{r}]}} |f(s,\tilde{x})| \leq \varepsilon\bigg(1 + 2\sup_{\substack{s\in[0,T],\\ \tilde{x}\in[-\tilde{r},\tilde{r}]}} |f(s,\tilde{x})|\bigg),$$

which tends to zero as $\varepsilon \to 0$. □

A suitable approximation, like the one provided in Lemma 3.13, ensures the convergence of associated Riemann–Stieltjes integrals. We denote by $C([0,T];\mathbb{R})$ the space of all continuous functions $g\colon [0,T] \to \mathbb{R}$, which is equipped with the supremum norm $\|\cdot\|_\infty$.

Lemma 3.14. *Let $f\colon [0,T]\times\mathbb{R} \to \mathbb{R}$ be a function such that for every compact set $\mathcal{K} \subset \mathbb{R}$ and every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$|f(t,x) - f(t,y)| \leq \varepsilon, \quad t \in [0,T],\ x,y \in \mathcal{K} \text{ with } |x-y| \leq \delta, \tag{3.10}$$

and $(f_k)_{k\in\mathbb{N}}$ be a sequence of functions such that $f_k\colon [0,T]\times\mathbb{R} \to \mathbb{R}$ and $|f(t,x)| + |f_k(t,x)| \leq C(1+|x|^2)$, $t \in [0,T]$, $x \in \mathbb{R}$, for all $k \in \mathbb{N}$ and for some $C > 0$, and $f_k \to f$ locally uniformly. Let $K\colon \Delta_T \to \mathbb{R}$ be measurable and bounded in $L^1([0,T])$ uniformly in the second variable, i.e. $\sup_{t\in[0,T]} \int_0^t |K(s,t)|\,\mathrm{d}s \leq M$ for some $M > 0$. If $(X^k)_{k\in\mathbb{N}}$ is a sequence of continuous stochastic processes such that $X^k \to X$ in $C([0,T];\mathbb{R})$ as $k \to \infty$ $\mathbb{P}$-a.s, then

$$\Big(\int_0^\cdot K(s,\cdot)f_k(s,X_s^k)\,\mathrm{d}s\Big)_{t\in[0,T]} \xrightarrow{\mathbb{P}} \Big(\int_0^\cdot K(s,\cdot)f(s,X_s)\,\mathrm{d}s\Big)_{t\in[0,T]} \quad \text{w.r.t. } \|\cdot\|_\infty, \quad k \to \infty,$$

where $\xrightarrow{\mathbb{P}}$ denotes convergence in probability.

Proof. First, note that due to the continuity condition (3.10), for every $n \in \mathbb{N}$ there exists some continuous non-decreasing function $g_n \colon [0,\infty) \to [0,\infty)$ with $g_n(0) = 0$, such that for all $x, y \in [-n, n]$,

$$|f(t,x) - f(t,y)| \leq g_n(|x-y|), \quad t \in [0,T].$$

Let $\varepsilon > 0$ and $\delta > 0$ be fixed but arbitrary. Choose $N \in \mathbb{N}$ and $K \in \mathbb{N}$ big enough such that

$$\mathbb{P}\big(\|X\|_\infty \geq n/2\big) \leq \delta/4 \qquad \text{and} \qquad \mathbb{P}\big(\|X^k - X\|_\infty \geq n/2\big) \leq \delta/4,$$

for all $n \geq N$ and $k \geq K$. Then,

$$\begin{aligned}\mathbb{P}\big(\|X\|_\infty \vee \|X^k\|_\infty \geq n\big) &\leq \mathbb{P}\Big(\{\|X\|_\infty \geq n\} \cup \{\|X^k - X\|_\infty + \|X\|_\infty \geq n\}\Big) \\ &\leq \mathbb{P}\big(\|X\|_\infty \geq n/2\big) + \mathbb{P}\big(\|X^k - X\|_\infty \geq n/2\big) \\ &\leq \delta/4 + \delta/4 = \delta/2.\end{aligned}$$

For every $n, k \in \mathbb{N}$, on $\{\|X\|_\infty \vee \|X^k\|_\infty \leq n\}$ we can bound for $t \in [0,T]$,

$$\begin{aligned}A^k_t - A_t &:= \int_0^t K(s,t) f_k(s, X^k_s)\,\mathrm{d}s - \int_0^t K(s,t) f(s, X_s)\,\mathrm{d}s \\ &\leq \int_0^t |K(s,t)|\big|f_k(s,X^k_s) - f(s,X^k_s)\big|\,\mathrm{d}s + \int_0^t |K(s,t)|\big|f(s,X^k_s) - f(s,X_s)\big|\,\mathrm{d}s \\ &\leq M\Bigg(\sup_{t\in[0,T],\, x\in[-n,n]} |f_k(t,x) - f(t,x)| + g_n\big(\|X^k - X\|_\infty\big)\Bigg), \end{aligned} \tag{3.11}$$

with $\sup_{t\in[0,T]} \int_0^t |K(s,t)|\,\mathrm{d}s \leq M$. For every $n \in \mathbb{N}$ we choose $K^n_{\varepsilon\delta} \in \mathbb{N}$ sufficiently large such that

$$\mathbb{P}\Bigg(\sup_{t\in[0,T],\, x\in[-n,n]} |f_k(t,x) - f(t,x)| + g_n(\|X^k - X\|_\infty) \geq \varepsilon/M\Bigg) \leq \delta/2, \quad k \geq K^n_{\varepsilon\delta}.$$

Setting $K_{\varepsilon\delta} := \max\{K^N_{\varepsilon\delta}, K\}$, we get

$$\begin{aligned}&\mathbb{P}\big(\|A^k - A\|_\infty \geq \varepsilon\big) \\ &\quad\leq \mathbb{P}\Big(\{\|A^k - A\|_\infty \geq \varepsilon\} \cap \{\|X\|_\infty \vee \|X^k\|_\infty < N\}\Big) + \mathbb{P}\big(\|X\|_\infty \vee \|X^k\|_\infty \geq N\big) \\ &\quad\leq \mathbb{P}\Bigg(\sup_{t\in[0,T],\, x\in[-N,N]} |f_k(t,x) - f(t,x)| + g_N(\|X^k - X\|_\infty) \geq \varepsilon/M\Bigg) + \delta/2 \leq \delta,\end{aligned}$$

for all $k \geq K_{\varepsilon\delta}$, which shows the desired convergence. □

Given coefficients μ, σ satisfying Assumption 3.8, we fix, relying on Lemma 3.13, two sequences $(\mu_n)_{n\in\mathbb{N}}$ and $(\sigma_n)_{n\in\mathbb{N}}$ with

$$\mu_n \colon [0,T] \times \mathbb{R} \to \mathbb{R} \quad \text{and} \quad \sigma_n \colon [0,T] \times \mathbb{R} \to \mathbb{R},$$

that fulfill properties (i)-(iii) of Lemma 3.13. For every $n \in \mathbb{N}$, we define $(X_t^n)_{t\in[0,T]}$ as the unique (strong) solution (see page 14 for the definition of unique strong solutions to SVEs) to the stochastic Volterra equation

$$X_t^n = x_0(t) + \int_0^t K_\mu(s,t)\mu_n(s,X_s^n)\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma_n(s,X_s^n)\,\mathrm{d}B_s, \quad t \in [0,T], \quad (3.12)$$

given a Brownian motion $(B_t)_{t\in[0,T]}$ on some probability space $(\Omega,\mathcal{F},\mathbb{P})$. Note that $(X_t^n)_{t\in[0,T]}$ exists by [Wan08, Theorem 1.1] due to the Lipschitz continuity of μ_n and σ_n. Furthermore, we introduce the sequences $(A^n)_{n\in\mathbb{N}}$ and $(M^n)_{n\in\mathbb{N}}$ by

$$A_t^n := \int_0^t \mu_n(s,X_s^n)\,\mathrm{d}s \quad \text{and} \quad M_t^n := \int_0^t \sigma_n(s,X_s^n)\,\mathrm{d}B_s, \qquad t\in[0,T]. \qquad (3.13)$$

In the following, we denote $X \overset{\mathscr{D}}{\sim} Y$ for equality in law of stochastic processes X and Y.

Lemma 3.15. *Suppose Assumption 3.8 and let $(X^n)_{n\in\mathbb{N}}$, $(A^n)_{n\in\mathbb{N}}$ and $(M^n)_{n\in\mathbb{N}}$ be given by (3.12) and (3.13). Then, there exist continuous stochastic processes $(\hat{X}^k)_{k\in\mathbb{N}}$, $(\hat{A}^k)_{k\in\mathbb{N}}$, $(\hat{M}^k)_{k\in\mathbb{N}}$, X, A, M and a Brownian motion $(\tilde{B}_t)_{t\in[0,T]}$ on a common probability space $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbb{P}})$ such that $(\hat{X}^k,\hat{A}^k,\hat{M}^k) \to (X,A,M)$ in $C([0,T];\mathbb{R}^3)$ as $k\to\infty$ $\tilde{\mathbb{P}}$-a.s., $(\hat{X}^k,\hat{A}^k,\hat{M}^k) \overset{\mathscr{D}}{\sim} (X^{n_k},A^{n_k},M^{n_k})$ and M is a local martingale with the representation*

$$M_t = \int_0^t \sigma(s,X_s)\,\mathrm{d}\tilde{B}_s, \quad t\in[0,T],$$

where $(X^{n_k},A^{n_k},M^{n_k})_{k\in\mathbb{N}}$ denotes some subsequence of $(X^n,A^n,M^n)_{n\in\mathbb{N}}$.

Proof. First we want to apply Kolmogorov's tightness criterion (see [KS91, Problem 2.4.11]) to the probability measures $(\mathbb{P}_{(X^n,A^n,M^n,B)})_{n\in\mathbb{N}}$ associated to the four-dimensional stochastic processes $(X^n,A^n,M^n,B)_{n\in\mathbb{N}}$. By Lemma 3.13 (i) we know, that the coefficients μ_n and σ_n fulfill the linear growth condition (3.8) with uniformly bounded constants, i.e. $C_{\mu_n,\sigma_n} \le 2C_{\mu,\sigma}$ for all $n\in\mathbb{N}$. Hence, using $p\in(4,\infty)$ from Assumption 3.8, we deduce, by Lemma 2.10, that

$$\sup_{n\in\mathbb{N}} \sup_{s\in[0,T]} \mathbb{E}[|X_s^n|^p] \le C\Big(1 + \sup_{s\in[0,T]} |x_0(s)|\Big)^p < \infty,$$

and, by Lemma 2.7 and Remark 2.9, that

$$\mathbb{E}[|X_{t'}^n - x_0(t') - X_t^n - x_0(t)|^p] \le C|t'-t|^{\beta p}, \quad n\in\mathbb{N},$$

for every $\beta\in(0,\gamma-1/p)$, where the constant $C>0$ depends only on p, T, K_μ, K_σ and $C_{\mu,\sigma}$. Moreover, it is straightforward to show that

$$\mathbb{E}[|A_{t'}^n - A_t^n|^p] \le C|t'-t|^{\frac{p}{2}} \quad \text{and} \quad \mathbb{E}[|M_{t'}^n - M_t^n|^p] \le C|t'-t|^{\frac{p}{2}}$$

for all $0 \le t \le t' \le T$ and some constant $C > 0$, by Hölder's inequality and Burkholder–Davis–Gundy's inequality, respectively. Choosing β sufficiently close to $\gamma - 1/p$ so that $\beta p > 1$, which is possible due to Assumption 3.8, and noting that the initial distributions $(X_0^n, A_0^n, M_0^n, B_0)_{n\in\mathbb{N}}$ are independent of n, we can apply Kolmogorov's tightness criterion to obtain the tightness of the sequence $(\mathbb{P}_{(X^n,A^n,M^n,B)})_{n\in\mathbb{N}}$. Hence, by Prohorov's theorem ([KS91, Theorem 2.4.7]) we get relative compactness ([KS91, Definition 2.4.6]) of the sequence of measures $(\mathbb{P}_{(X^n,A^n,M^n,B)})_{n\in\mathbb{N}}$ in $\mathcal{M}_1(C([0,T];\mathbb{R}^4))$, which denotes the space of all probability measures on $C([0,T];\mathbb{R}^4)$. Consequently, there exists a converging subsequence $(\mathbb{P}_{(X^{n_k},A^{n_k},M^{n_k},B)})_{k\in\mathbb{N}}$ such that

$$\mathbb{P}_{(X^{n_k},A^{n_k},M^{n_k},B)} \to \mathbb{P}_{(X,A,M,B)} \quad \text{weakly} \quad \text{as} \quad k \to \infty,$$

for some measure $\mathbb{P}_{(X,A,M,B)}$ in $\mathcal{M}_1(C([0,T];\mathbb{R}^4))$.
The Skorokhod representation theorem (see e.g. [Dud02, Theorem 11.7.2]) yields the existence of some probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$ with continuous stochastic processes $(\hat{X}^k)_{k\in\mathbb{N}}$, $(\hat{A}^k)_{k\in\mathbb{N}}$, $(\hat{M}^k)_{k\in\mathbb{N}}$, $(\hat{B}^k)_{k\in\mathbb{N}}$ and X, A, M, $\hat{B}$ on it such that

$$(X^{n_k}, A^{n_k}, M^{n_k}, B) \overset{\mathscr{D}}{\sim} (\hat{X}^k, \hat{A}^k, \hat{M}^k, \hat{B}^k), \qquad k \in \mathbb{N},$$

and

$$(\hat{X}^k, \hat{A}^k, \hat{M}^k, \hat{B}^k) \to (X, A, M, \hat{B}) \quad \text{in} \quad C([0,T];\mathbb{R}^4) \quad \text{as} \quad k \to \infty, \quad \hat{\mathbb{P}}\text{-a.s.}$$

From a general version of the Yamada–Watanabe result, see [Kur14, Theorem 1.5], we can deduce that $\hat{M}_t^k = \int_0^t \sigma_{n_k}(s, \hat{X}_s^k)\,\mathrm{d}\hat{B}_s^k$, for $t \in [0,T]$ and for all $k \in \mathbb{N}$, and the stochastic processes $(B^k)_{k\in\mathbb{N}}$ are Brownian motions as $\hat{B}^k \overset{\mathscr{D}}{\sim} B$. Thus, $\hat{M}^k$ is a local $\hat{\mathbb{P}}$-martingale with quadratic variation $\langle \hat{M}^k \rangle_t = \int_0^t \sigma_{n_k}(s, \hat{X}_s^k)^2\,\mathrm{d}s$.
Due to the $\hat{\mathbb{P}}$-a.s. convergence of $(\hat{M}^k)_{k\in\mathbb{N}}$ to M, [JS03, Proposition IX.1.17] implies that M is also a local $\hat{\mathbb{P}}$-martingale, and the convergence of $\int_0^t \sigma_{n_k}(s, X_s^{n_k})^2\,\mathrm{d}s$ in probability, see Lemma 3.14, together with [JS03, Corollary VI.6.29] implies that the quadratic variation of M is $\langle M \rangle_t = \int_0^t \sigma(s, X_s)^2\,\mathrm{d}s$. Therefore, the representation theorem for local martingales with absolutely continuous quadratic variations (see e.g. [KS91, Theorem 3.4.2]) yields the existence of some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$, which is an extension of $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$, and a Brownian motion $(\tilde{B}_t)_{t\in[0,T]}$ on it, such that $M_t = \int_0^t \sigma(s, X_s)\,\mathrm{d}\tilde{B}_s$ for $t \in [0,T]$. $\square$

Using the stochastic processes X, A and M from Lemma 3.15, we can construct a solution to the Volterra local martingale problem in the sense of Definition 3.4.

Lemma 3.16. *Suppose Assumptions 3.8 and 3.9. There exists a solution to the Volterra local martingale problem given* $(x_0, \mu, \sigma, K_\mu, K_\sigma)$.

Proof. Recall, the stochastic processes $(X^n)_{n\in\mathbb{N}}$, $(A^n)_{n\in\mathbb{N}}$ and $(M^n)_{n\in\mathbb{N}}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ are given in (3.13) and $(\hat{X}^k)_{k\in\mathbb{N}}$, $(\hat{A}^k)_{k\in\mathbb{N}}$, $(\hat{M}^k)_{k\in\mathbb{N}}$, X, A and M on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ are given by Lemma 3.15. We introduce the stochastic processes $(Z^n)_{n\in\mathbb{N}}$, $(\hat{Z}^k)_{k\in\mathbb{N}}$ and Z by

$$Z_t^n := A_t^n + M_t^n, \quad \hat{Z}_t^k := \hat{A}_t^k + \hat{M}_t^k \quad \text{and} \quad Z_t := A_t + M_t, \quad t \in [0,T].$$

We shall show that the triple (X, Z), $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$, $(\mathcal{F}_t^X)_{t\in[0,T]}$, where $(\mathcal{F}_t^X)_{t\in[0,T]}$ denotes the augmented natural filtration of X (cf. [KS91, Definition 2.7.2]), solves the Volterra local martingale problem given $(x_0, \mu, \sigma, K_\mu, K_\sigma)$. Since the properties (i)-(iii) of Definition 3.4 are fairly easy to check, we verify here that

(iv) the process $(\mathcal{M}_t^f)_{t\in[0,T]}$ defined by (3.5) is a local $\tilde{\mathbb{P}}$-martingale for every $f \in C_0^2(\mathbb{R})$,

(v) the equality (3.6) holds $\tilde{\mathbb{P}}$-a.s.

(iv) For $k \in \mathbb{N}$ and $f \in C_0^2(\mathbb{R})$, the stochastic process $(\mathcal{M}_t^{f,k})_{t\in[0,T]}$ is defined by

$$\mathcal{M}_t^{f,k} := f(\hat{Z}_t^k) - \int_0^t \mathcal{A}^{f,k}(s, \hat{X}_s^k, \hat{Z}_s^k)\,\mathrm{d}s, \quad t \in [0,T],$$

where $\mathcal{A}^{f,k}(t,x,z) := \mu_{n_k}(t,x)f'(z) + \frac{1}{2}\sigma_{n_k}(t,x)^2 f''(z)$. Due to $(\hat{X}^k, \hat{Z}^k) \overset{\mathscr{D}}{\sim} (X^{n_k}, Z^{n_k})$ and since (X^{n_k}, Z^{n_k}) solves the Volterra local martingale problem given $(x_0, \mu_{n_k}, \sigma_{n_k}, K_\mu, K_\sigma)$ on $(\Omega, \mathcal{F}, \mathbb{P})$ by construction and Lemma 3.7, it follows that $(\mathcal{M}_t^{f,k})_{t\in[0,T]}$ is a local martingale on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ for every $k \in \mathbb{N}$. Moreover, Lemma 3.14 implies that $\mathcal{M}^{f,k} \to \mathcal{M}^f$ weakly as $k \to \infty$ and, thus, by [JS03, Proposition IX.1.17], the limiting process $(\mathcal{M}_t^f)_{t\in[0,T]}$ is a local martingale on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$.

(v) Since $(\hat{X}^k, \hat{M}^k) \overset{\mathscr{D}}{\sim} (X^{n_k}, M^{n_k})$ for every $k \in \mathbb{N}$ and pathwise uniqueness holds for SVEs with Lipschitz continuous coefficients (see e.g. [Wan08, Theorem 1.1]), the general version of the Yamada–Watanabe result ([Kur14, Theorem 1.5]) yields that $\hat{X}^k$ can be represented as the stochastic output of the Volterra equation (3.2) from the stochastic input $\hat{M}^k$ in the same way as X^{n_k} from M^{n_k}, hence, we get that

$$\hat{X}_t^k = x_0(t) + \int_0^t K_\mu(s,t)\mu_{n_k}(s, \hat{X}_s^k)\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\,\mathrm{d}\hat{M}_s^k, \quad t \in [0,T], \quad \tilde{\mathbb{P}}\text{-a.s.}, \tag{3.14}$$

holds. To continue the proof of (v), we need to distinguish between (a) bounded kernels and (b) kernels of convolutional type.

(a) We start with the bounded kernels as in Assumption 3.9 (i). Due to the absolute continuity of K_σ in the first variable, we can apply the integration by part formula for semimartingales (see [RW00, Theorem (VI).38.3]) to rewrite (3.14) to

$$\hat{X}_t^k = x_0(t) + \int_0^t K_\mu(s,t)\mu_{n_k}(s, \hat{X}_s^k)\,\mathrm{d}s + K_\sigma(t,t)\hat{M}_t^k + \int_0^t \hat{M}_s^k \partial_1 K_\sigma(s,t)\,\mathrm{d}s. \tag{3.15}$$

Since $(\hat{X}^k, \hat{M}^k) \to (X, M)$ in $C([0,T];\mathbb{R}^2)$ as $k \to \infty$, $\tilde{\mathbb{P}}$-a.s., and K_σ is bounded, we obtain by Lemma 3.14 that $\hat{X}^k \to X$ and $K_\sigma \hat{M}^k \to K_\sigma M$ in $C([0,T];\mathbb{R})$ as $k \to \infty$, $\tilde{\mathbb{P}}$-a.s., and $\int_0^{\cdot} K_\mu(s,\cdot)\mu_{n_k}(s, \hat{X}_s^k)\,\mathrm{d}s \to \int_0^t K_\mu(s,\cdot)\,\mathrm{d}A_s$ in $C([0,T];\mathbb{R})$ in probability as $k \to \infty$. Furthermore, applying Hölder's inequality with $p > 4$ (see Assumption 3.9) and denoting

$q = p/(p-1)$, we get by the integrability of $\partial_1 K_\sigma$ that

$$\left\| \int_0^{\cdot} (\hat{M}_s^k - M_s)\partial_1 K_\sigma(s,\cdot)\,\mathrm{d}s \right\|_\infty \leq \left(\int_0^T |\hat{M}_s^k - M_s|^q \,\mathrm{d}s \right)^{\frac{1}{q}} \left\| \int_0^{\cdot} |\partial_1 K_\sigma(s,\cdot)|^p \,\mathrm{d}s \right\|_\infty^{\frac{1}{p}}$$
$$\leq C\|\hat{M}^k - M\|_\infty.$$

Hence, the $\tilde{\mathbb{P}}$-a.s. convergence $(\hat{M}^k)_{k\in\mathbb{N}}$ to M implies $\int_0^{\cdot} \hat{M}_s^k K_\sigma(s,\cdot)\,\mathrm{d}s \to \int_0^{\cdot} M_s K_\sigma(s,\cdot)\,\mathrm{d}s$ as $k\to\infty$ $\tilde{\mathbb{P}}$-a.s., and we can take the limit in probability in (3.15) or the $\tilde{\mathbb{P}}$-a.s. limit for some subsequence, to obtain that (3.6) holds $\tilde{\mathbb{P}}$-a.s.
(b) For convolution kernels as in Assumption 3.9 (ii), we integrate both sides of (3.14) and use the stochastic Fubini theorem (see e.g. [Ver12, Theorem 2.2]) twice to obtain

$$\begin{aligned}
\int_0^t \hat{X}_s^k \,\mathrm{d}s &= \int_0^t x_0(s)\,\mathrm{d}s + \int_0^t \int_0^s K_\mu(s,u)\,\mathrm{d}\hat{A}_u^k\,\mathrm{d}s + \int_0^t \int_0^s K_\sigma(s-u)\,\mathrm{d}\hat{M}_u^k\,\mathrm{d}s \\
&= \int_0^t x_0(s)\,\mathrm{d}s + \int_0^t \int_0^s K_\mu(s,u)\,\mathrm{d}\hat{A}_u^k\,\mathrm{d}s + \int_0^t \int_u^t K_\sigma(s-u)\,\mathrm{d}s\,\mathrm{d}\hat{M}_u^k \\
&= \int_0^t x_0(s)\,\mathrm{d}s + \int_0^t \int_0^s K_\mu(s,u)\,\mathrm{d}\hat{A}_u^k\,\mathrm{d}s + \int_0^t \int_0^{t-u} K_\sigma(s)\,\mathrm{d}s\,\mathrm{d}\hat{M}_u^k \\
&= \int_0^t x_0(s)\,\mathrm{d}s + \int_0^t \int_0^s K_\mu(s,u)\,\mathrm{d}\hat{A}_u^k\,\mathrm{d}s + \int_0^t K_\sigma(s) \int_0^{t-s} \mathrm{d}\hat{M}_u^k\,\mathrm{d}s \\
&= \int_0^t x_0(s)\,\mathrm{d}s + \int_0^t \int_0^s K_\mu(s,u)\,\mathrm{d}\hat{A}_u^k\,\mathrm{d}s + \int_0^t K_\sigma(t-s)\hat{M}_s^k\,\mathrm{d}s. \qquad (3.16)
\end{aligned}$$

Since

$$\left\| \int_0^{\cdot} K_\sigma(\cdot - s)(\hat{M}_s^k - M_s)\,\mathrm{d}s \right\|_\infty \leq \|\hat{M}^k - M\|_\infty \int_0^T |K_\sigma(T-s)|\,\mathrm{d}s \leq C\|\hat{M}^k - M\|_\infty$$

and $\hat{M}^k \to M$ as $k\to\infty$, $\tilde{\mathbb{P}}$-a.s, we obtain $\int_0^{\cdot} K_\sigma(t-s)\hat{M}_s^k\,\mathrm{d}s \to \int_0^{\cdot} K_\sigma(t-s)M_s\,\mathrm{d}s$ as $k\to\infty$, $\tilde{\mathbb{P}}$-a.s. The convergence of $\int_0^t\int_0^s K_\mu(s,u)\,\mathrm{d}\hat{A}_u^k\,\mathrm{d}s$ follows as in (a). Thus, taking the $\tilde{\mathbb{P}}$-a.s. limit of both sides of (3.16) and then taking the derivative yields that (3.6) holds for (X,Z), $\tilde{\mathbb{P}}$-a.s. $\square$

Chapter 4

Pathwise uniqueness for the fractional SVE

The content of this chapter is published in [PS22].

Introduction

In this chapter, we study the one-dimensional fractional stochastic Volterra equation (SVE) which has the form

$$X_t = x_0(t) + \int_0^t (t-s)^{-\alpha}\mu(s, X_s)\,\mathrm{d}s + \int_0^t (t-s)^{-\alpha}\sigma(s, X_s)\,\mathrm{d}B_s, \quad t \in [0,T], \qquad (4.1)$$

where $\alpha \in [0, \frac{1}{2})$, $x_0\colon [0,T] \to \mathbb{R}$ is a continuous function, $\mu, \sigma\colon [0,T] \times \mathbb{R} \to \mathbb{R}$ are measurable functions and $(B_t)_{t\in[0,T]}$ is a standard Brownian motion. Although the stochastic integral in (4.1) is defined as a classical stochastic Itô integral, a potential solution of this SVE is, in general, neither a semimartingale nor a Markov process. Assuming that μ is Lipschitz continuous and σ is ξ-Hölder continuous for $\xi \in (\frac{1}{2(1-\alpha)}, 1]$, we show that pathwise uniqueness for the SVE (4.1) holds and, consequently, that there exists a unique strong solution.

As long as the kernels of a one-dimensional SVE are sufficiently regular, i.e. excluding the singular kernel $(t-s)^{-\alpha}$ in (4.1), the existence of unique strong solutions can be still obtained when the diffusion coefficients are only 1/2-Hölder continuous, see Chapter 2 or [AJEE19b]. The latter results are crucially based on the observation that solutions to SVEs with sufficiently regular kernels are semimartingales, allowing to rather directly implement approaches in the spirit of Yamada–Watanabe [YW71].

A major challenge to prove pathwise uniqueness for the SVE (4.1) with its singular fractional kernel $K_\alpha(s,t) = (t-s)^{-\alpha}$ is the missing natural semimartingale representation of its potential solution. Assuming the drift coefficient μ does not depend on the solution $(X_t)_{t\in[0,T]}$ and the diffusion coefficient σ is ξ-Hölder continuous for $\xi \in (\frac{1}{2(1-\alpha)}, 1]$, Mytnik and Salisbury [MS15] established pathwise uniqueness for the SVE (4.1) by equivalently reformulating the SVE into a stochastic partial differential equation, which then allows to accomplish a proof of pathwise uniqueness in the spirit of Yamada–Watanabe relying on the methodology developed in [MPS06, MP11]. In this chapter, we generalize the results and method of Mytnik and Salisbury [MS15] to derive pathwise uniqueness for the stochastic Volterra equation (4.1) with general time-inhomogeneous coefficients. As classical transforms allowing to remove the drift of an SDE are not applicable to the SVE (4.1), the general time-inhomogeneous coefficients μ creates severe novel challenges. For the sake of readability, all proofs are presented in a self-contained manner although some intermediate steps can already be found in the work [MS15] of Mytnik and Salisbury. The existence of a unique strong solution to the stochastic Volterra equation (4.1) follows by a general version of Yamada–Watanabe theorem (see [YW71, Kur14]) stating that the combination of pathwise uniqueness and the existence of weak solutions to the SVE (4.1) (as obtained in Chapter 3) guarantees the existence of a strong solution. Let us remark that strong existence and pathwise uniqueness play a crucial role in the context of large deviation and as key ingredients to fully justify some numerical schemes, see e.g. [DE97, Mao94].

Organization of the chapter: Section 4.1 presents the main results on the pathwise uniqueness and strong existence of solutions to stochastic Volterra equations. Section 4.2 contains the main steps in the proof of pathwise uniqueness, while the remaining Sections 4.3-4.6 provide the necessary auxiliary results to implement these main steps.

4.1 Main results

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space, which satisfies the usual conditions, $(B_t)_{t\in[0,T]}$ be a standard Brownian motion and $T \in (0,\infty)$. We consider the one-dimensional stochastic Volterra equation (SVE)

$$X_t = x_0(t) + \int_0^t (t-s)^{-\alpha}\mu(s, X_s)\,\mathrm{d}s + \int_0^t (t-s)^{-\alpha}\sigma(s, X_s)\,\mathrm{d}B_s, \quad t \in [0,T], \tag{4.2}$$

where $\alpha \in [0, \frac{1}{2})$, $x_0\colon [0,T] \to \mathbb{R}$ is a continuous function and $\mu, \sigma\colon [0,T]\times\mathbb{R} \to \mathbb{R}$ are deterministic, measurable functions. Furthermore, $\int_0^t (t-s)^{-\alpha}\mu(s,X_s)\,\mathrm{d}s$ is defined as a Riemann–Stieltjes integral and $\int_0^t (t-s)^{-\alpha}\sigma(s,X_s)\,\mathrm{d}B_s$ as an Itô integral.

The regularity of the coefficients μ and σ and of the initial condition x_0 is determined in the following assumption.

Assumption 4.1. *Let $\alpha \in [0, \frac{1}{2})$, let x_0 be absolutely continuous and let $\mu, \sigma\colon [0,T]\times\mathbb{R} \to \mathbb{R}$ be measurable functions such that*

(i) μ and σ are of linear growth, i.e. there is a constant $C_{\mu,\sigma} > 0$ such that

$$|\mu(t,x)| + |\sigma(t,x)| \le C_{\mu,\sigma}(1 + |x|),$$

for all $t \in [0,T]$ and $x \in \mathbb{R}$.

(ii) *μ is Lipschitz continuous and σ is Hölder continuous in the space variable uniformly in time of order ξ for some $\xi \in [\frac{1}{2}, 1]$ such that*

$$\xi > \frac{1}{2(1-\alpha)},$$

where in the case of $\alpha = 0$ even equality is allowed. Hence, there are constants $C_\mu, C_\sigma > 0$ such that

$$|\mu(t,x) - \mu(t,y)| \leq C_\mu |x-y| \quad \text{and} \quad |\sigma(t,x) - \sigma(t,y)| \leq C_\sigma |x-y|^\xi$$

hold for all $t \in [0,T]$ and $x, y \in \mathbb{R}$.

(iii) *For every $K > 0$, there is some constant $C_K > 0$ such that, for every $t \in [0,T]$ and every $x, y \in [-K, K]$,*

$$\left| \frac{\mu(t,x) - \mu(t,y)}{\sigma(t,x) - \sigma(t,y)} \right| \leq C_K,$$

where we use the convention $0/0 := 1$.

Assumption 4.1 is a standing assumption throughout the entire chapter. Although not always explicitly stated all results are proven supposing Assumption 4.1.

Remark 4.2. *Given Assumption 4.1 (ii), Assumption 4.1 (iii) is a fairly mild restriction. Consider, for example, some Lipschitz continuous function μ and σ of the form $\sigma(t,x) = \operatorname{sgn}(x)|x|^\xi$ for $\xi \in [1/2, 1]$. Note that, in interesting cases like the rough Heston model in mathematical finance, solutions to (4.2) are non-negative (see [AJEE19a, Theorem A.2]), so that the sgn in the definition of σ does not influence the dynamics of the associated SVE. Then, for $|x|, |y| \leq K$, using the inequality $\left|\operatorname{sgn}(x)|x|^\xi - \operatorname{sgn}(y)|y|^\xi\right| \geq K^{-1}|x-y|$, we get*

$$\left| \frac{\mu(t,x) - \mu(t,y)}{\sigma(t,x) - \sigma(t,y)} \right| \leq C_\mu \frac{|x-y|}{\left|\operatorname{sgn}(x)|x|^\xi - \operatorname{sgn}(y)|y|^\xi\right|} \leq C_\mu \frac{|x-y|}{K^{-1}|x-y|} = C_\mu K < \infty.$$

Based on Assumption 4.1, we obtain a unique strong solution of the stochastic Volterra equation (4.2). Therefore, let us briefly recall the concepts of strong solutions and pathwise uniqueness. Let for $p \geq 1$, $L^p(\Omega \times [0,T])$ be the space of all real-valued, p-integrable functions on $\Omega \times [0,T]$. An $(\mathcal{F}_t)_{t \in [0,T]}$-progressively measurable stochastic process $(X_t)_{t \in [0,T]}$ in $L^p(\Omega \times [0,T])$, on the given probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T]}, \mathbb{P})$, is called *(strong) L^p-solution* to the SVE (4.2) if $\int_0^t (|(t-s)^{-\alpha}\mu(s, X_s)| + |(t-s)^{-\alpha}\sigma(s, X_s)|^2)\, \mathrm{d}s < \infty$ for all $t \in [0,T]$ and the integral equation (4.2) holds a.s. We call a strong L^1-solution often just *solution* to the SVE (4.2). We say *pathwise uniqueness* in $L^p(\Omega \times [0,T])$ holds for the SVE (4.2) if $\mathbb{P}(X_t = \tilde{X}_t, \forall t \in [0,T]) = 1$ for two L^p-solutions $(X_t)_{t \in [0,T]}$ and $(\tilde{X}_t)_{t \in [0,T]}$ to the SVE (4.2) defined on the same probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T]}, \mathbb{P})$. Moreover, we say there exists a *unique strong L^p-solution* $(X_t)_{t \in [0,T]}$ to the SVE (4.2) if $(X_t)_{t \in [0,T]}$ is a strong L^p-solution to the SVE (4.2) and pathwise uniqueness in L^p holds for the SVE (4.2).

We say $(X_t)_{t\in[0,T]}$ is β-Hölder continuous for $\beta \in (0,1]$ if there exists a modification of $(X_t)_{t\in[0,T]}$ with sample paths that are almost all β-Hölder continuous.

Note that the kernels $K_\mu(s,t) = K_\sigma(s,t) = (t-s)^{-\alpha}$ with $\alpha \in (0,1/2)$ fulfill the assumptions of Lemma 2.7 and Lemma 2.10 for every

$$\varepsilon \in \Big(0, \frac{1}{\alpha} - 2\Big) \tag{4.3}$$

with

$$\gamma = \frac{1}{2+\varepsilon} - \alpha. \tag{4.4}$$

This means that, to use the results of Lemma 2.7 and Lemma 2.10, we need to consider L^p-solutions with

$$p > \max\{\frac{1}{\gamma}, 1 + \frac{2}{\varepsilon}\} = \max\{\frac{2+\varepsilon}{1-2\alpha-\varepsilon\alpha}, 1+\frac{2}{\varepsilon}\}. \tag{4.5}$$

The maximum in (4.5) is attained for $\varepsilon^\star = \frac{1-2\alpha}{1+\alpha}$. Hence, inserting $\varepsilon^\star$ into (4.5), we consider in the following L^p-solutions and L^p-pathwise uniqueness for some

$$p > 3 + \frac{6\alpha}{1-2\alpha}. \tag{4.6}$$

The following theorem states that pathwise uniqueness for the stochastic Volterra equation (4.2) holds, which is the main result of the present work.

Theorem 4.3. *Suppose Assumption 4.1 and let p be given by* (4.6)*. Then, L^p-pathwise uniqueness holds for the stochastic Volterra equation* (4.2)*.*

The proof of Theorem 4.3 will be summarized in Section 4.2 and the subsequent Sections 4.3-4.6 provide the necessary auxiliary results. Relying on the pathwise uniqueness and the classical Yamada–Watanabe theorem, we get the existence of a unique strong solution.

Corollary 4.4. *Suppose Assumption 4.1 and let p be given by* (4.6)*. Then, there exists a unique strong L^p-solution to the stochastic Volterra equation* (4.2)*.*

Proof. The L^p-pathwise uniqueness is provided by Theorem 4.3. The existence of a strong L^p-solution follows by the existence of a weak L^p-solution to the stochastic Volterra equation (4.2), which is provided by Theorem 3.10, which is applicable since the kernel $(t-s)^{-\alpha}$, $\alpha \in [0, \frac{1}{2})$, fulfills Assumption 3.8, cf. Remark 3.12. Thanks to Yamada–Watanabe's theorem (see [YW71, Corollary 1], or [Kur14, Theorem 1.5] for a generalized version), the existence of a weak L^p-solution and pathwise L^p-uniqueness imply the existence of a unique strong L^p-solution. □

Furthermore, we obtain the following regularity properties of solutions to the SVE (4.2).

Lemma 4.5. *Suppose Assumption 4.1, and let $(X_t)_{t\in[0,T]}$ be a strong L^p-solution to the stochastic Volterra equation (4.2) with p given by (4.6). Then, $\sup_{t\in[0,T]}\mathbb{E}[|X_t|^q]<\infty$ for any $q\geq 1$ and the sample paths of $(X_t)_{t\in[0,T]}$ are β-Hölder continuous for any $\beta\in(0,\frac{1}{2}-\alpha)$.*

Proof. The statements follow by Lemma 2.7 and Lemma 2.10 since the kernel $(t-s)^{-\alpha}$ fulfills Assumption 3.8. □

For $k\in\mathbb{N}\cup\{\infty\}$, we write $C^k(\mathbb{R})$, $C^k(\mathbb{R}_+)$ and $C^k([0,T]\times\mathbb{R})$ for the spaces of continuous functions mapping from $\mathbb{R}$, $\mathbb{R}_+$ resp. $[0,T]\times\mathbb{R}$ to $\mathbb{R}$, that are k-times continuously differentiable. We use an index 0 to indicate compact support, e.g. $C_0^\infty(\mathbb{R})$ denotes the space of smooth functions with compact support on $\mathbb{R}$. The space of square integrable functions $f\colon\mathbb{R}\to\mathbb{R}$ is denoted by $L^2(\mathbb{R})$ and equipped with the usual scalar product $\langle\cdot,\cdot\rangle$. Moreover, a ball in $\mathbb{R}$ around x with radius $R>0$ is defined by $B(x,R):=\{y\in\mathbb{R}:|y-x|\leq R\}$ and we use the notation $A_\eta\lesssim B_\eta$ for a generic parameter η, meaning that $A_\eta\leq CB_\eta$ for some constant $C>0$ independent of η.

4.2 Proof of pathwise uniqueness

We prove Theorem 4.3 by generalizing the well-known techniques of Yamada–Watanabe (cf. [YW71, Theorem 1]) and the work of Mytnik and Salisbury [MS15]. One of the main challenges is the missing semimartingale property of a solution $(X_t)_{t\in[0,T]}$ to the SVE (4.2). Therefore, we transform (4.2) into a random field in Step 1, for which we can derive a semimartingale decomposition in (4.8). Then, we implement an approach in the spirit of Yamada–Watanabe in Step 2-5 and conclude the pathwise uniqueness by using a Grönwall inequality for weak singularities in Step 6.

Proof of Theorem 4.3. Suppose there are two strong L^p-solutions $(X_t^1)_{t\in[0,T]}$ and $(X_t^2)_{t\in[0,T]}$ to the stochastic Volterra equation (4.2).
Step 1: To induce a semimartingale structure, we introduce the random fields

$$X^i(t,x):=x_0(t)+\int_0^t p_{t-s}^\theta(x)\mu(s,X_s^i)\,\mathrm{d}s+\int_0^t p_{t-s}^\theta(x)\sigma(s,X_s^i)\,\mathrm{d}B_s, \tag{4.7}$$

for $t\in[0,T]$, $x\in\mathbb{R}$ and $i=1,2$, where the densities $p_t^\theta\colon\mathbb{R}\to\mathbb{R}$ and $\theta:=1/2-\alpha$ are defined in (4.11). By Proposition 4.17, we get that $X^i\in C([0,T]\times\mathbb{R})$ and

$$\begin{aligned}\int_{\mathbb{R}}X^i(t,x)\Phi_t(x)\,\mathrm{d}x=&\int_{\mathbb{R}}\left(x_0\Phi_0(x)+\int_0^t\Phi_s(x)\frac{\partial}{\partial s}x_0(s)\,\mathrm{d}s\right)\mathrm{d}x\\&+\int_0^t\int_{\mathbb{R}}X^i(s,x)\left(\Delta_\theta\Phi_s(x)+\frac{\partial}{\partial s}\Phi_s(x)\right)\mathrm{d}x\,\mathrm{d}s\\&+\int_0^t\mu(s,X^i(s,0))\Phi_s(0)\,\mathrm{d}s+\int_0^t\sigma(s,X^i(s,0))\Phi_s(0)\,\mathrm{d}B_s,\end{aligned}\tag{4.8}$$

for $t \in [0,T]$ and every $\Phi \in C_0^2([0,T] \times \mathbb{R})$, where the differential operator Δ_θ is defined in (4.16). Notice, due to (4.8), the stochastic process $t \mapsto \int_{\mathbb{R}} X^i(t,x)\Phi_t(x)\,\mathrm{d}x$ is a semimartingale and $X^i(t,0) = X^i_t$ for $t \in [0,T]$.

Step 2: We define suitable sequences $(\Phi_x^m) \subset C_0^2(\mathbb{R})$, for $x \in \mathbb{R}$, and $(\phi_n) \subset C^\infty(\mathbb{R})$ of test functions, see (4.36) and (4.31) for the precise definitions, such that

$$\Phi_x^m \to \delta_x \quad \text{as } m \to \infty, \quad \text{for every } x \in \mathbb{R}, \quad \text{and} \quad \phi_n \to |\cdot| \quad \text{as } n \to \infty.$$

Applying Proposition 4.18 (which is based on Itô's formula and (4.8)) and setting $\tilde{X}(t) := \tilde{X}(t,\cdot) := X^1(t,\cdot) - X^2(t,\cdot)$ for $t \in [0,T]$, we get

$$\begin{aligned}
\phi_n(\langle \tilde{X}(t), \Phi_x^m\rangle) &= \int_0^t \phi_n'(\langle \tilde{X}(s), \Phi_x^m\rangle)\langle \tilde{X}(s), \Delta_\theta \Phi_x^m\rangle\,\mathrm{d}s \\
&\quad + \int_0^t \phi_n'(\langle \tilde{X}(s), \Phi_x^m\rangle)\Phi_x^m(0)\big(\mu(s,X^1(s,0)) - \mu(s,X^2(s,0))\big)\,\mathrm{d}s \\
&\quad + \int_0^t \phi_n'(\langle \tilde{X}(s), \Phi_x^m\rangle)\Phi_x^m(0)\big(\sigma(s,X^1(s,0)) - \sigma(s,X^2(s,0))\big)\,\mathrm{d}B_s \\
&\quad + \frac{1}{2}\int_0^t \psi_n(|\langle \tilde{X}(s), \Phi_x^m\rangle|)\Phi_x^m(0)^2\big(\sigma(s,X^1(s,0)) - \sigma(s,X^2(s,0))\big)^2\,\mathrm{d}s,
\end{aligned}$$

where $\langle\cdot,\cdot\rangle$ denotes the scalar product on $L^2(\mathbb{R})$.

Step 3: To implement an approach in the spirit of Yamada–Watanabe, we need to introduce another suitable test function $\Psi \in C([0,T] \times \mathbb{R})$ (satisfying Assumption 4.19 below). Denoting by $\dot{\Psi} := \frac{\partial}{\partial s}\Psi$ the time derivative of Ψ, Proposition 4.20 leads to

$$\begin{aligned}
&\langle \phi_n(\langle \tilde{X}(t), \Phi_\cdot^m\rangle), \Psi_t\rangle \\
&\quad = \int_0^t \langle \phi_n'(\langle \tilde{X}(s), \Phi_\cdot^m\rangle)\langle \tilde{X}(s), \Delta_\theta \Phi_\cdot^m\rangle, \Psi_s\rangle\,\mathrm{d}s \\
&\qquad + \int_0^t \langle \phi_n'(\langle \tilde{X}(s), \Phi_\cdot^m\rangle)\Phi_\cdot^m(0), \Psi_s\rangle\big(\mu(s,X^1(s,0)) - \mu(s,X^2(s,0))\big)\,\mathrm{d}s \\
&\qquad + \int_0^t \langle \phi_n'(\langle \tilde{X}(s), \Phi_\cdot^m\rangle)\Phi_\cdot^m(0), \Psi_s\rangle\big(\sigma(s,X^1(s,0)) - \sigma(s,X^2(s,0))\big)\,\mathrm{d}B_s \\
&\qquad + \frac{1}{2}\int_0^t \langle \psi_n(|\langle \tilde{X}(s), \Phi_\cdot^m\rangle|)\Phi_\cdot^m(0)^2, \Psi_s\rangle\big(\sigma(s,X^1(s,0)) - \sigma(s,X^2(s,0))\big)\,\mathrm{d}s \\
&\qquad + \int_0^t \langle \phi_n(\langle \tilde{X}(s), \Phi_\cdot^m\rangle), \dot{\Psi}_s\rangle\,\mathrm{d}s.
\end{aligned}$$

Step 4: Using the stopping time $T_{\xi,K}$ defined in (4.83), taking expectations and sending

$n, m \to \infty$, Proposition 4.31 states that

$$\begin{aligned}
&\mathbb{E}\big[\langle|\langle\tilde{X}(t\wedge T_{\xi,K})|,\Psi_{t\wedge T_{\xi,K}}\rangle\big]\\
&\quad\lesssim \mathbb{E}\bigg[\int_0^{t\wedge T_{\xi,K}}\int_{\mathbb{R}}|\tilde{X}(s,x)|\Delta_\theta\Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s\bigg]\\
&\qquad+\int_0^{t\wedge T_{\xi,K}}\Psi_s(0)\mathbb{E}[|\tilde{X}(s,0)|]\,\mathrm{d}s+\mathbb{E}\bigg[\int_0^{t\wedge T_{\xi,K}}\int_{\mathbb{R}}|\tilde{X}(s,x)|\dot{\Psi}_s(x)\,\mathrm{d}x\,\mathrm{d}s\bigg].
\end{aligned}$$

Step 5: Since $T_{\xi,K} \to T$ as $K \to \infty$ a.s. by Corollary 4.28, applying Fatou's lemma yields

$$\begin{aligned}
\int_{\mathbb{R}}\mathbb{E}[|\tilde{X}(t,x)|]\Psi_t(x)\,\mathrm{d}x \lesssim \int_0^t\int_{\mathbb{R}}\mathbb{E}[|\tilde{X}(s,x)|]|\Delta_\theta\Psi_s(x)+\dot{\Psi}_s(x)|\,\mathrm{d}x\,\mathrm{d}s\\
+\int_0^t\Psi_s(0)\mathbb{E}[|\tilde{X}(s,0)|]\,\mathrm{d}s. \qquad (4.9)
\end{aligned}$$

Finally, we choose appropriate test functions $(\Psi_{N,M})_{N,M\in\mathbb{N}}$ (satisfying Assumption 4.19) to approximate the Dirac distribution around 0 with $\Psi_{N,M}(t,\cdot)$. Thus, choosing $\Psi_t(x) = \Psi_{N,M}(t,x)$ in (4.9) and sending $N, M \to \infty$ yields, by Proposition 4.34, that

$$\mathbb{E}[|\tilde{X}(t,0)|] \lesssim \int_0^t (t-s)^{-\alpha}\mathbb{E}[|\tilde{X}(s,0)|]\,\mathrm{d}s, \quad t\in[0,T].$$

Step 6: Due to $\alpha \in (0, \frac{1}{2})$, Grönwall's inequality for weak singularities (see e.g. [Kru14, Lemma A.2]) reveals

$$\mathbb{E}[|\tilde{X}(t,0)|] = 0, \quad t\in[0,T],$$

and therefore $X^1_t = X^2_t = 0$ a.s. By the continuity of X^1 and X^2 (see Lemma 4.5), we conclude the claimed pathwise uniqueness. □

4.3 Step 1: Transformation into an SPDE

Recall, in general, a solution $(X_t)_{t\in[0,T]}$ of the SVE (4.2) will not be a semimartingale due to the t-dependence of the kernel. In this section we will transform the SVE (4.2) into a stochastic partial differential equation (SPDE) in distributional form, see (4.8), which allows us to recover a semimartingale structure and, thus, to implement an approach in the spirit of Yamada–Watanabe.

To that end, we consider the evolution equation

$$\begin{aligned}
\frac{\partial u}{\partial t}(t,x) &= \Delta_\theta u(t,x), \quad t\in[0,T],\ x\in\mathbb{R},\\
u(0,x) &= \delta_0(x),
\end{aligned} \qquad (4.10)$$

where the differential operator Δ_θ is defined by

$$\Delta_\theta := \frac{2}{(2+\theta)^2}\frac{\partial}{\partial x}|x|^{-\theta}\frac{\partial}{\partial x}$$

for some constant $\theta > 0$. It can be seen that the following densities solve (4.10):

$$p_t^\theta(x) := c_\theta t^{-\frac{1}{2+\theta}} e^{-\frac{|x|^{2+\theta}}{2t}}, \quad t \in [0,T],\ x \in \mathbb{R}. \tag{4.11}$$

Since $\int_0^\infty p_t^\theta(x)\,\mathrm{d}x$ is independent of $t \in (0,T]$, one can verify that if we choose the constant

$$c_\theta := (2+\theta)2^{-\frac{1}{2+\theta}}\Gamma\left(\frac{1}{2+\theta}\right)^{-1}, \tag{4.12}$$

where Γ denotes the Gamma function, then p_t^θ defines a probability density on $\mathbb{R}_+$. The reason, why we consider (4.10), is that by the choice of $\theta > 0$ such that

$$\alpha = \frac{1}{2+\theta},$$

we get that for $x = 0$ the solution $p_{t-s}^\theta(0)$ represents the kernel in the SVE (4.2) up to a constant. Therefore, we obtain the following lemma.

Lemma 4.6. *Every strong L^p-solution $(X_t)_{t\in[0,T]}$ of the SVE (4.2) defines an a.s. continuous strong solution $(X(t,x))_{t\in[0,T],x\in\mathbb{R}}$ of*

$$\begin{aligned} X(t,x) = x_0(t) &+ \int_0^t p_{t-s}^\theta(x)\mu(s,X(s,0))\,\mathrm{d}s \\ &+ \int_0^t p_{t-s}^\theta(x)\sigma(s,X(s,0))\,\mathrm{d}B_s, \quad t \in [0,T], x \in \mathbb{R}, \end{aligned} \tag{4.13}$$

i.e., on the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, there is a random field $(X(t,x))_{t\in[0,T],x\in\mathbb{R}}$ such that $X \in C([0,T]\times\mathbb{R})$ a.s., $(X(t,x))_{t\in[0,T]}$ is $(\mathcal{F}_t)$-progressively measurable for $x \in \mathbb{R}$,

$$\int_0^t \left(|p_{t-s}^\theta(x)\mu(s,X(s,0))| + |p_{t-s}^\theta(x)\sigma(s,X(s,0))|^2\right)\mathrm{d}s < \infty$$

and (4.13) holds a.s. Conversely, every strong solution of (4.13) defines a strong solution of the stochastic Volterra equation (4.2).

Proof. First, we assume that there is a solution Y of the SVE

$$Y_t = x_0(t) + \int_0^t p_{t-s}^\theta(0)\mu(s,Y_s)\,\mathrm{d}s + \int_0^t p_{t-s}^\theta(0)\sigma(s,Y_s)\,\mathrm{d}B_s.$$

We define, for $t \in [0,T], x \in \mathbb{R}$,

$$X(t,x) := x_0(t) + \int_0^t p_{t-s}^\theta(x)\mu(s,Y_s)\,\mathrm{d}s + \int_0^t p_{t-s}^\theta(x)\sigma(s,Y_s)\,\mathrm{d}B_s.$$

Then, by obtaining $X(t,0) = Y_t$, X solves

$$X(t,x) = x_0(t) + \int_0^t p_{t-s}^\theta(x)\mu(s,X(s,0))\,\mathrm{d}s + \int_0^t p_{t-s}^\theta(x)\sigma(s,X(s,0))\,\mathrm{d}B_s.$$

By the adaptedness of the Itô integral and the Riemann–Stieltjes integral, $(X(t,x))_{t\in[0,T]}$ is $(\mathcal{F}_t)$-progressively measurable for every $x \in \mathbb{R}$. By the continuity of $p_t^\theta(x)$, $X(t,x)$ is continuous in x-direction. By the continuity of the initial condition x_0 and the integrals, it is also continuous in t-direction.
Conversely, if $X = (X(t,x))_{t\in[0,T],x\in\mathbb{R}}$ solves (4.13), $Y_t := X(t,0)$ is a solution of (4.2). □

Due to the transformation of the SVE (4.2) into the SPDE (4.13), we shall study continuous solutions $X \in C([0,T]\times\mathbb{R})$ of the SPDE (4.13) instead of solutions to the SVE (4.2) directly. The next goal is to derive a regularity result for solutions of the SPDE (4.13). For this purpose, we first investigate the densities p_t^θ. We introduce some auxiliary lemmas, which are helpful for a better understanding of the densities p_t^θ, and skip the dependence on θ by writing

$$p_t(x) := ct^{-\alpha}e^{-\frac{|x|^{\frac{1}{\alpha}}}{2t}} \quad \text{for a fixed } \alpha \in (0,1/2).$$

Lemma 4.7. *For any $x,y \in \mathbb{R}$, $t \in [0,T]$ and $\beta \in [0,1]$, one has*

$$|p_t(x) - p_t(y)| \lesssim t^{-\alpha}\left(\frac{|x-y|}{t}\right)^\beta \max(|x|,|y|)^{(\frac{1}{\alpha}-1)\beta}.$$

Proof. First, let us fix $t \in [0,T]$ and consider the function $x \mapsto e^{-\frac{|x|^{1/\alpha}}{2t}}$. By applying the mean value theorem and assuming w.l.o.g. $|y| < |x|$, we obtain, for some $z \in [|y|,|x|]$,

$$\frac{e^{-\frac{|x|^{\frac{1}{\alpha}}}{2t}} - e^{-\frac{|y|^{\frac{1}{\alpha}}}{2t}}}{|x|-|y|} = -\frac{z^{\frac{1}{\alpha}-1}}{2t\alpha}e^{-\frac{z^{1/\alpha}}{2t}},$$

which reveals that

$$\left|e^{-\frac{|x|^{\frac{1}{\alpha}}}{2t}} - e^{-\frac{|y|^{\frac{1}{\alpha}}}{2t}}\right| \le \frac{|x-y|}{2t\alpha}|x|^{\frac{1}{\alpha}-1}. \tag{4.14}$$

Using inequality (4.14) and $\beta \in [0,1]$, we bound

$$|p_t(x) - p_t(y)| \lesssim t^{-\alpha}\left|e^{-\frac{|x|^{\frac{1}{\alpha}}}{2t}} - e^{-\frac{|y|^{\frac{1}{\alpha}}}{2t}}\right|^\beta \lesssim t^{-\alpha}\left(\frac{|x-y|}{t}\right)^\beta \max(|x|,|y|)^{(\frac{1}{\alpha}-1)\beta}.$$

□

Corollary 4.8. *For any $x,y \in [-1,1]$, $t \in [0,T]$ and $\beta \in (0,1-\alpha)$, one has*

$$\int_0^t |p_s(x) - p_s(y)|\,\mathrm{d}s \lesssim |x-y|^\beta.$$

Proof. By Lemma 4.7, we see that

$$\begin{aligned}\int_0^t |p_s(x) - p_s(y)|\,\mathrm{d}s &\lesssim \int_0^t s^{-\alpha}\left(\frac{|x-y|}{s}\right)^\beta \max(|x|,|y|)^{(\frac{1}{\alpha}-1)\beta}\,\mathrm{d}s \\ &\lesssim |x-y|^\beta \int_0^t s^{-\alpha-\beta}\,\mathrm{d}s \lesssim |x-y|^\beta.\end{aligned}$$

□

Lemma 4.9. *For any $0 < t < t' \leq T$ and $x \in \mathbb{R}$, one has*

$$\int_0^t (p_{t'-s}(x) - p_{t-s}(x))^2 \, \mathrm{d}s \lesssim |t'-t|^{1-2\alpha}.$$

Proof. We assume w.l.o.g. that $t' - t \leq t$ and use the linearity of the integral together with $|e^{-x}| \leq 1$ for non-negative x to get

$$\begin{aligned}
\int_0^t |p_{t'-s}(x) - p_{t-s}(x)|^2 \, \mathrm{d}s &\lesssim \int_{t-|t'-t|}^{t} |(t'-s)^{-\alpha} - (t-s)^{-\alpha}|^2 \, \mathrm{d}s \\
&\quad + \int_0^{t-|t'-t|} |p_{t'-s}(x) - p_{t-s}(x)|^2 \, \mathrm{d}s \\
&\lesssim \int_{t-|t'-t|}^{t} (t-s)^{-2\alpha} \, \mathrm{d}s \\
&\quad + \int_0^{t-|t'-t|} |(t-s)^{-\alpha} - (t'-s)^{-\alpha}|^2 e^{-\frac{|x|^{\frac{1}{\alpha}}}{2(t-s)}} \, \mathrm{d}s \\
&\quad + \int_0^{t-|t'-t|} (t'-s)^{-2\alpha} \left| e^{-\frac{|x|^{\frac{1}{\alpha}}}{2(t-s)}} - e^{-\frac{|x|^{\frac{1}{\alpha}}}{2(t'-s)}} \right| \mathrm{d}s \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

For I_1, we directly compute

$$I_1 = \left[\frac{-(t-s)^{1-2\alpha}}{1-2\alpha} \right]_{t-|t'-t|}^{t} \lesssim |t'-t|^{1-2\alpha}.$$

For I_2, we use $|a-b|^2 \leq a^2 - b^2$ for $a > b$ to bound

$$\begin{aligned}
I_2 &\leq \int_0^{t-|t'-t|} (t-s)^{-2\alpha} \, \mathrm{d}s - \int_0^{t-|t'-t|} (t'-s)^{-2\alpha} \, \mathrm{d}s \\
&= \left[\frac{-(t-s)^{1-2\alpha}}{1-2\alpha} \right]_0^{t-|t'-t|} - \left[\frac{-(t'-s)^{1-2\alpha}}{1-2\alpha} \right]_0^{t-|t'-t|} \\
&\lesssim |t'-t|^{1-2\alpha}.
\end{aligned}$$

For I_3, we use the mean value theorem for the function $t \mapsto e^{-\frac{|x|^{\frac{1}{\alpha}}}{2(t-s)}}$, similarly as we did in (4.14), to get the inequality

$$\left| e^{-\frac{|x|^{\frac{1}{\alpha}}}{2(t-s)}} - e^{-\frac{|x|^{\frac{1}{\alpha}}}{2(t'-s)}} \right| \leq (t'-t) \frac{|x|^{\frac{1}{\alpha}}}{2(t-s)^2} e^{-\frac{|x|^{\frac{1}{\alpha}}}{2(t'-s)}}.$$

Using this and the inequality $e^{-x} \leq x^{-1}$ for all $x \geq 0$, such as $\frac{t'-t}{t-s} \leq 1$ and $\frac{t'-s}{t-s} \leq \frac{2(t-s)}{t-s} =$

2 due to $s \leq t - |t' - t|$, we get

$$
\begin{aligned}
I_3 &\leq (t'-t) \int_0^{t-|t'-t|} (t-s)^{-2\alpha} \left(\frac{|x|^{\frac{1}{\alpha}}}{2(t-s)^2} e^{-\frac{|x|^{\frac{1}{\alpha}}}{2(t'-s)}} \right) \mathrm{d}s \\
&\lesssim \int_0^{t-|t'-t|} (t-s)^{-2\alpha} \frac{(t'-t)(t'-s)}{(t-s)^2} \,\mathrm{d}s \\
&\lesssim \int_0^{t-|t'-t|} (t-s)^{-2\alpha} \,\mathrm{d}s \lesssim |t'-t|^{1-2\alpha},
\end{aligned}
$$

which yields the statement. □

Lemma 4.10. *For any $x, y \in [-1,1]$, $t \in [0,T]$ and $\beta \in (0, \frac{1}{2} - \alpha)$, one has*

$$
\int_0^t (p_{t-s}(x) - p_{t-s}(y))^2 \,\mathrm{d}s \lesssim \max\big(|x|,|y|\big)^{(\frac{1}{\alpha}-1)2\beta} |x-y|^{1-2\alpha}.
$$

Proof. W.l.o.g. we may assume $t \geq |x-y|$ and split the integral into

$$
\begin{aligned}
\int_0^t (p_{t-s}(x) - p_{t-s}(y))^2 \,\mathrm{d}s &\leq \int_0^{t-|x-y|} (p_{t-s}(x) - p_{t-s}(y))^2 \,\mathrm{d}s \\
&\quad + \int_{t-|x-y|}^t (p_{t-s}(x) - p_{t-s}(y))^2 \,\mathrm{d}s \\
&=: I_1 + I_2.
\end{aligned}
$$

For I_1, we apply Lemma 4.7 with $\beta = 1$ to get

$$
\begin{aligned}
I_1 &\lesssim \max(|x|,|y|)^{(\frac{1}{\alpha}-1)2} \int_0^{t-|x-y|} |x-y|^2 (t-s)^{-2\alpha-2} \,\mathrm{d}s \\
&= \max(|x|,|y|)^{(\frac{1}{\alpha}-1)2} |x-y|^2 \left[\frac{-(t-s)^{1-2\alpha-2}}{1-2\alpha-2} \right]_0^{t-|x-y|} \\
&\lesssim \max(|x|,|y|)^{(\frac{1}{\alpha}-1)2} |x-y|^2 \big(t^{-2\alpha-1} + |x-y|^{-2\alpha-1} \big) \\
&\lesssim \max(|x|,|y|)^{(\frac{1}{\alpha}-1)2\beta} |x-y|^{1-2\alpha}
\end{aligned}
$$

with $t \geq |x-y|$.
For I_2, Lemma 4.7 again, but with $\beta \in (0, 1/2 - \alpha)$ such that $2\alpha + 2\beta < 1$, yields

$$
\begin{aligned}
I_2 &\lesssim \max(|x|,|y|)^{(\frac{1}{\alpha}-1)2\beta} |x-y|^{2\beta} \int_{t-|x-y|}^t (t-s)^{-2\alpha-2\beta} \,\mathrm{d}s \\
&\lesssim \max(|x|,|y|)^{(\frac{1}{\alpha}-1)2\beta} |x-y|^{2\beta} \left[\frac{-(t-s)^{1-2\alpha-2\beta}}{1-2\alpha-2\beta} \right]_{t-|x-y|}^t \\
&\lesssim \max(|x|,|y|)^{(\frac{1}{\alpha}-1)2\beta} |x-y|^{2\beta} |x-y|^{1-2\alpha-2\beta} \\
&\lesssim \max(|x|,|y|)^{(\frac{1}{\alpha}-1)2\beta} |x-y|^{1-2\alpha}.
\end{aligned}
$$

□

With these auxiliary results at hand, we are ready to prove the following regularity result for solutions of the SPDE (4.13).

Proposition 4.11. *Suppose Assumption 4.1 and let* $X \in C([0,T]\times\mathbb{R})$ *be a strong solution of the SPDE* (4.13).

(i) For any $p \in (0,\infty)$*, one has*

$$\sup_{t\in[0,T]} \sup_{x\in\mathbb{R}} \mathbb{E}[|X(t,x)|^p] < \infty.$$

(ii) We define the random field $(Z(t,x))_{t\in[0,T],x\in\mathbb{R}}$ *by*

$$\begin{aligned} Z(t,x) &:= X(t,x) - x_0(t) \\ &= \int_0^t p^\theta_{t-s}(x)\mu(s,X(s,0))\,\mathrm{d}s + \int_0^t p^\theta_{t-s}(x)\sigma(s,X_s(s,0))\,\mathrm{d}B_s. \end{aligned}$$

For any $0 \le t,t' \le T$*,* $|x|,|y| \le 1$ *and* $p \in [2,\infty)$*, we get*

$$\mathbb{E}\big[|Z(t,x) - Z(t',y)|^p\big] \lesssim |t'-t|^{(\frac{1}{2}-\alpha)p} + |x-y|^{(\frac{1}{2}-\alpha)p}.$$

Proof. (i) Let us assume that $p \ge 2$. For $p \in (0,2)$, the statement then follows by the orderedness of the L^p-spaces. From Lemma 4.6 we know that $Y_t := X(t,0)$ is a solution of the SVE (4.2) and from Lemma 4.5 we know that its moment are finite. Thus, applying Hölder's and the Burkholder–Davis–Gundy inequality, the linear growth condition on μ and σ from Assumption 4.1, such as Lemma 4.5, we get

$$\begin{aligned} \mathbb{E}[|X(t,x)|^p] &\lesssim 1 + \mathbb{E}\bigg[\bigg|\int_0^t p^\theta_{t-s}(x)\mu(s,Y_s)\,\mathrm{d}s\bigg|^p\bigg] + \mathbb{E}\bigg[\bigg|\int_0^t p^\theta_{t-s}(x)\sigma(s,Y_s)\,\mathrm{d}B_s\bigg|^p\bigg] \\ &\lesssim 1 + \bigg(\int_0^t \big(p^\theta_{t-s}(x)\big)^2\,\mathrm{d}s\bigg)^{\frac{p}{2}} + \bigg(\int_0^t \big(p^\theta_{t-s}(x)\big)^2\,\mathrm{d}s\bigg)^{\frac{p}{2}} \\ &\lesssim 1 + \bigg(\int_0^t c_\theta^2 (t-s)^{-2\alpha} e^{-2\frac{|x|^{2+\theta}}{2(t-s)}}\,\mathrm{d}s\bigg)^{\frac{p}{2}} \\ &\lesssim 1 + \bigg(\int_0^t (t-s)^{-2\alpha}\,\mathrm{d}s\bigg)^{\frac{p}{2}} < \infty. \end{aligned}$$

(ii) With

$$Z(t,x) = \int_0^t p^\theta_{t-s}(x)\mu(s,X(s,0))\,\mathrm{d}s + \int_0^t p^\theta_{t-s}(x)\sigma(s,X_s(s,0))\,\mathrm{d}B_s$$

and by splitting the integrals, we get

$$\begin{aligned}
&|Z(t',x)-Z(t,y)| \\
&= \int_0^t \big(p^\theta_{t'-s}(x)-p^\theta_{t-s}(x)\big)\mu(s,X(s,0))\,\mathrm{d}s + \int_0^t \big(p^\theta_{t-s}(x)-p^\theta_{t-s}(y)\big)\mu(s,X(s,0))\,\mathrm{d}s \\
&\quad + \int_t^{t'} p^\theta_{t'-s}(x)\mu(s,X(s,0))\,\mathrm{d}s \\
&\quad + \int_0^t \big(p^\theta_{t'-s}(x)-p^\theta_{t-s}(x)\big)\sigma(s,X(s,0))\,\mathrm{d}B_s + \int_0^t \big(p^\theta_{t-s}(x)-p^\theta_{t-s}(y)\big)\sigma(s,X(s,0))\,\mathrm{d}B_s \\
&\quad + \int_t^{t'} p^\theta_{t'-s}(x)\sigma(s,X(s,0))\,\mathrm{d}B_s \\
&=: D_1+D_2+D_3+S_1+S_2+S_3.
\end{aligned}$$

We use Lemma 4.9, Lemma 4.10, Hölder's and the Burkholder–Davis–Gundy inequality, Fubini's theorem as well as (i) to get the following estimates:

$$\begin{aligned}
&\mathbb{E}[|D_1|^p] \le \Big(\int_0^t \big(p^\theta_{t'-s}(x)-p^\theta_{t-s}(x)\big)^2\,\mathrm{d}s\Big)^{\frac{p}{2}} \lesssim |t'-t|^{p(\frac{1}{2}-\alpha)}, \\
&\mathbb{E}[|S_1|^p] \le \Big(\int_0^t \big(p^\theta_{t'-s}(x)-p^\theta_{t-s}(x)\big)^2\,\mathrm{d}s\Big)^{\frac{p}{2}} \lesssim |t'-t|^{p(\frac{1}{2}-\alpha)}, \\
&\mathbb{E}[|D_2|^p] \le \Big(\int_0^t \big(p^\theta_{t-s}(x)-p^\theta_{t-s}(y)\big)^2\,\mathrm{d}s\Big)^{\frac{p}{2}} \lesssim |x-y|^{p(\frac{1}{2}-\alpha)}, \\
&\mathbb{E}[|S_2|^p] \lesssim |x-y|^{p(\frac{1}{2}-\alpha)}, \\
&\mathbb{E}[|D_3|^p] \le \Big(\int_t^{t'} p^\theta_{t'-s}(x)^2\,\mathrm{d}s\Big)^{\frac{p}{2}} \lesssim \Big(\int_t^{t'} (t'-s)^{-2\alpha}\,\mathrm{d}s\Big)^{\frac{p}{2}} \lesssim |t'-t|^{p(\frac{1}{2}-\alpha)}, \\
&\mathbb{E}[|S_3|^p] \lesssim |t'-t|^{p(\frac{1}{2}-\alpha)}.
\end{aligned}$$

Hence, we obtain the desired statement. □

4.3.1 Transformation to an SPDE in distributional form

The next aim is to transform the SPDE (4.13) into an SPDE in distributional form. To that end, we want to derive explicit formulas for the fundamental solution $p^\theta\colon [0,T]\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ of (4.10), in the sense that for any $g\colon \mathbb{R}\to\mathbb{R}$, $\big(\int_{\mathbb{R}} p^\theta_t(x,y)g(y)\,\mathrm{d}y\big)_{t\in[0,T],x\in\mathbb{R}}$ is a solution of (4.10) with initial condition g instead of δ_0.
The semigroup $(S_t)_{t\in[0,T]}$ generated by Δ_θ is defined by $S_t\colon C_0^\infty(\mathbb{R})\to C_0^\infty(\mathbb{R})$ via

$$S_t\phi(x) := \int_{\mathbb{R}} p^\theta_t(x,y)\phi(y)\,\mathrm{d}y, \quad \phi\in C_0^\infty(\mathbb{R}). \tag{4.15}$$

First, we only consider $x \in \mathbb{R}_+$ and, hence, search for the fundamental solutions of

$$\frac{\partial u}{\partial t}(t,x) = \Delta_\theta u(t,x), \quad u(0,x) = \delta_0(x), \quad t \in [0,T], x \in \mathbb{R}_+, \tag{4.16}$$

where $\Delta_\theta := \frac{2}{(2+\theta^2)} \frac{\partial}{\partial x} |x|^{-\theta} \frac{\partial}{\partial x}$. We will later, more precisely in Proposition 4.14 below, allow for $[0,T] \times \mathbb{R}$ as the domain of (4.16).

As preparations we need:

- A squared Bessel process $Z_t \geq 0$ of dimension $n \in \mathbb{R}$ is given by the stochastic differential equation

$$\mathrm{d}Z_t = 2\sqrt{Z_t}\, \mathrm{d}B_t + n\, \mathrm{d}t, \quad t \in [0,T].$$

- The generator of a squared Bessel process of dimension n is given by

$$(Lf)(x) = n\frac{\partial}{\partial x} f(x) + 2x \frac{\partial^2}{\partial x^2} f(x), \quad x \in \mathbb{R}_+, \tag{4.17}$$

  for $f \in C_0^\infty(\mathbb{R}_+)$, see [RY99, page 443].

- The semigroup $(S_t)_{t\in[0,T]}$, defined in (4.15), fulfills

$$\frac{\partial}{\partial t}(S_t f) = \Delta_\theta (S_t f) \tag{4.18}$$

  for $f \in C_0^\infty(\mathbb{R}_+)$, since p^θ is the fundamental solution of (4.16).

- Denote by $(\xi_t)_{t\in[0,T]}$ the Markov process that is generated by the semigroup S_t, that is, it has the transition densities $(p_t^\theta)_{t\in[0,T]}$. We define the semigroup $(T_t)_{t\in[0,T]}$ by

$$(T_t g)(x) := (S_t(g \circ \tilde{f}))(x) = \mathbb{E}_x[g(\tilde{f}(\xi_t))]$$

  for the fixed function $\tilde{f}(x) := |x|^{2+\theta}$ and for $g \in C_0^\infty(\mathbb{R}_+)$.

Lemma 4.12. *The process $(|\xi_t|^{2+\theta})_{t\in[0,T]}$ is a squared Bessel process of dimension $\frac{2}{2+\theta} < 1$.*

Proof. We show that the generator of $\tilde{f}(\xi_t)$ is the same as the one of the squared Bessel process in (4.17) with dimension $\frac{2}{2+\theta}$. Therefore, we use the semigroup T_t and denote by G its generator. For appropriate functions g we get, by the definition of the generator and by (4.18),

$$\begin{aligned}(Gg)(x) &= \frac{\partial}{\partial t}(T_t g)|_{t\to 0}(x) = \frac{\partial}{\partial t}(S_t(g\circ\tilde{f}))|_{t\to 0}(x) = \Delta_\theta S_0(g\circ\tilde{f})(x) \\ &= \Delta_\theta(g\circ\tilde{f})(x).\end{aligned}$$

Note that the set $\{t \in [0,T]\colon \xi_t = 0\}$ has Lebesque measure zero. Therefore, we can explicitly calculate, for $x \neq 0$,

$$\begin{aligned}
(Gg)(x) &= \frac{2}{(2+\theta)^2}\frac{\partial}{\partial x}\left(|x|^{-\theta}\frac{\partial}{\partial x}(g(|x|^{2+\theta}))\right)\\
&= \frac{2}{(2+\theta)^2}\frac{\partial}{\partial x}(|x|^{-\theta}g'(|x|^{2+\theta})(2+\theta)|x|^{1+\theta}\operatorname{sgn}(x))\\
&= \frac{2}{(2+\theta)}\frac{\partial}{\partial x}(xg'(|x|^{2+\theta}))\\
&= \frac{2}{(2+\theta)}(g'(|x|^{2+\theta}) + xg''(|x|^{2+\theta})(2+\theta)|x|^{1+\theta}\operatorname{sgn}(x))\\
&= \frac{2}{(2+\theta)}\frac{\partial g}{\partial x}(|x|^{2+\theta}) + 2|x|^{2+\theta}\frac{\partial g^2}{\partial x^2}(|x|^{2+\theta})\\
&= (Lg)(u),
\end{aligned}$$

where L is the generator of a squared Bessel process of dimension $\frac{2}{2+\theta}$ and $u := |x|^{2+\theta}$. $\square$

Next, we want to find explicit formulas for the transition densities of $(\xi_t)_{t\in[0,T]}$. Note that the transition densities for the squared Bessel process of dimension n are for $t > 0$ and $y > 0$ given by (see e.g. [RY99, Corollary XI.1.4])

$$q_t^n(x,y) = \frac{1}{2t}\left(\frac{y}{x}\right)^{\frac{\nu}{2}} e^{-\frac{x+y}{2t}} I_\nu\left(\frac{\sqrt{xy}}{t}\right) \quad \text{for } x > 0 \quad \text{and} \tag{4.19}$$

$$q_t^n(0,y) = 2^{-\nu}t^{-(\nu+1)}\Gamma(\nu+1)^{-1}y^{2\nu+1}e^{-\frac{y^2}{2t}}, \tag{4.20}$$

where $\nu := \frac{n}{2} - 1$ denotes the index of the Bessel process and I_ν is the modified Bessel function that is given by

$$I_\nu(x) := \sum_{k=0}^{\infty} \frac{(x/2)^{2k+\nu}}{k!\Gamma(\nu+k+1)} \tag{4.21}$$

for $\nu \geq -1$ and $x > 0$.

Lemma 4.13. *The transition densities of the Markov process $(|\xi_t|)_{t\in[0,T]}$ are, for $t > 0$, given by*

$$p_t^\theta(x,y) = \frac{(2+\theta)}{2t}|xy|^{\frac{(1+\theta)}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}} I_\nu\left(\frac{|xy|^{1+\frac{\theta}{2}}}{t}\right) \quad \text{for } x,y > 0, \tag{4.22}$$

and for $x = 0$, $y > 0$ with $p_t^\theta(0,y) = p_t^\theta(y)$ defined in (4.11).

Proof. Denote for fixed $\theta > 0$ by q_t the density function of the Bessel process $|\xi_t|^{2+\theta}$ with dimension $\frac{2}{2+\theta}$, that is given by (4.19) with $\nu = \frac{1}{2+\theta} - 1$.
Now, by noting that, for all $x,t,s > 0$ and Borel sets $A \subset B(\mathbb{R}_+)$,

$$\mathbb{E}\Big[\mathbb{1}_A(|\xi_{t+s}|^{2+\theta})|\xi_{t+s}|^{2+\theta}\,\Big|\,|\xi_t|^{2+\theta} = x\Big] = \mathbb{E}\Big[\mathbb{1}_A(|\xi_{t+s}|^{2+\theta})|\xi_{t+s}|^{2+\theta}\,\Big|\,|\xi_t| = x^{\frac{1}{2+\theta}}\Big]$$

holds, we get with the notation $B := \{b \in \mathbb{R}_+ : b^{2+\theta} \in A\}$ the relation

$$
\begin{aligned}
\int_A q_t(x,y) y \,\mathrm{d}y &= \int_B p_t^\theta\Big(x^{\frac{1}{2+\theta}}, y\Big) y^{2+\theta} \,\mathrm{d}y \\
&= \frac{1}{2+\theta} \int_A p_t^\theta\Big(x^{\frac{1}{2+\theta}}, z^{\frac{1}{2+\theta}}\Big) z \, z^{\frac{1}{2+\theta}-1} \,\mathrm{d}z \\
&= \frac{1}{2+\theta} \int_A p_t^\theta\Big(x^{\frac{1}{2+\theta}}, y^{\frac{1}{2+\theta}}\Big) y^{\frac{1}{2+\theta}-1} y \,\mathrm{d}y, \qquad (4.23)
\end{aligned}
$$

where we substituted $z := y^{2+\theta}$ and thus $\mathrm{d}y = \frac{1}{2+\theta} z^{\frac{1}{2+\theta}-1} \,\mathrm{d}z$. Since (4.23) must hold for all Borel sets A, we can compare both sides of the equation to see with the notation

$$
\hat{x} := x^{\frac{1}{2+\theta}} \quad \text{and} \quad \hat{y} := y^{\frac{1}{2+\theta}}
$$

that, with $\nu = \frac{1}{2+\theta} - 1 = -(\frac{1+\theta}{2+\theta})$,

$$
\begin{aligned}
p_t^\theta(\hat{x}, \hat{y}) &= (2+\theta) q_t\Big(\hat{x}^{2+\theta}, \hat{y}^{2+\theta}\Big) y^{1-\frac{1}{2+\theta}} \\
&= \frac{(2+\theta)}{2t} \left|\frac{\hat{y}}{\hat{x}}\right|^{\frac{(2+\theta)\nu}{2}} e^{-\frac{|\hat{x}|^{2+\theta} + |\hat{y}|^{2+\theta}}{2t}} I_\nu\left(\frac{|\hat{x}\hat{y}|^{1+\frac{\theta}{2}}}{t}\right) |\hat{y}|^{1+\theta} \\
&= \frac{(2+\theta)}{2t} \left|\frac{\hat{y}}{\hat{x}}\right|^{-\frac{(1+\theta)}{2}} e^{-\frac{|\hat{x}|^{2+\theta} + |\hat{y}|^{2+\theta}}{2t}} I_\nu\left(\frac{|\hat{x}\hat{y}|^{1+\frac{\theta}{2}}}{t}\right) |\hat{y}|^{1+\theta} \\
&= \frac{(2+\theta)}{2t} |\hat{x}\hat{y}|^{\frac{(1+\theta)}{2}} e^{-\frac{|\hat{x}|^{2+\theta} + |\hat{y}|^{2+\theta}}{2t}} I_\nu\left(\frac{|\hat{x}\hat{y}|^{1+\frac{\theta}{2}}}{t}\right).
\end{aligned}
$$

By a very similar calculation, (4.20) can be used to derive (4.11) in the case of $x = 0$:

$$
\begin{aligned}
\int_B q_t^\theta(0,y) y \,\mathrm{d}y &= \int_A q_t^\theta(0, z^{1+\theta/2}) z^{\theta/2} (1+\theta/2) z^{1+\theta/2} \,\mathrm{d}z \\
&= (1+\theta/2) 2^{\frac{1+\theta}{2+\theta}} \Gamma(\nu+1)^{-1} \int_A t^{-(\nu+1)} z^{-\theta/2} e^{-\frac{|z|^{2+\theta}}{2t}} z^{\theta/2} z^{1+\theta/2} \,\mathrm{d}z \\
&= (2+\theta) 2^{-\frac{1}{2+\theta}} \Gamma\left(\frac{1}{2+\theta}\right)^{-1} \int_A t^{-\frac{1}{2+\theta}} e^{-\frac{|z|^{2+\theta}}{2t}} z^{1+\theta/2} \,\mathrm{d}z \\
&= \int_A p_t^\theta(0,z) z^{1+\theta/2} \,\mathrm{d}z
\end{aligned}
$$

with $p_t^\theta(0,z) = p_t^\theta(z)$ as in (4.11) and choosing c_θ as in (4.12). □

We can define the fundamental solution p^θ also for negative x, y to extend the domain of $\Delta_\theta u = \frac{\partial}{\partial t} u$ to the domain $[0,T] \times \mathbb{R}$.

Proposition 4.14. *Consider* (4.10) *on the domain* $[0,T] \times \mathbb{R}$. *Then, if we define* $p_t^\theta(x,y)$ *as in* (4.22) *also for* $x, y < 0$, *then* p^θ *is still a fundamental solution.*

Proof. Assume $p_t^\theta(x,y)$ to be defined as in (4.22) also for negative x, y. Due to the absolute values in p_t^θ, it is clear that

$$\frac{\partial}{\partial t} p_t^\theta(x,y) = \frac{\partial}{\partial t} p_t^\theta(-x,y) = \frac{\partial}{\partial t} p_t^\theta(-x,-y) = \frac{\partial}{\partial t} p_t^\theta(x,-y).$$

Similarly, due to the 2nd derivative, we get in the space dimension

$$\frac{\partial}{\partial x}|x|^{-\theta}\frac{\partial}{\partial x}g(|x|) = \frac{\partial}{\partial x}|x|^{-\theta}\operatorname{sgn}(x)g'(|x|) = -\theta|x|^{-\theta-1}g'(|x|) + |x|^{-\theta}g''(|x|)$$

for all sufficiently smooth functions g and thus, where Δ_θ acts w.r.t. x,

$$\begin{aligned}\Delta_\theta \int_{\mathbb{R}} p_t^\theta(-x,y)g(y)\,\mathrm{d}y &= \int_{\mathbb{R}} \Delta_\theta p_t^\theta(x,y)g(y)\,\mathrm{d}y \\ &= \frac{\partial}{\partial t}\int_{\mathbb{R}} p_t^\theta(x,y)g(y)\,\mathrm{d}y = \frac{\partial}{\partial t}\int_{\mathbb{R}} p_t^\theta(-x,y)g(y)\,\mathrm{d}y,\end{aligned}$$

and analogue for the second variable. Hence, $p^\theta(x,y)$ also defines a fundamental solution of (4.10) if we allow negative x. □

Let us introduce a partial integration formula for the operator Δ_θ.

Lemma 4.15. *For $\Delta_\theta = \frac{2}{(2+\theta)^2}\frac{\partial}{\partial x}|x|^{-\theta}\frac{\partial}{\partial x}$, the partial integration formula*

$$\int_{\mathbb{R}} p_t(x,y)\Delta_\theta\phi(x)\,\mathrm{d}x = \int_{\mathbb{R}} \big(\Delta_\theta p_t(x,y)\big)\phi(x)\,\mathrm{d}x, \quad t\in[0,T], y\in\mathbb{R},$$

holds for any $\phi \in C_0^2(\mathbb{R})$.

Proof. Denoting $\phi_{2,t}(x) := |x|^{-\theta}\frac{\partial}{\partial x}\phi(x)$, then $\phi_{2,t}$ has also compact support and we get, by the classical partial integration formula,

$$\begin{aligned}&\int_{\mathbb{R}} p_t(x,y)\frac{\partial}{\partial x}|x|^{-\theta}\frac{\partial}{\partial x}\phi(x)\,\mathrm{d}x = \int_{\mathbb{R}} p_t(x,y)\frac{\partial}{\partial x}\phi_{2,t}(x)\,\mathrm{d}x \\ &\quad = -\int_{\mathbb{R}} \frac{\partial}{\partial x}p_t(x,y)\phi_{2,t}(x)\,\mathrm{d}x = -\int_{\mathbb{R}}\left(\frac{\partial}{\partial x}p_t(x,y)\right)|x|^{-\theta}\frac{\partial}{\partial x}\phi(x)\,\mathrm{d}x.\end{aligned}$$

Then, again by partial integration, we get, as claimed,

$$\int_{\mathbb{R}} p_t(x,y)\frac{\partial}{\partial x}|x|^{-\theta}\frac{\partial}{\partial x}\phi(x)\,\mathrm{d}x = \int_{\mathbb{R}}\frac{\partial}{\partial x}\left(\left(\frac{\partial}{\partial x}p_t(x,y)\right)|x|^{-\theta}\right)\phi(x)\,\mathrm{d}x.$$

□

With these auxiliary results at hand, we are in a position to do the transformation into an SPDE in distributional form. We consider test functions $\Phi \in C_0^2([0,T]\times\mathbb{R})$, to which we can apply the operator Δ_θ such that

$$\Delta_\theta\Phi_t(x) = \frac{\partial}{\partial x}|x|^{-\theta}\frac{\partial}{\partial x}\Phi_t(x)$$

is well-defined for all $t\in[0,T]$ and $x \in \mathbb{R}\setminus\{0\}$.

Lemma 4.16. *Every strong solution* $(X(t,x))_{t\in[0,T],x\in\mathbb{R}}$ *of* (4.13) *is a strong solution to the following SPDE in distributional form*

$$\begin{aligned}
&\int_{\mathbb{R}} X(t,x)\Phi_t(x)\,\mathrm{d}x \\
&\quad= \int_{\mathbb{R}} \left(x_0\Phi_0(x) + \int_0^t \Phi_s(x)\frac{\partial}{\partial s}x_0(s)\,\mathrm{d}s \right)\mathrm{d}x \\
&\quad\quad+ \int_0^t \int_{\mathbb{R}} X(s,x)\Big(\Delta_\theta\Phi_s(x) + \frac{\partial}{\partial s}\Phi_s(x)\Big)\,\mathrm{d}x\,\mathrm{d}s \\
&\quad\quad+ \int_0^t \mu(s,X(s,0))\Phi_s(0)\,\mathrm{d}s + \int_0^t \sigma(s,X(s,0))\Phi_s(0)\,\mathrm{d}B_s, \quad t\in[0,T],
\end{aligned} \tag{4.24}$$

for every test function $\Phi \in C_0^2([0,T]\times\mathbb{R})$.

Proof. Let X be a solution to (4.13) and Φ be as in the statement. We first observe that

$$\begin{aligned}
&\int_0^t \langle X(s,\cdot), \Delta_\theta\Phi_s\rangle\,\mathrm{d}s \\
&\quad= \int_0^t\int_{\mathbb{R}} x_0(s)\Delta_\theta\Phi_s(x)\,\mathrm{d}x\,\mathrm{d}s + \int_0^t\int_{\mathbb{R}}\int_0^s p^\theta_{s-u}(x)\sigma(u,X(u,0))\,\mathrm{d}B_u\,\Delta_\theta\Phi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&\quad\quad+ \int_0^t\int_{\mathbb{R}}\int_0^s p^\theta_{s-u}(x)\mu(u,X(u,0))\,\mathrm{d}u\,\Delta_\theta\Phi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&\quad=: I_1 + I_2 + I_3.
\end{aligned} \tag{4.25}$$

Use the fact that $p^\theta_s(x,\cdot)$ is a probability density to write $x_0(s) = \int_{\mathbb{R}} p^\theta_s(x,y)x_0(s)\,\mathrm{d}y$ and use Fubini's theorem, the partial integration formula from Lemma 4.15 and the fact that p^θ_t is a fundamental solution, to get

$$\begin{aligned}
I_1 &= \int_0^t\int_{\mathbb{R}}\int_{\mathbb{R}} p^\theta_s(x,y)x_0(s)\,\mathrm{d}y\,\Delta_\theta\Phi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&= \int_0^t x_0(s)\int_{\mathbb{R}}\int_{\mathbb{R}} p^\theta_s(x,y)\Delta_\theta\Phi_s(x)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}}\int_0^t x_0(s)\Big(\Delta_\theta p^\theta_s(x,y)\Big)\Phi_s(x)\,\mathrm{d}s\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}}\int_0^t \left(\frac{\partial}{\partial s}p^\theta_s(x,y)\right)x_0(s)\Phi_s(x)\,\mathrm{d}s\,\mathrm{d}y\,\mathrm{d}x.
\end{aligned}$$

We denote the summands on the right-hand side of (4.13) as $X_i(t,x)$ for $i=2,3$, that is, $X(t,x) = x_0 + X_2(t,x) + X_3(t,x)$. Due to the s-dependence in $x_0(s)$ and Φ_s, we apply the product rule to get

$$\begin{aligned}
I_1 &= \int_{\mathbb{R}}\int_{\mathbb{R}}\int_0^t \frac{\partial}{\partial s}\Big((x_0(s)p_s^\theta(x,y)\Phi_s(x)\Big)\,\mathrm{d}s\,\mathrm{d}y\,\mathrm{d}x \\
&\quad - \int_{\mathbb{R}}\int_{\mathbb{R}}\int_0^t p_s^\theta(x,y)\frac{\partial}{\partial s}\Big(x_0(s)\Phi_s(x)\Big)\,\mathrm{d}s\,\mathrm{d}y\,\mathrm{d}x \\
&= \langle x_0(t),\Phi_t\rangle - \langle x_0(0),\Phi_0\rangle \\
&\quad - \int_0^t\int_{\mathbb{R}} x_0(s)\frac{\partial}{\partial s}\Phi_s(x)\,\mathrm{d}x\,\mathrm{d}s - \int_0^t\int_{\mathbb{R}} \Phi_s(x)\frac{\partial}{\partial s}x_0(s)\,\mathrm{d}x\,\mathrm{d}s. \qquad (4.26)
\end{aligned}$$

Similarly, using the stochastic Fubini theorem, we get

$$\begin{aligned}
I_2 &= \int_0^t\int_{\mathbb{R}}\int_0^s p_{s-u}^\theta(x)\sigma(u,X(u,0))\,\mathrm{d}B_u\,\Delta_\theta\Phi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&= \int_0^t\int_{\mathbb{R}}\int_u^t \left(\frac{\partial}{\partial s}p_{s-u}^\theta(x)\right)\Phi_s(x)\,\mathrm{d}s\,\mathrm{d}x\,\sigma(u,X(u,0))\,\mathrm{d}B_u \\
&= \int_0^t\int_{\mathbb{R}}\int_u^t \frac{\partial}{\partial s}\left(p_{s-u}^\theta(x)\Phi_s(x)\right)\mathrm{d}s\,\mathrm{d}x\,\sigma(u,X(u,0))\,\mathrm{d}B_u \\
&\quad - \int_0^t\int_{\mathbb{R}}\int_u^t p_{s-u}^\theta(x)\left(\frac{\partial}{\partial s}\Phi_s(x)\right)\mathrm{d}s\,\mathrm{d}x\,\sigma(u,X(u,0))\,\mathrm{d}B_u \\
&= \langle X_2(t,\cdot),\Phi_t\rangle - \int_0^t\int_{\mathbb{R}} p_0^\theta(x,0)\Phi_u(x)\,\mathrm{d}x\,\sigma(u,X(u,0))\,\mathrm{d}B_u \\
&\quad - \int_0^t\int_{\mathbb{R}}\int_0^s p_{s-u}^\theta(x)\sigma(u,X(u,0))\,\mathrm{d}B_u\left(\frac{\partial}{\partial s}\Phi_s(x)\right)\mathrm{d}x\,\mathrm{d}s \\
&= \langle X_2(t,\cdot),\Phi_t\rangle - \int_0^t \Phi_u(0)\sigma(u,X(u,0))\,\mathrm{d}B_u \\
&\quad - \int_0^t\int_{\mathbb{R}} X_2(s,x)\left(\frac{\partial}{\partial s}\Phi_s(x)\right)\mathrm{d}x\,\mathrm{d}s \qquad (4.27)
\end{aligned}$$

and

$$\begin{aligned}
I_3 &= \int_0^t\int_{\mathbb{R}}\int_0^s p_{s-u}^\theta(x)\mu(u,X(u,0))\,\mathrm{d}u\,\Delta_\theta\Phi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&= \int_0^t\int_{\mathbb{R}}\int_u^t \frac{\partial}{\partial s}\left(p_{s-u}^\theta(x)\Phi_s(x)\right)\mathrm{d}s\,\mathrm{d}x\,\mu(u,X(u,0))\,\mathrm{d}u \\
&\quad - \int_0^t\int_{\mathbb{R}}\int_u^t p_{s-u}^\theta(x)\left(\frac{\partial}{\partial s}\Phi_s(x)\right)\mathrm{d}s\,\mathrm{d}x\,\mu(u,X(u,0))\,\mathrm{d}u \\
&= \langle X_3(t,\cdot),\Phi_t\rangle - \int_0^t \Phi_u(0)\mu(u,X(u,0))\,\mathrm{d}u \\
&\quad - \int_0^t\int_{\mathbb{R}} X_3(s,x)\left(\frac{\partial}{\partial s}\Phi_s(x)\right)\mathrm{d}x\,\mathrm{d}s. \qquad (4.28)
\end{aligned}$$

Plugging (4.26), (4.27) and (4.28) into (4.25) and rearranging the terms yields

$$\begin{aligned}\langle X(t,\cdot),\Phi_t\rangle &= \int_{\mathbb{R}}\left(x_0(0)\Phi_0(x)+\int_0^t \Phi_s(x)\frac{\partial}{\partial s}x_0(s)\,\mathrm{d}s\right)\mathrm{d}x \\ &\quad+\int_0^t\int_{\mathbb{R}} X(s,x)\left(\Delta_\theta\Phi_s(x)+\frac{\partial}{\partial s}\Phi_s(x)\right)\mathrm{d}x\,\mathrm{d}s \\ &\quad+\int_0^t \mu(s,X(s,0))\Phi_s(0)\,\mathrm{d}s+\int_0^t \sigma(s,X(s,0))\Phi_s(0)\,\mathrm{d}B_s,\end{aligned}$$

for $t\in[0,T]$, which shows that (4.24) holds. □

We summarize the findings of Step 1 in the following proposition.

Proposition 4.17. *Every strong L^p-solution $(X_t)_{t\in[0,T]}$ to the stochastic Volterra equation* (4.2) *with p given by* (4.6) *generates a strong solution $(X_t)_{t\in[0,T],x\in\mathbb{R}}$, as defined in* (4.7)*, to the distributional SPDE* (4.24) *with $X\in C([0,T]\times\mathbb{R})$ a.s. Furthermore, $\sup_{t\in[0,T],x\in\mathbb{R}}\mathbb{E}[|X(t,x)|^q]<\infty$ for all $q\in(0,\infty)$ and, for $Z(t,x):=X(t,x)-x_0(t)$ and $q\in[2,\infty)$,*

$$\mathbb{E}[|Z(t,x)-Z(t',x')|^q]\lesssim |t'-t|^{(\frac{1}{2}-\alpha)q}+|x-x'|^{(\frac{1}{2}-\alpha)q},$$

for all $t,t'\in[0,T]$ and $x,x'\in[-1,1]$.

Proof. The implication of the solution to (4.24) by the one to (4.2) is given by Lemma 4.6 and Lemma 4.16, the continuity by Lemma 4.6 and the remaining properties by Proposition 4.11. □

4.4 Step 2 and 3: Implementing Yamada–Watanabe's approach

The next steps are to use the classical approximation of the absolute value function introduced by Yamada–Watanabe [YW71], allowing us to apply Itô's formula. Recall that, by Assumption 4.1 (ii), σ is ξ-Hölder continuous for some $\xi\in[\frac{1}{2},1]$. Hence, there exists a strictly increasing function $\rho\colon[0,\infty)\to[0,\infty)$ such that $\rho(0)=0$,

$$|\sigma(t,x)-\sigma(t,y)|\le C_\sigma|x-y|^\xi\le\rho(|x-y|)\quad\text{for } t\in[0,T]\text{ and } x,y\in\mathbb{R}$$

and

$$\int_0^\varepsilon \frac{1}{\rho(x)^2}\,\mathrm{d}x=\infty\quad\text{for all }\varepsilon>0.$$

Based on ρ, we define a sequence $(\phi_n)_{n\in\mathbb{N}}$ of functions mapping from $\mathbb{R}$ to $\mathbb{R}$ that approximates the absolute value in the following way: Let $(a_n)_{n\in\mathbb{N}}$ be a strictly decreasing sequence with $a_0=1$ such that $a_n\to0$ as $n\to\infty$ and

$$\int_{a_n}^{a_{n-1}}\frac{1}{\rho(x)^2}\,\mathrm{d}x=n. \tag{4.29}$$

Furthermore, we define a sequence of mollifiers: let $(\psi_n)_{n\in\mathbb{N}} \in C_0^\infty(\mathbb{R})$ be smooth functions with compact support such that $\operatorname{supp}(\psi_n) \subset (a_n, a_{n-1})$,

$$0 \le \psi_n(x) \le \frac{2}{n\rho(x)^2} \le \frac{2}{nx}, \quad x \in \mathbb{R}, \quad \text{and} \quad \int_{a_n}^{a_{n-1}} \psi_n(x)\,\mathrm{d}x = 1. \tag{4.30}$$

We set

$$\phi_n(x) := \int_0^{|x|} \left(\int_0^y \psi_n(z)\,\mathrm{d}z \right) \mathrm{d}y, \quad x \in \mathbb{R}. \tag{4.31}$$

By (4.30) and the compact support of ψ_n, it follows that $\phi_n(\cdot) \to |\cdot|$ uniformly as $n \to \infty$. Since every ψ_n and, thus, every ϕ_n is zero in a neighborhood around zero, the functions ϕ_n are smooth with

$$\|\phi_n'\|_\infty \le 1, \quad \phi_n'(x) = \operatorname{sgn}(x) \int_0^{|x|} \psi_n(y)\,\mathrm{d}y \quad \text{and} \quad \phi_n''(x) = \psi_n(|x|), \quad \text{for } x \in \mathbb{R}.$$

Let X^1 and X^2 be two strong solutions to the SPDE (4.24) for a given Brownian motion $(B_t)_{t\in[0,T]}$ such that $X^1, X^2 \in C([0,T]\times\mathbb{R})$ a.s. We define $\tilde{X} := X^1 - X^2$ and consider, for some $\Phi_x^m \in C_0^2(\mathbb{R})$ for fixed $x \in \mathbb{R}$ and $m \in \mathbb{R}_+$ (we will later define m depending on n and Φ_x^m is independent of t):

$$\langle \tilde{X}_t, \Phi_x^m \rangle = \int_{\mathbb{R}} \tilde{X}(t,y)\Phi_x^m(y)\,\mathrm{d}y,$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product on $L^2(\mathbb{R})$.

Proposition 4.18. *For a fixed $x \in \mathbb{R}$ and $m \in \mathbb{R}_+$, let $\Phi_x^m \in C_0^2(\mathbb{R})$ be such that $\Delta_\theta \Phi_x^m$ is well-defined. Then, for $t \in [0,T]$, one has*

$$\begin{aligned}
\phi_n(\langle \tilde{X}_t, \Phi_x^m \rangle) = &\int_0^t \phi_n'(\langle \tilde{X}_s, \Phi_x^m \rangle)\langle \tilde{X}_s, \Delta_\theta \Phi_x^m \rangle\,\mathrm{d}s \\
&+ \int_0^t \phi_n'(\langle \tilde{X}_s, \Phi_x^m \rangle)\Phi_x^m(0)(\mu(s, X^1(s,0)) - \mu(s, X^2(s,0)))\,\mathrm{d}s \\
&+ \int_0^t \phi_n'(\langle \tilde{X}_s, \Phi_x^m \rangle)\Phi_x^m(0)(\sigma(s, X^1(s,0)) - \sigma(s, X^2(s,0)))\,\mathrm{d}B_s \\
&+ \frac{1}{2}\int_0^t \psi_n(|\langle \tilde{X}_s, \Phi_x^m \rangle|)\Phi_x^m(0)^2(\sigma(s, X^1(s,0)) - \sigma(s, X^2(s,0)))^2\,\mathrm{d}s.
\end{aligned} \tag{4.32}$$

Proof. By (4.24), $(\langle \tilde{X}_t, \Phi_x^m \rangle)_{t\in[0,T]}$ is a semimartingale. Therefore, we are able to apply Itô's formula to ϕ_n, which yields the result. □

Note that (4.32) defines a function in x. We want to integrate this against another non-negative test function with the following properties.

Assumption 4.19. *Let $\Psi \in C^2([0,T]\times\mathbb{R})$ be twice continuously differentiable such that*

(i) $\Psi_t(0) > 0$ *for all* $t \in [0,T]$,

(ii) $\Gamma(t) := \{x \in \mathbb{R} : \exists s \le t \text{ s.t. } |\Psi_s(x)| > 0\} \subset B(0, J(t))$ *for some* $0 < J(t) < \infty$,

(iii)
$$\sup_{s\le t}\left|\int_{\mathbb{R}} |x|^{-\theta}\left(\frac{\partial \Psi_s(x)}{\partial x}\right)^2 \mathrm{d}x\right| < \infty, \quad t \in [0,T].$$

We will later choose an explicit function Ψ and show that it fulfills Assumption 4.19. Then, we get the following equality, where the extra term $I_5^{m,n}$ arises due to the t-dependence of Ψ.

Proposition 4.20. *For Ψ fulfilling Assumption 4.19, we have*

$$\begin{aligned}
&\langle \phi_n(\langle \tilde{X}_t, \Phi^m_\cdot\rangle), \Psi_t\rangle \\
&= \int_0^t \langle \phi_n'(\langle \tilde{X}_s, \Phi^m_\cdot\rangle)\langle \tilde{X}_s, \Delta_\theta \Phi^m_\cdot\rangle, \Psi_s\rangle \,\mathrm{d}s \\
&\quad + \int_0^t \langle \phi_n'(\langle \tilde{X}_s, \Phi^m_\cdot\rangle)\Phi^m_\cdot(0), \Psi_s\rangle(\mu(s, X^1(s,0)) - \mu(s, X^2(s,0)))\,\mathrm{d}s \\
&\quad + \int_0^t \langle \phi_n'(\langle \tilde{X}_s, \Phi^m_\cdot\rangle)\Phi^m_\cdot(0), \Psi_s\rangle(\sigma(s, X^1(s,0)) - \sigma(s, X^2(s,0)))\,\mathrm{d}B_s \\
&\quad + \frac{1}{2}\int_0^t \langle \psi_n(|\langle \tilde{X}_s, \Phi^m_\cdot\rangle|)\Phi^m_\cdot(0)^2, \Psi_s\rangle(\sigma(s, X^1(s,0)) - \sigma(s, X^2(s,0)))^2\,\mathrm{d}s \\
&\quad + \int_0^t \langle \phi_n(\langle \tilde{X}_s, \Phi^m_\cdot\rangle), \dot{\Psi}_s\rangle \,\mathrm{d}s \\
&=: I_1^{m,n}(t) + I_2^{m,n}(t) + I_3^{m,n}(t) + I_4^{m,n}(t) + I_5^{m,n}(t),
\end{aligned} \tag{4.33}$$

for $t \in [0,T]$, where $\dot{\Psi}_s(x) := \frac{\partial}{\partial s}\Psi_s(x)$.

Proof. We discretize $\Psi_t(x)$ in its time variable, then let the grid size go to zero and show that the resulting term converges to (4.33). Therefore, let $t_i = i2^{-k}$, $i = 0, 1, \dots, \lfloor t2^k\rfloor + 1 =: K_t^k$, where $\lfloor\cdot\rfloor$ denotes rounding down to the next integer, such that $t_{\lfloor t2^k\rfloor} \le t < t_{K_t^k}$, and denote

$$\Psi_t^k(x) := 2^k \int_{t_{i-1}}^{t_i} \Psi_s(x)\,\mathrm{d}s, \quad t \in [t_{i-1}, t_i), x \in \mathbb{R}. \tag{4.34}$$

Then, we can build the telescope sum

$$\begin{aligned}
\langle \phi_n(\langle \tilde{X}_t, \Phi^m_\cdot\rangle), \Psi_t\rangle = &\sum_{i=1}^{K_t^k} \langle \phi_n(\langle \tilde{X}_{t_i}, \Phi^m_\cdot\rangle), \Psi^k_{t_i}\rangle - \langle \phi_n(\langle \tilde{X}_{t_{i-1}}, \Phi^m_\cdot\rangle), \Psi^k_{t_{i-1}}\rangle \\
&- \langle \phi_n(\langle \tilde{X}_{t_{K_t^k}}, \Phi^m_\cdot\rangle), \Psi^k_{t_{K_t^k}}\rangle + \langle \phi_n(\langle \tilde{X}_t, \Phi^m_\cdot\rangle), \Psi_t\rangle.
\end{aligned} \tag{4.35}$$

By the continuity of $\tilde{X}$, Ψ and ϕ_n, the sum of the last two terms approaches zero as $t_{K_t^k} \to t$ and thus as $k \to \infty$.
For the terms in the summation, we use the continuity of $\tilde{X}$ and the notation $f(t_i-) := \lim_{s<t_i, s\to t_i} f(s)$, to get the equality

$$\langle \phi_n(\langle \tilde{X}_{t_i}, \Phi^m_\cdot \rangle), \Psi^k_{t_i} \rangle = \langle \phi_n(\langle \tilde{X}_{t_i-}, \Phi^m_\cdot \rangle), \Psi^k_{t_i-} \rangle + \langle \phi_n(\langle \tilde{X}_{t_i}, \Phi^m_\cdot \rangle), \Psi^k_{t_i} - \Psi^k_{t_{i-1}} \rangle.$$

By plugging this into (4.35), we get

$$\begin{aligned} \langle \phi_n(\langle \tilde{X}_t, \Phi^m_\cdot \rangle), \Psi_t \rangle &= \sum_{i=1}^{K_t^k} \langle \phi_n(\langle \tilde{X}_{t_i-}, \Phi^m_\cdot \rangle), \Psi^k_{t_i-} \rangle - \langle \phi_n(\langle \tilde{X}_{t_{i-1}}, \Phi^m_\cdot \rangle), \Psi^k_{t_{i-1}} \rangle \\ &\quad + \sum_{i=1}^{K_t^k} \langle \phi_n(\langle \tilde{X}_{t_i}, \Phi^m_\cdot \rangle), \Psi^k_{t_i} - \Psi^k_{t_{i-1}} \rangle =: A_t^k + C_t^k. \end{aligned}$$

For A_t^k, we get, by applying Itô's formula, that

$$\begin{aligned} A_t^k &= \sum_{i=1}^{K_t} \langle \phi_n(\langle \tilde{X}_{t_i}, \Phi^m_\cdot \rangle), \Psi^k_{t_{i-1}} \rangle - \langle \phi_n(\langle \tilde{X}_{t_{i-1}}, \Phi^m_\cdot \rangle), \Psi^k_{t_{i-1}} \rangle \\ &\to I_1^{m,n}(t) + I_2^{m,n}(t) + I_3^{m,n}(t) + I_4^{m,n}(t) \quad \text{as } k \to \infty, \end{aligned}$$

by the continuity of Ψ.
Thus, it remains to show that C_t^k converges to $I_5^{m,n}(t)$. To that end, we use the construction (4.34) and Fubini's theorem to conclude that

$$\begin{aligned} C_t^k &= \sum_{i=1}^{K_t^k} \left\langle \phi_n(\langle \tilde{X}_{t_i}, \Phi^m_\cdot \rangle), 2^k \int_{t_{i-1}}^{t_i} (\Psi_s - \Psi_{s-2^{-k}}) \,\mathrm{d}s \right\rangle \\ &= \sum_{i=1}^{K_t^k} \left\langle \phi_n(\langle \tilde{X}_{t_i}, \Phi^m_\cdot \rangle), 2^k \int_{t_{i-1}}^{t_i} \int_{s-2^{-k}}^{s} \dot{\Psi}_r \,\mathrm{d}r \,\mathrm{d}s \right\rangle \\ &= 2^k \sum_{i=1}^{K_t^k} \int_{t_{i-1}}^{t_i} \int_{s-2^{-k}}^{s} \langle \phi_n(\langle \tilde{X}_{t_i}, \Phi^m_\cdot \rangle), \dot{\Psi}_r \rangle \,\mathrm{d}r \,\mathrm{d}s \\ &= 2^k \sum_{i=1}^{K_t^k} \int_{t_{i-1}}^{t_i} \int_{s-2^{-k}}^{s} \langle \phi_n(\langle \tilde{X}_{t_i}, \Phi^m_\cdot \rangle), \dot{\Psi}_r \rangle - \langle \phi_n(\langle \tilde{X}_r, \Phi^m_\cdot \rangle), \dot{\Psi}_r \rangle \,\mathrm{d}r \,\mathrm{d}s \\ &\quad + 2^k \sum_{i=1}^{K_t^k} \int_{t_{i-1}}^{t_i} \int_{s-2^{-k}}^{s} \langle \phi_n(\langle \tilde{X}_r, \Phi^m_\cdot \rangle), \dot{\Psi}_r \rangle \,\mathrm{d}r \,\mathrm{d}s. \end{aligned}$$

The first summand can be bounded by

$$\int_0^t \sup_{u\le t, |u-r|\le 2^{-k}} \left| \langle \phi_n(\langle \tilde{X}_u, \Phi^m_\cdot \rangle), \dot{\Psi}_r \rangle - \langle \phi_n(\langle \tilde{X}_r, \Phi^m_\cdot \rangle), \dot{\Psi}_r \rangle \right| \mathrm{d}r,$$

which converges to zero a.s. as $k \to \infty$ by the continuity and boundedness of $\tilde{X}$. Furthermore, we get, by

$$2^k \int_{s-2^{-k}}^{s} \langle \phi_n(\langle \tilde{X}_r, \Phi_\cdot^m \rangle), \dot{\Psi}_r \rangle \, \mathrm{d}r \to \langle \phi_n(\langle \tilde{X}_s, \Phi_\cdot^m \rangle), \dot{\Psi}_s \rangle \quad \text{as } k \to \infty$$

and the dominated convergence theorem, that

$$C_t^k \to \int_0^t \langle \phi_n(\langle \tilde{X}_s, \Phi_\cdot^m \rangle), \dot{\Psi}_s \rangle \, \mathrm{d}s \quad \text{as } k \to \infty,$$

which proves the proposition. □

We will bound the expectation of the terms $I_1^{m,n}$ to $I_5^{m,n}$ as $m, n \to \infty$ in Section 4.5.

4.5 Step 4: Passing to the limit

Before we can pass to the limit in (4.33), we need to choose a sequence $(\Phi_x^{m,n})_{n\in\mathbb{N}}$ of smooth functions $\Phi_x^{m,n} \in C_0^\infty(\mathbb{R})$ for some $x \in \mathbb{R}$ and for $m \in \mathbb{R}_+$, which approximates the Dirac distribution δ_x explicitly. We will choose some $m = m^{(n)}$ dependent on the index n of the Yamada–Watanabe approximation and, for notational simplicity, will skip the m-dependence and shortly write $(\Phi_x^n)_{n\in\mathbb{N}}$.

4.5.1 Explicit choice of the test function

We want to approximate with Φ_x^n a Dirac distribution centered around $x \in \mathbb{R}$. Therefore, we choose it to coincide with the sum of two Gaussian kernels with mean x and y, respectively, and standard deviation m^{-1}, when x and y are close. The reason for this construction is that we want to keep the mass of Φ in $B(0, \frac{1}{m^{(n)}})$ constant as $n \to \infty$. For this purpose, we define

$$\tilde{\Phi}_x^m(y) := \frac{1}{\sqrt{2\pi m^{-2}}} e^{-\frac{(y-x)^2}{2m^{-2}}}$$

and, to construct the compact support, let $\tilde{\psi}_x^{m,n}$ be smooth functions for $n \in \mathbb{N}$ and fixed $x \in \mathbb{R}$ with

$$\tilde{\psi}_x^{m,n}(y) := \begin{cases} 1, & \text{if } y \in B(x, \frac{1}{m}) \\ 0, & \text{if } y \in \mathbb{R} \setminus B(x, \frac{1}{m} + b_n) \end{cases}$$

and $0 \le \tilde{\psi}_x^{m,n}(y) \le 1$ for y elsewhere such that $\tilde{\psi}_x^{m,n}$ is smooth. Here, let $(b_n)_{n\in\mathbb{N}}$ be a sequence such that $b_n > 0$ and

$$\mu_n\left(B\left(x, \frac{1}{m} + b_n\right) \setminus B\left(x, \frac{1}{m}\right)\right) = \frac{a_n}{2},$$

where $\mu_n(A) := \int_A \tilde{\Phi}_x^m(y) \, \mathrm{d}y$ denotes the measure in terms of the above normal distribution and $a_n := e^{-\frac{n(n+1)}{2}}$ comes from the Yamada–Watanabe sequence. It is always possible

to find such a $b_n > 0$ since the mass of $\tilde{\Phi}^m_x$ in $B(x, \frac{1}{m})$ is ≈ 0.6827, which is independent of n, and $\frac{a_n}{2} < 0.3$ for all $n \in \mathbb{N}$.
Then, we define

$$\Phi^n_x(y) := c\Big(\tilde{\psi}^{m,n}_x(y)\tilde{\Phi}^m_x(y) + \tilde{\Phi}^m_y(x)\tilde{\psi}^{m,n}_y(x)\Big), \tag{4.36}$$

with $c := 1/(2m_\sigma)$, where $m_\sigma \approx 0.6827$ denotes the mass of a normal distribution $\mathcal{N}(\mu, \sigma^2)$ inside the interval $[\mu - \sigma, \mu + \sigma]$. With that choice of c, Φ^n_x approximates the Dirac distribution δ_x around x as $n \to \infty$. Note that $\Phi^n_x(y)$ is identical in terms of x and y. Furthermore, Φ^n_x owes the following properties that we will need later. To that end, let us introduce the following stopping time for $K > 0$:

$$T_K := \inf_{t\in[0,T]} \Bigg\{ \sup_{x\in[-\frac{1}{2},\frac{1}{2}]} (|X^1(t,x)| + |X^2(t,x)|) > K \Bigg\}, \tag{4.37}$$

where we use the convention $\inf \emptyset := \infty$. Note that, by the continuity of X^1 and X^2, $T_K \to \infty$ a.s. as $K \to \infty$.

Proposition 4.21. *For fixed $x \in \mathbb{R}$, Φ^n_x, as defined in (4.36), fulfills:*

(i) $\Delta_{\theta,x}\Phi^n_x(y) = \Delta_{\theta,y}\Phi^n_x(y)$ *for all $x, y \in \mathbb{R}$, where $\Delta_{\theta,x}$ denotes Δ_θ acting on x;*

(ii) $\int_{\mathbb{R}} \Phi^n_x(0)^2 \,\mathrm{d}x \lesssim m^{(n)}$ *for all $n \in \mathbb{N}$;*

(iii) $\int_{\mathbb{R}} \Phi^n_x(0) \,\mathrm{d}x \leq 2$ *for all $n \in \mathbb{N}$;*

(iv) for all $(s, x) \in [0, T] \times \mathbb{R}$,

$$\langle \tilde{X}_s, \Phi^n_x \rangle \to \tilde{X}(s, x) \quad \textit{and} \quad \phi'_n(\langle \tilde{X}_s, \Phi^n_x \rangle)\langle \tilde{X}_s, \Phi^n_x \rangle \to |\tilde{X}(s, x)|, \quad \textit{as } n \to \infty;$$

(v) given $s \in [0, T_K]$, there exists a constant $C_K > 0$ that is independent from n, such that, if

$$\left| \int_{\mathbb{R}} \tilde{X}(s, y)\Phi^n_x(y) \,\mathrm{d}y \right| \leq a_{n-1} \tag{4.38}$$

holds, then there is some $\hat{x} \in B(x, \frac{1}{m})$ such that $|\tilde{X}(s, \hat{x})| \leq C_K a_{n-1}$.

Proof. (i) This statement is clear since Φ^n_x is identical in x and y.
(ii) We denote $c := \frac{1}{\sqrt{2\pi}}$ to get

$$\int_{\mathbb{R}} \Phi^n_x(0)^2 \,\mathrm{d}x \leq \int_{\mathbb{R}} \left(cme^{-\frac{|x|^2}{2m^{-2}}} \right)^2 \mathrm{d}x \leq cm \int_{\mathbb{R}} cme^{-\frac{|x|^2}{2m^{-2}}} \,\mathrm{d}x = cm.$$

(iii) $\int_{\mathbb{R}} \Phi^n_x(0) \,\mathrm{d}x \leq 2 \int_{\mathbb{R}} \tilde{\Phi}^m_x(0) \,\mathrm{d}x = 2$.
(iv) From the construction of Φ^n_x we get that

$$\int_{\mathbb{R}} \tilde{X}(s, y)\Phi^n_x(y) \,\mathrm{d}y \to \int_{\mathbb{R}} \tilde{X}(s, y)\delta_x(y) \,\mathrm{d}y = \tilde{X}(s, x) \quad \text{as } n \to \infty.$$

Furthermore, we know that $\phi_n'(x)x \to |x|$ as $n \to \infty$ uniformly in $x \in \mathbb{R}$ and thus the second statement follows.
(v) Let us write

$$\int_{\mathbb{R}} \tilde{X}(s,y)\Phi_x^n(y)\,\mathrm{d}y = \int_{B(x,\frac{1}{m})} \tilde{X}(s,y)\Phi_x^n(y)\,\mathrm{d}y + \int_{\mathbb{R}\setminus B(x,\frac{1}{m})} \tilde{X}(s,y)\Phi_x^n(y)\,\mathrm{d}y. \tag{4.39}$$

By the construction of $\tilde{\psi}_x^{m,n}$ we know that Φ_x^n vanishes outside the ball $B(x, \frac{1}{m} + b_n)$, and, by the choice of b_n, we know that the mass of Φ_x^n in $B(x, \frac{1}{m} + b_n) \setminus B(x, \frac{1}{m})$ is $a_{n-1}/2$. Since we have that $s \leq T_K$, we can bound

$$\left| \int_{\mathbb{R}\setminus B(x,\frac{1}{m})} \tilde{X}(s,y)\Phi_x^n(y)\,\mathrm{d}y \right| \leq 2K \int_{\mathbb{R}\setminus B(x,\frac{1}{m})} \Phi_x^n(y)\,\mathrm{d}y \leq K a_{n-1}.$$

Thus, by assumption and (4.39), we have that

$$\left| \int_{B(x,\frac{1}{m})} \tilde{X}(s,y)\Phi_x^n(y)\,\mathrm{d}y \right| \leq (K+1)a_{n-1},$$

and, since Φ_x^n is the sum of two Gaussian densities with standard deviation $\frac{1}{m}$, we know that its mass inside the ball is $\approx 2 \cdot 0.6827$ and can conclude, using the continuity of $\tilde{X}$, that

$$(K+1)a_{n-1} \geq \int_{B(x,\frac{1}{m})} \Phi_x^n(y)\,\mathrm{d}y \inf_{y\in B(x,\frac{1}{m})} |\tilde{X}(s,y)| \geq 1.3 \inf_{y\in B(x,\frac{1}{m})} |\tilde{X}(s,y)|,$$

and thus, the statement holds with $C_K = (K+1)/1.3$. □

4.5.2 Bounding the Yamada–Watanabe terms

We start with the summands $I_1^{m,n}$, $I_2^{m,n}$, $I_3^{m,n}$ and $I_5^{m,n}$ in (4.33) and will analyze $I_4^{m,n}$ later. To that end, we need the following elementary estimate.

Lemma 4.22. *If $f \in C_0^2(\mathbb{R})$ is non-negative and not identically zero, then*

$$\sup_{x\in\mathbb{R}\colon f(x)>0} \{(f'(x))^2 f(x)^{-1}\} \leq 2\|f''(x)\|_\infty.$$

Proof. Choose some $x \in \mathbb{R}$ with $f(x) > 0$ and assume w.l.o.g. that $f'(x) > 0$. Let

$$x_1 := \sup\{x' < x \colon f'(x') = 0\},$$

which exists due to the compact support of f. By the extended mean value theorem (see [Apo67, Theorem 4.6]), applied to f and $(f')^2$, there exists an $x_2 \in (x_1, x)$ such that

$$(f'(x)^2 - f'(x_1)^2)f'(x_2) = (f(x) - f(x_1))\frac{\partial (f')^2}{\partial x}(x_2).$$

By the choice of x_1, we know that $f'(x_2) > 0$, and thus with $f'(x_1) = 0$,

$$f'(x)^2 = (f(x) - f(x_1))2f''(x_2).$$

Since f is strictly increasing on (x_1, x) and non-negative, we conclude

$$\frac{f'(x)^2}{f(x)} \leq \frac{f'(x)^2}{f(x) - f(x_1)} = 2f''(x_2) \leq 2\|f''\|_\infty.$$

□

We want to take expectations on both sides of (4.33) and then send $m, n \to \infty$.

Lemma 4.23. *For any stopping time $\mathcal{T}$ and fixed $t \in [0, T]$ we have:*

(i) $\lim\limits_{m,n\to\infty} \mathbb{E}[I_1^{m,n}(t \wedge \mathcal{T})] \leq \mathbb{E}\big[\int_0^{t\wedge\mathcal{T}} \int_{\mathbb{R}} |\tilde{X}(s,x)| \Delta_\theta \Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s\big]$*;*

(ii) $\lim_{m,n\to\infty} \mathbb{E}[I_2^{m,n}(t \wedge \mathcal{T})] \lesssim \int_0^{t\wedge\mathcal{T}} \Psi_s(0)\,\mathbb{E}[|\tilde{X}(s,0)|]\,\mathrm{d}s$*;*

(iii) $\mathbb{E}[I_3^{m,n}(t \wedge \mathcal{T})] = 0$ *for all* $m, n \in \mathbb{N}$*;*

(iv) $\lim\limits_{m,n\to\infty} \mathbb{E}[I_5^{m,n}(t \wedge \mathcal{T})] = \mathbb{E}\big[\int_0^{t\wedge\mathcal{T}} \int_{\mathbb{R}} |\tilde{X}(s,x)| \dot{\Psi}_s(x)\,\mathrm{d}x\,\mathrm{d}s\big]$.

Proof. (i) We need to rewrite $I_1^{m,n}$. We use the property of Φ_x^n from Proposition 4.21 (i) and the product rule to get

$$\begin{aligned}
I_1^{m,n}(t) &= \int_0^t \int_{\mathbb{R}} \phi_n'(\langle \tilde{X}_s, \Phi_x^n \rangle) \int_{\mathbb{R}} \tilde{X}(s,y) \Delta_{y,\theta} \Phi_x^n(y)\,\mathrm{d}y\,\Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&= \int_0^t \int_{\mathbb{R}} \phi_n'(\langle \tilde{X}_s, \Phi_x^n \rangle) \Delta_{x,\theta}(\langle \tilde{X}_s, \Phi_x^n \rangle) \Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&= 2\alpha^2 \int_0^t \int_{\mathbb{R}} \phi_n'(\langle \tilde{X}_s, \Phi_x^n \rangle) \Big(\frac{\partial}{\partial x} |x|^{-\theta} \frac{\partial}{\partial x} \langle \tilde{X}_s, \Phi_x^n \rangle \Big) \Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&\quad + 2\alpha^2 \int_0^t \int_{\mathbb{R}} \phi_n'(\langle \tilde{X}_s, \Phi_x^n \rangle) |x|^{-\theta} \Big(\frac{\partial^2}{\partial x^2} \langle \tilde{X}_s, \Phi_x^n \rangle \Big) \Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s.
\end{aligned}$$

Now, we use integration by parts for both summands and the compact support of Ψ_s for every $s \in [0, T]$ to get

$$\begin{aligned}
I_1^{m,n}(t) &= -2\alpha^2 \int_0^t \int_{\mathbb{R}} \psi_n(\langle \tilde{X}_s, \Phi_x^n \rangle) |x|^{-\theta} \Big(\frac{\partial}{\partial x} \langle \tilde{X}_s \Phi_x^n \rangle \Big)^2 \Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&\quad - 2\alpha^2 \int_0^t \int_{\mathbb{R}} \phi_n'(\langle \tilde{X}_s, \Phi_x^n \rangle) |x|^{-\theta} \frac{\partial}{\partial x} \langle \tilde{X}_s \Phi_x^n \rangle \frac{\partial}{\partial x} \Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s. \qquad (4.40)
\end{aligned}$$

By a very similar partial integration we see that

$$\begin{aligned}
&\int_0^t \int_{\mathbb{R}} \phi_n'(\langle \tilde{X}_s, \Phi_x^n\rangle)\langle \tilde{X}_s, \Phi_x^n\rangle \Delta_\theta \Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&\quad = -2\alpha^2 \int_0^t \int_{\mathbb{R}} \psi_n(\langle \tilde{X}_s, \Phi_x^n\rangle) \frac{\partial}{\partial x}\langle \tilde{X}_s, \Phi_x^n\rangle \langle \tilde{X}_s, \Phi_x^n\rangle |x|^{-\theta} \frac{\partial}{\partial x}\Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&\qquad - 2\alpha^2 \int_0^t \int_{\mathbb{R}} \phi_n'(\langle \tilde{X}_s, \Phi_x^n\rangle) \frac{\partial}{\partial x}\langle \tilde{X}_s, \Phi_x^n\rangle |x|^{-\theta} \frac{\partial}{\partial x}\Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s. \qquad (4.41)
\end{aligned}$$

By identifying that the second term in (4.40) coincides with the second term in (4.41), we can plug in the latter one into the first one to get

$$\begin{aligned}
I_1^{m,n}(t) &= -2\alpha^2 \int_0^t \int_{\mathbb{R}} \psi_n(\langle \tilde{X}_s, \Phi_x^n\rangle)|x|^{-\theta}\left(\frac{\partial}{\partial x}\langle \tilde{X}_s \Phi_x^n\rangle\right)^2 \Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&\quad + 2\alpha^2 \int_0^t \int_{\mathbb{R}} \psi_n(\langle \tilde{X}_s, \Phi_x^n\rangle) \frac{\partial}{\partial x}\langle \tilde{X}_s, \Phi_x^n\rangle \langle \tilde{X}_s, \Phi_x^n\rangle |x|^{-\theta} \frac{\partial}{\partial x}\Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&\quad + \int_0^t \int_{\mathbb{R}} \phi_n'(\langle \tilde{X}_s, \Phi_x^n\rangle)\langle \tilde{X}_s, \Phi_x^n\rangle \Delta_\theta \Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s \\
&= \int_0^t \big(I_{1,1}^{m,n}(s) + I_{1,2}^{m,n}(s) + I_{1,3}^{m,n}(s)\big)\,\mathrm{d}s. \qquad (4.42)
\end{aligned}$$

In order to deal with the various parts of $I_1^{m,n}$, we start with treating $I_{1,1}^{m,n}$ and $I_{1,2}^{m,n}$. Since we want to show that these parts are less than or equal to 0, we define for fixed $s \in [0,t]$:

$$\begin{aligned}
A^s &:= \left\{x \in \mathbb{R} : \left(\frac{\partial}{\partial x}\langle \tilde{X}_s, \Phi_x^n\rangle\right)^2 \Psi_s(x) \le \langle \tilde{X}_s, \Phi_x^n\rangle \frac{\partial}{\partial x}\langle \tilde{X}_s, \Phi_x^n\rangle \frac{\partial}{\partial x}\Psi_s(x)\right\} \\
&\qquad \cap \{x \in \mathbb{R} : \Psi_s(x) > 0\} \\
&= A^{+,s} \cup A^{-,s} \cup A^{0,s},
\end{aligned}$$

with

$$\begin{aligned}
&A^{+,s} := A^s \cap \left\{\frac{\partial}{\partial x}\langle \tilde{X}_s, \Phi_x^n\rangle > 0\right\}, \quad A^{-,s} := A^s \cap \left\{\frac{\partial}{\partial x}\langle \tilde{X}_s, \Phi_x^n\rangle < 0\right\} \quad \text{and} \\
&A^{0,s} := A^s \cap \left\{\frac{\partial}{\partial x}\langle \tilde{X}_s, \Phi_x^n\rangle = 0\right\}.
\end{aligned}$$

By Assumption 4.19 (i) and (iii), we can find an $\varepsilon > 0$ such that

$$B(0,\varepsilon) \subset \Gamma(t) \quad \text{and} \quad \inf_{s \le t, x \in B(0,\varepsilon)} \Psi_s(x) > 0. \qquad (4.43)$$

On $A^{+,s}$ we have, by the definition of A^s, that

$$0 < \left(\frac{\partial}{\partial x}\langle \tilde{X}_s, \Phi_x^n\rangle\right)\Psi_s(x) \le \langle \tilde{X}_s, \Phi_x^n\rangle \frac{\partial}{\partial x}\Psi_s(x),$$

and, therefore, we can bound the $A^{+,s}$-part of $I_{1,2}^{m,n}$ for any $t \in [0,T]$ by

$$\begin{aligned}
&\int_0^t \int_{A^{+,s}} \psi_n(\langle \tilde{X}_s, \Phi_x^n \rangle) \frac{\partial}{\partial x} \langle \tilde{X}_s, \Phi_x^n \rangle \langle \tilde{X}_s, \Phi_x^n \rangle |x|^{-\theta} \frac{\partial}{\partial x} \Psi_s(x) \, \mathrm{d}x \, \mathrm{d}s \\
&\quad \leq \int_0^t \int_{A^{+,s}} \psi_n(\langle \tilde{X}_s, \Phi_x^n \rangle) |x|^{-\theta} \langle \tilde{X}_s, \Phi_x^n \rangle^2 \frac{(\frac{\partial}{\partial x} \Psi_s(x))^2}{\Psi_s(x)} \, \mathrm{d}x \, \mathrm{d}s \\
&\quad \leq \int_0^t \int_{A^{+,s}} \frac{2}{n} \mathbb{1}_{\{a_{n-1} \leq |\langle \tilde{X}_s, \Phi_x^n \rangle| \leq a_n\}} |x|^{-\theta} \langle \tilde{X}_s, \Phi_x^n \rangle \frac{(\frac{\partial}{\partial x} \Psi_s(x))^2}{\Psi_s(x)} \, \mathrm{d}x \, \mathrm{d}s \\
&\quad \leq \frac{2a_n}{n} \int_0^t \int_{\mathbb{R}} \mathbb{1}_{\{\Psi_s(x) > 0\}} |x|^{-\theta} \frac{(\frac{\partial}{\partial x} \Psi_s(x))^2}{\Psi_s(x)} \, \mathrm{d}x \, \mathrm{d}s.
\end{aligned}$$

Next, we split the integral by using ε from (4.43) to be able to apply Assumption 4.19 and Lemma 4.22 and get

$$\begin{aligned}
&\int_0^t \int_{A^{+,s}} \psi_n(\langle \tilde{X}_s, \Phi_x^n \rangle) \frac{\partial}{\partial x} \langle \tilde{X}_s, \Phi_x^n \rangle \langle \tilde{X}_s, \Phi_x^n \rangle |x|^{-\theta} \frac{\partial}{\partial x} \Psi_s(x) \, \mathrm{d}x \, \mathrm{d}s \\
&\quad \leq \frac{2a_n}{n} \int_0^t \left(\int_{B(0,\varepsilon)} |x|^{-\theta} \frac{(\frac{\partial}{\partial x} \Psi_s(x))^2}{\Psi_s(x)} \, \mathrm{d}x + 2\|D^2 \Psi_s\|_\infty \int_{\Gamma(t) \setminus B(0,\varepsilon)} |x|^{-\theta} \, \mathrm{d}x \right) \mathrm{d}s \\
&\quad =: \frac{2a_n}{n} C(\Psi, t).
\end{aligned}$$

Note that $\varepsilon > 0$ is fixed and thus the ε-dependence of $C(\Psi, t)$ does not matter.
On the set $A^{-,s}$,

$$0 > \left(\frac{\partial}{\partial x} \langle \tilde{X}_s, \Phi_x^n \rangle \right) \Psi_s(x) \geq \langle \tilde{X}_s, \Phi_x^n \rangle \frac{\partial}{\partial x} \Psi_s(x), \tag{4.44}$$

holds and, since both terms in (4.44) are negative, we can use the same calculation as above to get

$$\int_0^t \int_{A^{+,s}} \psi_n(\langle \tilde{X}_s, \Phi_x^n \rangle) \frac{\partial}{\partial x} \langle \tilde{X}_s, \Phi_x^n \rangle \langle \tilde{X}_s, \Phi_x^n \rangle |x|^{-\theta} \frac{\partial}{\partial x} \Psi_s(x) \, \mathrm{d}x \, \mathrm{d}s \leq \frac{2a_n}{n} C(\Psi, t).$$

Finally, on the set $A^{0,s}$,

$$\int_0^t \int_{A^{+,s}} \psi_n(\langle \tilde{X}_s, \Phi_x^n \rangle) \frac{\partial}{\partial x} \langle \tilde{X}_s, \Phi_x^n \rangle \langle \tilde{X}_s, \Phi_x^n \rangle |x|^{-\theta} \frac{\partial}{\partial x} \Psi_s(x) \, \mathrm{d}x \, \mathrm{d}s = 0$$

and thus

$$\mathbb{E}[I_{1,1}^{m,n}(t \wedge \mathcal{T}) + I_{1,2}^{m,n}(t \wedge \mathcal{T})] \leq 4\alpha^2 C(\Psi, t) \frac{a_n}{n} \to 0 \quad \text{as } n \to \infty.$$

The remaining term in (4.42), we have to deal with, is

$$I_{1,3}^{m,n} = \int_0^t \int_{\mathbb{R}} \phi_n'(\langle \tilde{X}_s, \Phi_x^n \rangle) \langle \tilde{X}_s, \Phi_x^n \rangle \Delta_\theta \Psi_s(x) \, \mathrm{d}x \, \mathrm{d}s.$$

Therefore, we apply Proposition 4.21 (iv) to get the pointwise convergence

$$\phi_n'(\langle \tilde{X}_s, \Phi_x^n \rangle)\langle \tilde{X}_s, \Phi_x^n \rangle \to \tilde{X}(s,x) \quad \text{as } m,n \to \infty.$$

To complete our proof, we only need to show uniform integrability of $|\phi_n'(\langle \tilde{X}_s, \Phi_x^n \rangle)\langle \tilde{X}_s, \Phi_x^n \rangle|$ in terms of $m,n \in \mathbb{N}$ on $([0,T] \times B(0,J(t)) \times \Omega)$, since Ψ vanishes outside $B(0,J(t))$. First, by the inequality $|\phi_n'| \leq 1$, we can bound

$$|\phi_n'(\langle \tilde{X}_s, \Phi_x^n \rangle)\langle \tilde{X}_s, \Phi_x^n \rangle| \leq \langle |\tilde{X}_s|, \Phi_x^n \rangle.$$

Inserting the function Φ^n from (4.36), taking the mean and using Proposition 4.11 (i), we can bound

$$\mathbb{E}[|\langle |\tilde{X}_s|, \Phi_x^n \rangle|] \leq \mathbb{E}\left[\int_{\mathbb{R}} |\tilde{X}(s,y)| 2\tilde{\Phi}_x^m(y)\,\mathrm{d}y\right] \leq 2\sup_{y\in\mathbb{R}} \mathbb{E}[|\tilde{X}(s,y)|] \int_{\mathbb{R}} \tilde{\Phi}_x^{m(n)}(y)\,\mathrm{d}y < \infty, \tag{4.45}$$

thus the claimed integrability holds and we get

$$\lim_{m,n\to\infty} \mathbb{E}[I_{1,3}^{m,n}(t \wedge \mathcal{T})] \leq \mathbb{E}\left[\int_0^{t\wedge\mathcal{T}} \int_{\mathbb{R}} |\tilde{X}(s,x)| \Delta_\theta \Psi_s(x)\,\mathrm{d}x\,\mathrm{d}s\right]$$

and, altogether, we have shown the statement.
(ii) Again the inequality $|\phi_n'| \leq 1$ and the Lipschitz continuity of μ yield

$$\mathbb{E}[I_2^{m,n}(t \wedge \mathcal{T})] \lesssim \int_0^{t\wedge\mathcal{T}} \left(\int_{\mathbb{R}} \Phi_x^n(0)\Psi_s(x)\,\mathrm{d}x\right) \mathbb{E}[|\tilde{X}(s,0)|]\,\mathrm{d}s.$$

Sending $m,n \to \infty$ gives the statement as $\Phi_x^n(0) \to \delta_0(x)$.
(iii) We set $g_{m,n}(s) := \langle \phi_n'(\langle \tilde{X}_s, \Phi_\cdot^n \rangle)\Phi_\cdot^n(0), \Psi_s \rangle$. Then, by $|\phi_n'| \leq 1$, one has

$$|g_{m,n}(s)| = \left|\int_{\mathbb{R}} \phi_n'(\langle \tilde{X}_s, \Phi_x^n \rangle)\Phi_x^n(0)\Psi_s(x)\,\mathrm{d}x\right| \leq \|\Psi\|_\infty \int_{\mathbb{R}} 2\tilde{\Phi}_0^m(x)\,\mathrm{d}x = 2\|\Psi\|_\infty$$

by the construction of Φ^n in (4.36). Thus, $I_3^{m,n}(t \wedge \mathcal{T})$ is a continuous local martingale with quadratic variation

$$\begin{aligned}\langle I_3^{m,n} \rangle_{t\wedge\mathcal{T}} &\leq 4\|\Psi\|_\infty^2 \int_0^{t\wedge\mathcal{T}} (\sigma(s,X^1(s,0)) - \sigma(s,X^2(s,0)))^2\,\mathrm{d}s \\ &\lesssim \int_0^{t\wedge\mathcal{T}} (|X^1(s,0)| + |X^2(s,0)| + 2)^2\,\mathrm{d}s\end{aligned}$$

by the growth condition on σ and, consequently, by Proposition 4.11,

$$\mathbb{E}[\langle I_3^{m,n} \rangle_{t\wedge\mathcal{T}}] < \infty,$$

such that $I_3^{m,n}(t \wedge \mathcal{T})$ is a square integrable martingale with mean 0.

(iv) We want to calculate the limit as $n, m \to \infty$ of the term

$$\mathbb{E}[I_5^{m,n}(t \wedge \mathcal{T})] = \mathbb{E}\bigg[\int_0^{t\wedge\mathcal{T}} \langle \phi_n(\langle \tilde{X}_s, \Phi_\cdot^n\rangle), \dot{\Psi}_s\rangle \, \mathrm{d}s\bigg].$$

Therefore, the same argumentation as in (i) with the uniform integrability in (4.45) and the boundedness of $|\dot{\Psi}_s|$ as a continuous function with compact support yield

$$\lim_{m,n\to\infty} \mathbb{E}[I_5^{m,n}(t \wedge \mathcal{T})] = \mathbb{E}\bigg[\int_0^{t\wedge\mathcal{T}} \int_{\mathbb{R}} |\tilde{X}(s,x)| \dot{\Psi}_s(x) \, \mathrm{d}x \, \mathrm{d}s\bigg].$$

□

4.5.3 Key argument: Bounding the quadratic variation term

What is left to bound in line (4.33), is the expectation of the quadratic variation term $I_4^{m,n}$. The main ingredient to be able to do this, will be the following Theorem 4.24.
Let us first introduce some definitions that we need to formulate the Theorem 4.24. Recall the definition of T_K in (4.37). Moreover, we define a semimetric on $[0,T]\times\mathbb{R}$ by

$$d((t,x),(t',x')) := |t-t'|^\alpha + |x-x'|, \quad t,t' \in [0,T], x,x' \in \mathbb{R},$$

and, for $K>0$, $N\in\mathbb{N}$ and $\zeta\in(0,1)$, the set

$$Z_{K,N,\zeta} := \left\{ (t,x) \in [0,T]\times[-1/2,1/2] : \begin{array}{l} t \le T_K,\ |x| \le 2^{-N\alpha-1}, \\ |t-\hat{t}| \le 2^{-N}\ |x-\hat{x}| \le 2^{-N\alpha}, \\ \text{for some } (\hat{t},\hat{x}) \in [0,T_K]\times[-1/2,1/2] \\ \text{satisfying } |\tilde{X}(\hat{t},\hat{x})| \le 2^{-N\zeta} \end{array} \right\}. \tag{4.46}$$

The following theorem improves the regularity of $\tilde{X}(t,x)$ when $|x|$ is small. For two measures $\mathbb{Q}_1$ and $\mathbb{Q}_2$ on some measurable space $(\tilde{\Omega}, \tilde{\mathscr{F}})$, we call $\mathbb{Q}_1$ absolutely continuous with respect to $\mathbb{Q}_2$, denoted by $\mathbb{Q}_1 \ll \mathbb{Q}_2$, if $\mathscr{N}_1 \supseteq \mathscr{N}_2$, where $\mathscr{N}_i \in \tilde{\mathscr{F}}$ denotes the zero sets of $\mathbb{Q}_i$ in $(\tilde{\Omega}, \tilde{\mathscr{F}})$.

Theorem 4.24. *Suppose Assumption 4.1 and let $\tilde{X} := X^1 - X^2$, where X^i is a solution of the SPDE (4.13) with $X^i \in C([0,T]\times\mathbb{R})$ a.s. for $i=1,2$. Let $\zeta\in(0,1)$ satisfy:*

$$\exists N_\zeta = N_\zeta(K,\omega) \in \mathbb{N} \text{ a.s. such that, for any } N \ge N_\zeta \text{ and any } (t,x) \in Z_{K,N,\zeta}: \\ \left.\begin{array}{c} |t'-t| \le 2^{-N}, t' \le T_K \\ |y-x| \le 2^{-N\alpha} \end{array}\right\} \quad \Rightarrow \quad |\tilde{X}(t,x) - \tilde{X}(t',y)| \le 2^{-N\zeta}. \tag{4.47}$$

Let $\frac{1}{2}-\alpha < \zeta^1 < (\zeta\xi + \frac{1}{2} - \alpha)\wedge 1$. Then, there is an $N_{\zeta^1}(K,\omega,\zeta)\in\mathbb{N}$ a.s. such that, for any $N \ge N_{\zeta^1}$ and any $(t,x)\in Z_{K,N,\zeta^1}$:

$$\left.\begin{array}{c} |t'-t| \le 2^{-N}, t' \le T_K \\ |y-x| \le 2^{-N\alpha} \end{array}\right\} \quad \Rightarrow \quad |\tilde{X}(t,x) - \tilde{X}(t',y)| \le 2^{-N\zeta^1}. \tag{4.48}$$

Moreover, there is some measure $\mathbb{Q}^{X,K}$ on $(\Omega, \mathscr{F})$ such that $\mathbb{Q}^{X,K} \ll \mathbb{P}$ on $(\Omega, \mathscr{F})$ and $\mathbb{P} \ll \mathbb{Q}^{X,K}$ on $(\Omega, \mathscr{F}^K)$, where $\mathscr{F}^K := \{A \cap \{T_K \geq T\} \colon A \in \mathscr{F}\} \subseteq \mathscr{F}$ is the σ-algebra restricted to $\{T_K \geq T\}$, and there are constants $R > 1$ and $\delta, C, c_2 > 0$ depending on ζ and ζ^1 (not on K) and $N(K) \in \mathbb{N}$ such that

$$\mathbb{Q}^{X,K}(N_{\zeta^1} \geq N) \leq C\Big(\mathbb{Q}^{X,K}\Big(N_\zeta \geq \frac{N}{R}\Big) + Ke^{-c_2 2^{N\delta}}\Big) \tag{4.49}$$

for $N \geq N(K)$.

Proof of Theorem 4.24. From the assumptions of Theorem 4.24 and Assumption 4.1, we are given the variables $\alpha \in [0, \frac{1}{2})$, $\zeta \in (0,1)$, $\xi \in (\frac{1}{2(1-\alpha)}, 1]$ and $\zeta_1 < (\zeta\xi + \frac{1}{2} - \alpha) \wedge 1$. Moreover, fix arbitrary $(t,x), (t',y) \in [0, T_K] \times [-\frac{1}{2}, \frac{1}{2}]$ such that w.l.o.g. $t \leq t'$ and given some $N \geq N_\zeta$,

$$|t - t'| \leq \varepsilon := 2^{-N}, \quad |x| \leq 2^{-N\alpha} \quad \text{and} \quad |x - y| \leq 2^{-N\alpha}. \tag{4.50}$$

We define small numbers $\delta, \delta', \delta_1, \delta_2 > 0$ in the following way. We choose $\delta \in (0, \frac{1}{2} - \alpha)$ such that

$$\zeta_1 < \Big(\Big(\zeta\xi + \frac{1}{2} - \alpha\Big) \wedge 1\Big) - \alpha\delta < 1.$$

Fixing $\delta' \in (0, \delta)$, we choose $\delta_1 \in (0, \delta')$ sufficiently small that

$$\zeta_1 < \Big(\Big(\zeta\xi + \frac{1}{2} - \alpha\Big) \wedge 1\Big) - \alpha\delta + \alpha\delta_1 < 1. \tag{4.51}$$

Furthermore, we define $\delta_2 > 0$ sufficiently small such that

$$\delta' - \delta_2 > \delta_1, \tag{4.52}$$

and we set

$$p := \Big(\Big(\zeta\xi + \frac{1}{2} - \alpha\Big) \wedge 1\Big) - \alpha\Big(\frac{1}{2} - \alpha\Big) + \alpha\delta_1 \tag{4.53}$$

and

$$\hat{p} := p + \alpha(\delta' - \delta_2 - \delta_1) = \Big(\Big(\zeta\xi + \frac{1}{2} - \alpha\Big) \wedge 1\Big) - \alpha\Big(\frac{1}{2} - \alpha\Big) + \alpha(\delta' - \delta_2). \tag{4.54}$$

By (4.52), we see that $\hat{p} > p$.
Moreover, we introduce

$$D^{x,y,t,t'}(s) := |p_{t-s}(x) - p_{t'-s}(y)|^2 |\tilde{X}(s,0)|^{2\xi} \quad \text{and} \quad D^{x,t'}(s) := p_{t'-s}(x)^2 |\tilde{X}(s,0)|^{2\xi}. \tag{4.55}$$

Our goal is to bound the following expression, where we will explicitly determine the measure $\mathbb{Q}$ as in the statement of the theorem and the random variable $N_1 := N_1(\omega)$ (in (4.73)), later:

$$
\begin{aligned}
&\mathbb{Q}\Big(|\tilde{X}(t,x)-\tilde{X}(t,y)| \geq |x-y|^{\frac{1}{2}-\alpha-\delta}\varepsilon^p, (t,x)\in Z_{K,N,\zeta}, N\geq N_1\Big)\\
&\quad+\mathbb{Q}\Big(|\tilde{X}(t',x)-\tilde{X}(t,x)| \geq |t'-t|^{\alpha(\frac{1}{2}-\alpha-\delta)}\varepsilon^p,\ (t,x)\in Z_{K,N,\zeta}, N\geq N_1\Big)\\
&\quad\leq \mathbb{Q}\Big(|\tilde{X}(t,x)-\tilde{X}(t,y)| \geq |x-y|^{\frac{1}{2}-\alpha-\delta}\varepsilon^p, (t,x)\in Z_{K,N,\zeta}, N\geq N_1,\\
&\qquad\qquad \int_0^t D^{x,y,t,t}(s)\,\mathrm{d}s \leq |x-y|^{1-2\alpha-2\delta'}\varepsilon^{2p}\Big)\\
&\qquad+\mathbb{Q}\Big(|\tilde{X}(t',x)-\tilde{X}(t,x)| \geq |t'-t|^{\alpha(\frac{1}{2}-\alpha-\delta)}\varepsilon^p, (t,x)\in Z_{K,N,\zeta}, N\geq N_1,\\
&\qquad\qquad \int_t^{t'} D^{x,t'}(s)\,\mathrm{d}s+\int_0^t D^{x,x,t,t'}(s)\,\mathrm{d}s \leq (t'-t)^{2\alpha(\frac{1}{2}-\alpha-\delta')}\varepsilon^{2p}\Big)\\
&\qquad+\mathbb{Q}\Big(\int_0^t D^{x,y,t,t}(s)\,\mathrm{d}s > |x-y|^{1-2\alpha-2\delta'}\varepsilon^{2p}, (t,x)\in Z_{K,N,\zeta}, N\geq N_1\Big)\\
&\qquad+\mathbb{Q}\Big(\int_t^{t'} D^{x,t'}(s)\,\mathrm{d}s+\int_0^t D^{x,x,t,t'}(s)\,\mathrm{d}s > (t'-t)^{2\alpha(\frac{1}{2}-\alpha-\delta')}\varepsilon^{2p},\\
&\qquad\qquad (t,x)\in Z_{K,N,\zeta}, N\geq N_1\Big)\\
&\quad=: Q_1+Q_2+Q_3+Q_4.
\end{aligned}
\tag{4.56}
$$

We will proceed in three steps to prove the theorem:

Step (i): explicitly choosing a measure $\mathbb{Q}^{X,K}$ as in the statement of the theorem, such that Q_1 and Q_2 in (4.56) fulfill $Q_1+Q_2 \leq ce^{-c'|t'-t|^{-2\alpha\delta''}}$ for some $c,c'>0$,

Step (ii): showing that $Q_3 = Q_4 = 0$ holds w.r.t. $\mathbb{P}$ (and hence also w.r.t. $\mathbb{Q}^{X,K}$, since $\mathbb{Q}^{X,K} \ll \mathbb{P}$), if we choose the random variable $N_1 := cN_\zeta$ for some large enough deterministic constant $c>0$,

Step (iii): completing the proof, using Step (i) and Step (ii).

Step (i): Consider first the term Q_1. Note that on the measurable space $(\Omega, \mathscr{F}^K)$, where the restricted σ-algebra $\mathscr{F}^K$ on $\{T_K \geq T\}$ is defined in the statement of the theorem, Assumption 4.1 (iii) yields the existence of some constant $C_K > 0$ such that

$$
\left|\frac{\mu(s,X^1(s,0))-\mu(s,X^2(s,0))}{\sigma(s,X^1(s,0))-\sigma(s,X^2(s,0))}\right| \leq C_K < \infty,
$$

for all $s \in [0,T]$ $\mathbb{P}$-a.s. on $(\Omega, \mathscr{F}^K)$ and, thus, we can apply Girsanov's theorem (see [KS91, Theorem 3.5.1]) with the adapted process $(L_t)_{t\in[0,T]}$ defined by

$$L_t := -\int_0^t \frac{\mu(s, X^1(s,0)) - \mu(s, X^2(s,0))}{\sigma(s, X^1(s,0)) - \sigma(s, X^2(s,0))} \,\mathrm{d}B_s,$$

whose stochastic exponential process $\mathscr{E}(L_t)$ is a martingale due to Novikov's condition (see [KS91, Proposition 3.5.12]). We define $\mathbb{Q}^{X,K}$ via the Radon–Nikodym derivative $\mathscr{E}(L_T)$ of the measure $\mathbb{Q}^{X,K}$ with respect to $\mathbb{P}$, under which the process $(\tilde{B}_t^{X,K})_{t\in[0,T]}$ is a Brownian motion, where $\tilde{B}_t^{X,K} = B_t - \langle B, L\rangle_t = B_t + A_t$ with $A_t := \int_0^t \frac{\mu(s,X^1(s,0))-\mu(s,X^2(s,0))}{\sigma(s,X^1(s,0))-\sigma(s,X^2(s,0))}\,\mathrm{d}s$ on $[0, T_K]$.
To avoid measurability problems we re-define $\mathbb{Q}^{X,K}$ as a measure on $(\Omega, \mathscr{F})$ by setting

$$\mathbb{Q}^{X,K}(A) := \mathbb{Q}^{X,K}(A \cap \{T_K \geq T\})$$

for $A \in \mathscr{F}$. Girsanov's theorem implies that $\mathbb{Q}^{X,K} \ll \mathbb{P}$ on $(\Omega, \mathscr{F})$ and $\mathbb{P} \ll \mathbb{Q}^{X,K}$ on $(\Omega, \mathscr{F}^K)$. With this notation, we see that

$$\begin{aligned}
&\tilde{X}(t,x) - \tilde{X}(t,y) \\
&\quad = \int_0^t p_{t-s}^\theta(x)\Big(\sigma(s, X^1(s,0)) - \sigma(s, X^2(s,0))\Big)\,\mathrm{d}(B_s + A_s) \\
&\quad\quad - \int_0^t p_{t-s}^\theta(y)\Big(\sigma(s, X^1(s,0)) - \sigma(s, X^2(s,0))\Big)\,\mathrm{d}(B_s + A_s) \\
&\quad = \int_0^t \big(p_{t-s}^\theta(x) - p_{t-s}^\theta(y)\big)\Big(\sigma(s, X^1(s,0)) - \sigma(s, X^2(s,0))\Big)\,\mathrm{d}\tilde{B}_s^{X,K}.
\end{aligned}$$

For fixed $t \in [0,T]$ and $x, y \in [-\frac{1}{2}, \frac{1}{2}]$, the process

$$S_{\tilde{t}}^{x,y} = \int_0^{\tilde{t}} (p_{t-s}^\theta(x) - p_{t-s}^\theta(y))(\sigma(s, X^1(s,0)) - \sigma(s, X^2(s,0)))\,\mathrm{d}\tilde{B}_s^{X,K}, \quad \tilde{t} \in [0,t],$$

is a local $\mathbb{Q}^{X,K}$-martingale with quadratic variation

$$\begin{aligned}
\langle S^{x,y}\rangle_{\tilde{t}} &= \int_0^{\tilde{t}} (p_{t-s}^\theta(x) - p_{t-s}^\theta(y))^2(\sigma(s, X^1(s,0)) - \sigma(s, X^2(s,0)))^2\,\mathrm{d}s \\
&\leq C_\sigma^2 \int_0^{\tilde{t}} (p_{t-s}^\theta(x) - p_{t-s}^\theta(y))^2 |\tilde{X}(s,0)|^{2\xi}\,\mathrm{d}s \\
&= C_\sigma^2 \int_0^{\tilde{t}} D^{x,y,t,t}(s)\,\mathrm{d}s.
\end{aligned}$$

Thus, working under $\mathbb{Q}^{X,K}$ in (4.56), we can bound the term Q_1 as follows:

$$\begin{aligned}
Q_1 &\leq \mathbb{Q}^{X,K}\bigg(|S_t^{x,y}| \geq |x-y|^{\frac{1}{2}-\alpha-\delta}\varepsilon^p,\ \int_0^t D^{x,y,t,t}(s)\,\mathrm{d}s \leq |x-y|^{1-2\alpha-2\delta'}\varepsilon^{2p}\bigg) \\
&\leq \mathbb{Q}^{X,K}\big(|S_t^{x,y}| \geq |x-y|^{\frac{1}{2}-\alpha-\delta}\varepsilon^p,\ \langle S^{x,y}\rangle_t \leq C_\sigma^2|x-y|^{1-2\alpha-2\delta'}\varepsilon^{2p}\big)
\end{aligned}$$

by the definition of $D^{x,y,t,t}$.
Next, we apply the Dambis–Dubins–Schwarz theorem, which states that the local $\mathbb{Q}^{X,K}$-martingale $S^{x,y}_{\tilde{t}}$ can be embedded into a $\mathbb{Q}^{X,K}$-Brownian motion $(\tilde{W}_{\tilde{t}})_{\tilde{t}\in[0,t]}$ such that $S^{x,y}_{\tilde{t}} = \tilde{W}_{\langle S^{x,y}\rangle_{\tilde{t}}}$ holds for all $\tilde{t}\in[0,t]$. Thus, with $z := C_\sigma^2|x- \ y|^{1-2\alpha-2\delta'}\varepsilon^{2p}$ we obtain

$$
\begin{aligned}
Q_1 &\le \mathbb{Q}^{X,K}\Big(|\tilde{W}_{\langle S^{x,y}\rangle_t}| \ge |x-y|^{\frac{1}{2}-\alpha-\delta}\varepsilon^p,\ \langle S^{x,y}\rangle_t \le z\Big)\\
&\le \mathbb{Q}^{X,K}\Big(\sup_{0\le s\le z}|\tilde{W}_s| \ge |x-y|^{\frac{1}{2}-\alpha-\delta}\varepsilon^p\Big),
\end{aligned}
$$

since from the first event follows always the second one. Thus, with the notation $\tilde{W}^*(t) := \sup_{0\le s\le t}|\tilde{W}_s|$, the scaling property of Brownian motion and the reflection principle, we get

$$
\begin{aligned}
Q_1 &\le \mathbb{Q}^{X,K}\big(\tilde{W}^*(C_\sigma^2|x-y|^{1-2\alpha-2\delta'}\varepsilon^{2p}) \ge |x-y|^{\frac{1}{2}-\alpha-\delta}\varepsilon^p\big)\\
&= \mathbb{Q}^{X,K}\big(\tilde{W}^*(1)C_\sigma|x-y|^{\frac{1}{2}-\alpha-\delta'}\varepsilon^p \ge |x-y|^{\frac{1}{2}-\alpha-\delta}\varepsilon^p\big)\\
&= 2\mathbb{Q}^{X,K}\big(\tilde{W}(1) \ge C_\sigma^{-1}|x-y|^{-\delta''}\big)
\end{aligned}
$$

with $\delta'' := \delta-\delta' > 0$ and, applying the concentration inequality $\mathbb{Q}^{X,K}(N>a) \le e^{-\frac{a^2}{2}}$ for standard normal distributed N, we get

$$
Q_1 \le 2e^{-\frac{1}{2C_\sigma^2}|x-y|^{-2\delta''}} =: ce^{-c'|x-y|^{-2\delta''}}, \tag{4.57}
$$

for some constants $c, c' > 0$. With a very similar argumentation, we can use the probability measure $\mathbb{Q}^{X,K}$ and proceed as above to derive the bound

$$
Q_2 \le ce^{-c'|t'-t|^{-2\alpha\delta''}},
$$

where c and c' are the same constants as in (4.57).

Step (ii): We want to show that the terms Q_3 and Q_4 in (4.56) vanish $\mathbb{P}$-a.s., if we choose N_1 large enough. Therefore, we consider $(t,x)\in Z_{K,N,\zeta}$ and (t',y) as in (4.50) and begin by showing the following bound on $|\tilde{X}(s,0)|$ for $s\le t'$:

$$
|\tilde{X}(s,0)| \le \begin{cases} 3\varepsilon^\zeta & \text{if } s\in[t-\varepsilon,t'],\\ (4+K)2^{\zeta N_\zeta}(t-s)^\zeta & \text{if } s\in[0,t-\varepsilon]. \end{cases} \tag{4.58}
$$

To see (4.58), we choose for $(t,x)\in Z_{K,N,\zeta}$ some $(\hat{t},\hat{x})$ as in the definition of $Z_{K,N,\zeta}$ in (4.46) such that

$$
|t-\hat{t}| \le \varepsilon = 2^{-N}, \quad |x-\hat{x}| \le \varepsilon^\alpha \quad \text{and} \quad |\tilde{X}(\hat{t},\hat{x})| \le 2^{-N\zeta} = \varepsilon^\zeta.
$$

Then, for $s\in[t-\varepsilon,t']$, we see that $|t-s|\le\varepsilon$ by (4.50). Thus, by (4.47), we obtain that

$$
\begin{aligned}
|\tilde{X}(s,0)| &\le |\tilde{X}(\hat{t},\hat{x})| + |\tilde{X}(\hat{t},\hat{x})-\tilde{X}(t,x)| + |\tilde{X}(t,x)-\tilde{X}(s,0)|\\
&\le 3\cdot 2^{-N\zeta} = 3\varepsilon^\zeta.
\end{aligned}
$$

For $s \in [t-2^{-N_\zeta}, t-\varepsilon]$, we can choose some $\tilde{N} \geq N_\zeta$ such that $2^{-(\tilde{N}+1)} \leq t-s \leq 2^{-\tilde{N}}$ due to $t-\varepsilon \geq s$, i.e. $t-s \geq 2^{-N}$. Thus, we get

$$\begin{aligned}
|\tilde{X}(s,0)| &\leq |\tilde{X}(\hat{t},\hat{x})| + |\tilde{X}(\hat{t},\hat{x}) - \tilde{X}(t,x)| + |\tilde{X}(t,x) - \tilde{X}(s,0)| \\
&\leq 2^{-N\zeta} + 2^{-N\zeta} + 2^{-\tilde{N}\zeta} \leq 2 \cdot (t-s)^\zeta + 2^\zeta 2^{-(\tilde{N}+1)\zeta} \\
&\leq 4(t-s)^\zeta.
\end{aligned}$$

Last, for $s \in [0, t-2^{-N_\zeta}]$ with $s \leq T_K$, i.e. $\tilde{X}$ is bounded by $K > 0$, and $t-s \geq 2^{-N_\zeta}$, we can bound

$$|\tilde{X}(s,0)| \ \leq K \leq K(t-s)^{-\zeta}(t-s)^\zeta \leq K 2^{N_\zeta \zeta}(t-s)^\zeta,$$

which shows the bound (4.58).
For Q_3, using (4.58) and the definition of $D^{x,y,t,t'}$ in (4.55), we can bound the term inside Q_3 by

$$\begin{aligned}
\int_0^t D^{x,y,t,t}(s)\,\mathrm{d}s &\leq 3^{2\xi} \int_{t-\varepsilon}^t (p_{t-s}(x) - p_{t-s}(y))^2 \varepsilon^{2\zeta\xi}\,\mathrm{d}s \\
&\quad + (4+K)^{2\xi} 2^{2\xi\zeta N_\zeta} \int_0^{t-\varepsilon} (p_{t-s}(x) - p_{t-s}(y))^2 (t-s)^{2\zeta\xi}\,\mathrm{d}s \\
&=: D_1(t) + D_2(t).
\end{aligned} \tag{4.59}$$

Now, by Lemma 4.10 with $\beta = \frac{1}{2} - \alpha - \delta'$ and $\max(|x|,|y|) \leq 2\varepsilon^\alpha$, we can bound

$$\begin{aligned}
D_1(t) &\lesssim \varepsilon^{2\zeta\xi}|x-y|^{1-2\alpha} \max(|x|,|y|)^{(\frac{1}{\alpha}-1)2\beta} \\
&\lesssim \varepsilon^{2\zeta\xi+2\delta'}|x-y|^{1-2\alpha-2\delta'}\varepsilon^{(1-\alpha)2\beta} \\
&= \varepsilon^{2(\frac{1}{2}-\alpha(\frac{3}{2}-\alpha)+\alpha\delta'+\xi\zeta)}|x-y|^{1-2\alpha-2\delta'} \\
&\leq \varepsilon^{2\hat{p}}|x-y|^{1-2\alpha-2\delta'}
\end{aligned} \tag{4.60}$$

by the definition of $\hat{p}$ in (4.54). For $D_2(t)$, we use Lemma 4.7 with $\beta = 1$ to bound

$$\begin{aligned}
D_2(t) &\lesssim 2^{2\xi\zeta N_\zeta} \int_0^{t-\varepsilon} |x-y|^2 (t-s)^{2\zeta\xi-2\alpha-2} \varepsilon^{2(1-\alpha)}\,\mathrm{d}s \\
&= 2^{2\xi\zeta N_\zeta}|x-y|^{1-2\alpha-2\delta'}|x-y|^{1+2\alpha+2\delta'}\varepsilon^{2(1-\alpha)}\left[\frac{(t-s)^{-2\alpha-1+2\xi\zeta}}{-2\alpha-1+2\xi\zeta}\right]_0^{t-\varepsilon} \\
&\lesssim 2^{2\xi\zeta N_\zeta}|x-y|^{1-2\alpha-2\delta'}\varepsilon^{\alpha(1+2\alpha+2\delta')}\varepsilon^{2(1-\alpha)}\varepsilon^{((-2\alpha-1+2\xi\zeta)\wedge 0)-2\alpha\delta_2} \\
&= 2^{2\xi\zeta N_\zeta}|x-y|^{1-2\alpha-2\delta'}\varepsilon^{2\hat{p}}.
\end{aligned} \tag{4.61}$$

Hence, by inserting (4.60) and (4.61) into (4.59), we obtain

$$\int_0^t D^{x,y,t,t}(s)\,\mathrm{d}s \lesssim 2^{2\xi\zeta N_\zeta}|x-y|^{1-2\alpha-2\delta'}\varepsilon^{2\hat{p}}. \tag{4.62}$$

For Q_4, we can use (4.58) to bound the first summand in the definition of Q_4 by

$$\begin{aligned}\int_t^{t'} D^{x,t'}(s)\,\mathrm{d}s &= \int_t^{t'} p_{t'-s}(x)^2|\tilde{X}(s,0)|^{2\xi}\,\mathrm{d}s\\ &\lesssim \int_t^{t'} (t'-s)^{-2\alpha}\varepsilon^{2\zeta\xi}\,\mathrm{d}s\\ &\lesssim \varepsilon^{2\zeta\xi}|t'-t|^{1-2\alpha}\\ &\lesssim \varepsilon^{2\xi\zeta}\varepsilon^{2(\frac{1}{2}-\alpha-\alpha(\frac{1}{2}-\alpha)+\alpha\delta')}|t'-t|^{2\alpha(\frac{1}{2}-\alpha-\delta')}\\ &\lesssim \varepsilon^{2\hat{p}}|t'-t|^{2\alpha(\frac{1}{2}-\alpha-\delta')},\end{aligned} \tag{4.63}$$

where we used that $|t-t'|\leq\varepsilon$ and $\hat{p}<\frac{1}{2}-\alpha-\alpha(\frac{1}{2}-\alpha)+\alpha\delta'$. We split the second summand similar as before:

$$\int_0^t D^{x,x,t,t'}(s)\,\mathrm{d}s = \int_{t-\varepsilon}^t D^{x,x,t,t'}(s)\,\mathrm{d}s + \int_0^{t-\varepsilon} D^{x,x,t,t'}(s)\,\mathrm{d}s =: D_3(t)+D_4(t). \tag{4.64}$$

By Lemma 4.9, we estimate

$$\begin{aligned}D_3(t) &= \int_{t-\varepsilon}^t |p_{t-s}(x)-p_{t'-s}(x)|^2|\tilde{X}(s,0)|^{2\xi}\,\mathrm{d}s\\ &\lesssim \varepsilon^{2\xi\zeta}|t'-t|^{1-2\alpha}\\ &\lesssim \varepsilon^{2\hat{p}}|t'-t|^{2\alpha(\frac{1}{2}-\alpha-\delta')},\end{aligned} \tag{4.65}$$

where the last estimate follows as in (4.63).
For $D_4(t)$, using the inequality $(a+b)^2\leq 2(a^2+b^2)$, we obtain

$$\begin{aligned}D_4(t) &= \int_0^{t-\varepsilon} |p_{t-s}(x)-p_{t'-s}(x)|^2|\tilde{X}(s,0)|^{2\xi}\,\mathrm{d}s\\ &\leq 2(4+K)^{2\xi}2^{2\xi\zeta N_\zeta}\int_0^{t-\varepsilon}\left|((t-s)^{-\alpha}-(t'-s)^{-\alpha})e^{-\frac{|x|^{1/\alpha}}{t-s}}\right|^2(t-s)^{2\xi\zeta}\,\mathrm{d}s\\ &\quad+2(4+K)^{2\xi}2^{2\xi\zeta N_\zeta}\int_0^{t-\varepsilon}\left|(t'-s)^{-\alpha}\left(e^{-\frac{|x|^{1/\alpha}}{t-s}}-e^{-\frac{|x|^{1/\alpha}}{t'-s}}\right)\right|^2(t-s)^{2\xi\zeta}\,\mathrm{d}s\\ &=: D_{4,1}+D_{4,2}.\end{aligned} \tag{4.66}$$

For $D_{4,1}$, we use the inequality

$$((t-s)^{-\alpha}-(t'-s)^{-\alpha})e^{-\frac{|x|^{1/\alpha}}{t-s}} \leq (t-s)^{-\alpha-1}(t'-t). \tag{4.67}$$

To see this, note that

$$e^{-\frac{|x|^{1/\alpha}}{t-s}} \leq \left(\frac{t-s}{t'-s}\right)^{\alpha}e^{-\frac{|x|^{1/\alpha}}{t-s}}+\frac{t'-t}{t-s},$$

which holds since

$$\left(\frac{t-s}{t'-s}\right)^{\alpha}+\frac{t'-t}{t-s}\geq\frac{t-s}{t'-s}+\frac{t'-t}{t-s}$$
$$=\frac{t-s}{t'-s}+\frac{t'-s}{t-s}-1\geq 1 \tag{4.68}$$

as $x\mapsto\frac{1}{x}+x\geq 2$ on $[0,1]$. Thus, using (4.67), we get

$$\begin{aligned}D_{4,1}&\lesssim 2^{2\xi\zeta N_\zeta}\int_0^{t-\varepsilon}(t-s)^{-2\alpha-2}(t'-t)^2(t-s)^{2\xi\zeta}\,\mathrm{d}s\\&\lesssim 2^{2\xi\zeta N_\zeta}(t'-t)^2\varepsilon^{((-2\alpha-1+\xi\zeta)\wedge 0)-2\alpha\delta_2}\\&\lesssim 2^{2\xi\zeta N_\zeta}(t'-t)^{2\alpha(\frac12-\alpha-\delta')}\varepsilon^{2-2\alpha(\frac12-\alpha-\delta')}\varepsilon^{((-2\alpha-1+\xi\zeta)\wedge 0)-2\alpha\delta_2}\\&=2^{2\xi\zeta N_\zeta}(t'-t)^{\alpha(1-2\alpha-2\delta')}\varepsilon^{2((-\alpha+\frac12+\xi\zeta)\wedge 1)-\alpha\delta_2-\alpha(\frac12-\alpha-\delta')}\\&=2^{2\xi\zeta N_\zeta}(t'-t)^{\alpha(1-2\alpha-2\delta')}\varepsilon^{2\hat p}.\end{aligned}\tag{4.69}$$

For $D_{4,2}$, we use the inequality $|e^{-a}-e^{-b}|\leq|a-b|$ and then the bound $\frac{1}{t-s}-\frac{1}{t'-s}\leq\frac{t'-t}{(t-s)^2}$, which holds as in (4.68), to get

$$\begin{aligned}D_{4,2}&\lesssim 2^{2\xi\zeta N_\zeta}\int_0^{t-\varepsilon}(t'-s)^{-2\alpha}\left|\frac{|x|^{1/\alpha}}{t-s}-\frac{|x|^{1/\alpha}}{t'-s}\right|^2(t-s)^{2\xi\zeta}\,\mathrm{d}s\\&\lesssim 2^{2\xi\zeta N_\zeta}|x|^{2/\alpha}\int_0^{t-\varepsilon}(t'-s)^{-2\alpha}(t-s)^{-4}(t'-t)^2(t-s)^{2\xi\zeta}\,\mathrm{d}s\\&\lesssim 2^{2\xi\zeta N_\zeta}|x|^{2/\alpha}\varepsilon^{-3-2\alpha+2\xi\zeta}(t'-t)^2\\&\lesssim 2^{2\xi\zeta N_\zeta}|x|^{2/\alpha}\varepsilon^{-3-2\alpha+2\xi\zeta}(t'-t)^{2\alpha(\frac12-\alpha-\delta')}\varepsilon^{2-2\alpha(\frac12-\alpha-\delta')}\\&=2^{2\xi\zeta N_\zeta}|x|^{2/\alpha}\varepsilon^{2(\frac12-\alpha+\xi\zeta-\alpha(\frac12-\alpha)+\alpha\delta')}(t'-t)^{\alpha(1-2\alpha-2\delta')}\\&=2^{2\xi\zeta N_\zeta}|x|^{2/\alpha}\varepsilon^{2\hat p}(t'-t)^{\alpha(1-2\alpha-2\delta')}.\end{aligned}\tag{4.70}$$

Hence, (4.63) and plugging (4.65), (4.66), (4.69) and (4.70) into (4.64), we obtain

$$\int_t^{t'}D^{x,t'}(s)\,\mathrm{d}s+\int_0^t D^{x,x,t,t'}(s)\,\mathrm{d}s\lesssim 2^{2\xi\zeta N_\zeta}|t'-t|^{\alpha(1-2\alpha-2\delta')}\varepsilon^{2\hat p}. \tag{4.71}$$

Combining (4.62) and (4.71), we can denote $C>0$ to be the maximum of the two generic constants occuring in the estimates, to conclude, that if we can secure that

$$C2^{2\xi\zeta N_\zeta}\varepsilon^{2\hat p}<\varepsilon^{2p}, \tag{4.72}$$

then the conditions inside of Q_3 and Q_4 are never fulfilled and, thus, we get that $Q_3=Q_4=0$. By $\varepsilon=2^{-N}$, (4.72) is equivalent to

$$C<2^{2N(\hat p-p)-2N_\zeta\xi\zeta},$$

and, since $\hat{p} - p > 0$, fulfilled for all

$$N > \frac{2\xi\zeta N_\zeta + log_2(C)}{2(\hat{p} - p)}.$$

Therefore, we can find a deterministic constant $c_{K,\zeta,\delta,\delta_1,\delta',\delta_2}$ such that, for all

$$N \geq N_1(\omega) := c_{K,\zeta,\delta,\delta_1,\delta',\delta_2} N_\zeta(\omega), \tag{4.73}$$

$Q_3 = Q_4 = 0$ holds.

Step (iii): We discretize $\tilde{X}(t,y)$ for $t \in [0, T_K]$ and $y \in [-\frac{1}{2}, \frac{1}{2}]$ as follows:

$$\begin{aligned} M_{n,N,K} := \max\Big\{ &\Big|\tilde{X}(j2^{-n}, (z+1)2^{-\alpha n}) - \tilde{X}(j2^{-n}, z2^{-\alpha n})\Big| \\ &+ \Big|\tilde{X}((j+1)2^{-n}, z2^{-\alpha n}) - \tilde{X}(j2^{-n}, z2^{-\alpha n})\Big| : \\ &|z| \leq 2^{\alpha n - 1}, (j+1)2^{-n} \leq T_K, j \in \mathbb{Z}_+, z \in \mathbb{Z}, \\ &(j2^{-n}, z2^{-\alpha n}) \in Z_{K,N,\zeta} \Big\}. \end{aligned}$$

Moreover, we define the event

$$A_N := \big\{\omega \in \Omega \,:\, \text{for some } n \geq N,\, M_{n,N,K} \geq 2^{-n\alpha(\frac{1}{2} - \alpha - \delta)} 2^{-Np},\, N \geq N_1\big\}.$$

Then, we get, by using (4.56), Step (i) and Step (ii), that for all $N \geq N_1$ as in (4.73):

$$\begin{aligned} \mathbb{Q}^{X,K}\Big(\bigcup_{N' \geq N} A_{N'}\Big) &\leq \sum_{N'=N}^{\infty} \sum_{n=N'}^{\infty} \mathbb{Q}^{X,K}(M_{n,N',K} \geq 2 \cdot 2^{-n\alpha(\frac{1}{2} - \alpha - \delta)} 2^{-Np}) \\ &\lesssim \sum_{N'=N}^{\infty} \sum_{n=N'}^{\infty} 2^{(\alpha+1)n} e^{-c' 2^{n\delta'' \alpha}}, \end{aligned}$$

since the total number of partition elements in each $M_{n,N,K}$ is at most $2 \cdot 2^{\alpha n - 1} \cdot K \cdot 2^n \lesssim K 2^{(\alpha+1)n}$ (if $T_K = T$). Furthermore, we used that $|t - \hat{t}| \leq 2^{-n}$ and $|x - \hat{x}| \leq 2^{-n\alpha}$, which follows by the construction of $M_{n,N,K}$.

We use the convexity $2^{x+y} \geq 2^x + 2^y$ for $x, y \geq 0$ to estimate

$$\begin{aligned}
\mathbb{Q}^{X,K}\Big(\bigcup_{N' \geq N} A_{N'}\Big) &\lesssim \sum_{N'=N}^{\infty}\sum_{n=0}^{\infty} 2^{(\alpha+1)(n+N')} e^{-c' 2^{(n+N')\delta''\alpha}} \\
&\leq \sum_{N'=N}^{\infty} 2^{(\alpha+1)N'} \sum_{n=0}^{\infty} 2^{(\alpha+1)n} e^{-c'(2^{n\delta''\alpha} + 2^{N'\delta''\alpha})} \\
&= \sum_{N'=N}^{\infty} 2^{(\alpha+1)N'} e^{-c' 2^{N'\delta''\alpha}} \sum_{n=0}^{\infty} 2^{(\alpha+1)n} e^{-c' 2^{n\delta''\alpha}} \\
&= 2^{(\alpha+1)N} e^{-c' 2^{N\delta''\alpha}} \sum_{N'=0}^{\infty} 2^{(\alpha+1)N'} e^{-c' 2^{N'\delta''\alpha}} \sum_{n=0}^{\infty} 2^{(\alpha+1)n} e^{-c' 2^{n\delta''\alpha}} \\
&\lesssim e^{(\alpha+1)N} e^{-c' 2^{N\delta''\alpha}} \\
&\lesssim e^{-c_2 2^{N\delta''\alpha}},
\end{aligned}$$

for some constant $c_2 > 0$, where we used convergence and thus finiteness of the two series in the fourth line by applying the ratio test

$$\lim_{n\to\infty} \left| 2^{\alpha+1} e^{-c'(2^{(n+1)\delta''\alpha} - 2^{n\delta''\alpha})} \right| = 0.$$

Therefore, we get for

$$N_2(\omega) := \min\{N \in \mathbb{N} : \omega \in A^c_{N'}\, \forall N' \geq N\},$$

where the superscript c denotes the complement of a set, that

$$\mathbb{Q}^{X,K}(N_2 > N) = \mathbb{Q}^{X,K}\Big(\bigcup_{N' \geq N} A_{N'}\Big) \lesssim e^{-c_2 2^{N\delta''\alpha}}, \tag{4.74}$$

and thus $N_2 < \infty$ $\mathbb{Q}^{X,K}$-a.s.
We fix some $m \in \mathbb{N}$ with $m > 3/\alpha$ and choose $N(\omega) \geq (N_2(\omega) + m) \wedge (N_1 + m)$, which is finite a.s., such that holds:

$$\forall n \geq N : M_{n,N,K} < 2^{-n\alpha(\frac{1}{2}-\alpha-\delta)} 2^{-Np} \quad \text{a.s.} \tag{4.75}$$

and $Q_3 = Q_4 = 0$.
Furthermore, we choose $(t, x) \in Z_{K,N,\zeta}$ and (t', y) such that

$$d((t', y), (t, x)) := |t' - t|^\alpha + |y - x| \leq 2^{-N\alpha},$$

and we choose points near (t, x) as follows: for $n \geq N$, we denote by $t_n \in 2^{-n}\mathbb{Z}_+$ and $x_n \in 2^{-\alpha n}\mathbb{Z}$ for the unique points such that

$$\begin{aligned}
&t_n \leq t < t_n + 2^{-n}, \\
&x_n \leq x < x_n + 2^{-\alpha n} \text{ for } x \geq 0 \quad \text{or} \quad x_n - 2^{-\alpha n} < x \leq x_n \text{ for } x < 0.
\end{aligned}$$

We define t'_n, y_n analogously. Let $(\hat{t}, \hat{x})$ be the points from the definition of $Z_{K,N,\zeta}$ with $|\tilde{X}(\hat{t}, \hat{x})| \leq 2^{-N\zeta}$. Then, for $n \geq N$, we observe that

$$\begin{aligned}
d((t'_n, y_n), (\hat{t}, \hat{x})) &\leq d((t'_n, y_n), (t', y)) + d((t', y), (t, x)) + d((t, x), (\hat{t}, \hat{x})) \\
&\leq |t'_n - t|^\alpha + |y - y_n| + 2^{-N\alpha} + 2 \cdot 2^{-N\alpha} \\
&\leq 6 \cdot 2^{-N\alpha} < 2^{3-N\alpha} = 2^{-\alpha(N - \frac{3}{\alpha})} \\
&< 2^{-\alpha(N-m)},
\end{aligned} \tag{4.76}$$

which implies $(t'_n, y_n) \in Z_{K,N-m,\zeta}$. We use that to finally formulate our bound. We also use the continuity of $\tilde{X}$ and our construction of the t_n, x_n to get that

$$\lim_{n \to \infty} \tilde{X}(t_n, x_n) = \tilde{X}(t, x) \quad \text{a.s.}$$

and the same for t'_n, y_n. Thus, by the triangle inequality:

$$\begin{aligned}
|\tilde{X}(t, x) - \tilde{X}(t', y)| &= \Bigg| \sum_{n=N}^{\infty} \Big((\tilde{X}(t_{n+1}, x_{n+1}) - \tilde{X}(t_n, x_n)) + (\tilde{X}(t'_n, y_n) - \tilde{X}(t'_{n+1}, y_{n+1})) \Big) \\
&\qquad + \tilde{X}(t_N, x_N) - \tilde{X}(t'_N, y_N) \Bigg| \\
&\leq \sum_{n=N}^{\infty} |\tilde{X}(t_{n+1}, x_{n+1}) - \tilde{X}(t_n, x_n)| + |\tilde{X}(t'_n, y_n) - \tilde{X}(t'_{n+1}, y_{n+1})| \\
&\qquad + |\tilde{X}(t_N, x_N) - \tilde{X}(t'_N, y_N)|.
\end{aligned}$$

Since we choose t_n, x_n and t'_n, y_n to be of the form of the discrete points in $M_{n,N,K}$ and, since we have (4.76), we can continue to estimate

$$|\tilde{X}(t, x) - \tilde{X}(t', y)| \leq \sum_{n=N}^{\infty} 2M_{n+1,N-m,K} + |\tilde{X}(t_N, x_N) - \tilde{X}(t'_N, y_N)|.$$

Because of $|t - t'| \leq 2^{-N}$ and our construction of t_N, t'_N, they must be equal or adjacent in $2^{-N}\mathbb{Z}_+$ and analogue for x_N, y_N. Thus, we get

$$\begin{aligned}
|\tilde{X}(t, x) - \tilde{X}(t', y)| &\leq \sum_{n=N}^{\infty} 2M_{n+1,N-m,K} + M_{N,N-m,K} \\
&\leq 2 \sum_{n=N}^{\infty} M_{n,N-m,K} \\
&\lesssim \sum_{n=N}^{\infty} 2^{-n\alpha(\frac{1}{2} - \alpha - \delta)} 2^{-(N-m)p} \\
&= 2^{-(N-m)p} \sum_{n=0}^{\infty} 2^{-(n+N)\alpha(\frac{1}{2} - \alpha - \delta)}
\end{aligned}$$

$$\lesssim 2^{mp} 2^{-N(\alpha(\frac{1}{2}-\alpha-\delta)+p)}$$
$$< 2^{-N\zeta_1},$$

where the last line follows with $\alpha(\frac{1}{2}-\alpha-\delta)+p > \zeta_1$, which holds by (4.51) and (4.53), and for all

$$N \geq N_3 \tag{4.77}$$

for some N_3 that is large enough such that 2^{mp} is dominated and thus depends deterministically on p. Therefore, we have proven Theorem 4.24 with

$$N_{\zeta_1}(\omega) := \max\{N_2(\omega)+m, N_\zeta(\omega)+m, c_{K,\zeta,\delta,\delta_1,\delta',\delta_2} N_\zeta(\omega)+m, N_3\}$$

by N_{ζ_1} chosen in that way due to (4.75), Step (ii), (4.73) and (4.77). If we denote $R' := 1 \vee c_{K,\zeta,\delta,\delta_1,\delta',\delta_2}$ and consider some $N \geq 2m \vee N_3$, (4.74) implies

$$\begin{aligned}
\mathbb{Q}^{X,K}(N_{\zeta_1} \geq N) &\leq \mathbb{Q}^{X,K}(N_2 \geq N-m) + 2\mathbb{Q}^{X,K}\left(N_\zeta \geq \frac{N-m}{R'}\right) \\
&\leq CKe^{-c_2 2^{(N-m)\delta''\alpha}} + 2\mathbb{Q}^{X,K}(N_\zeta \geq N/R)
\end{aligned}$$

for $R = 2R'$ and $C > 0$ not depending on K, which shows the probability bound in (4.49) by re-defining $\delta := \delta''\alpha > 0$ and thus completes the proof. □

In the following we sometimes only write a.s. when we mean $\mathbb{P}$-a.s. Since $\mathbb{Q}^{X,K} \ll \mathbb{P}$, this implies $\mathbb{Q}^{X,K}$-a.s.

Corollary 4.25. *With the hypotheses of Theorem 4.24 and $\frac{1}{2}-\alpha < \zeta < \frac{\frac{1}{2}-\alpha}{1-\xi} \wedge 1$, there is an a.s. finite positive random variable $C_{\zeta,K}(\omega)$ such that, for any $\varepsilon \in (0,1]$, $t \in [0,T_K]$ and $|x| < \varepsilon^\alpha$, if $|\tilde{X}(t,\hat{x})| \leq \varepsilon^\zeta$ for some $|\hat{x}-x| \leq \varepsilon^\alpha$, then*

$$|\tilde{X}(t,y)| \leq C_{\zeta,K}\varepsilon^\zeta, \tag{4.78}$$

whenever $|x-y| \leq \varepsilon^\alpha$.
Moreover, there are constants $\delta, C_1, c_2, \tilde{R} > 0$, depending on ζ (but not on K), and $r_0(K) > 0$ such that

$$\mathbb{Q}^{X,K}(C_{\zeta,K} \geq r) \leq C_1\left[\mathbb{Q}^{X,K}\left(N_{\frac{\alpha}{2}(\frac{1}{2}-\alpha)} \geq \frac{1}{\tilde{R}}\log_2\left(\frac{r-6}{K+1}\right)\right) + Ke^{-c_2\left(\frac{r-6}{K+1}\right)^\delta}\right] \tag{4.79}$$

for all $r \geq r_0(K) > 6+(K+1)$, where $\mathbb{Q}^{X,K}$ is the probability measure from Theorem 4.24.

Proof. We will derive the statement by an appropriate induction. We start by choosing

$$\zeta_0 := \frac{\alpha}{2}\left(\frac{1}{2}-\alpha\right),$$

to be able to use the regularity result from Proposition 4.11. Indeed, by 4.11 (ii) we get the inequality (4.47) with ζ_0 by Kolmogorov's continuity theorem.

Now, we define

$$\zeta_{n+1} := \left[\left(\zeta_n\xi + \frac{1}{2} - \alpha\right) \wedge 1\right]\left(1 - \frac{1}{n+d}\right)$$

for some $d \in \mathbb{R}$. We chose that d given ζ_0 big enough such that $\zeta_1 > \frac{1}{2} - \alpha$. Moreover, it is clearly $\zeta_{n+1} > \zeta_n$. Thus, we get inductively that $\zeta_n \uparrow \frac{\frac{1}{2}-\alpha}{1-\xi} \wedge 1$ and, for every fixed $\zeta \in \left(\frac{1}{2} - \alpha, \frac{\frac{1}{2}-\alpha}{1-\xi} \wedge 1\right)$ as in the statement, we can find $n_0 \in \mathbb{N}$ such that $\zeta_{n_0} \geq \zeta > \zeta_{n_0-1}$. By applying Theorem 4.24 n_0-times, we get (4.47) for ζ_{n_0-1} and, hence, (4.48) for ζ_{n_0}. We derive the estimation (4.78) for all $0 < \varepsilon \leq 1$. Therefore, we consider first $\varepsilon \leq 2^{-N_{\zeta_{n_0}}}$, where we got $N_{\zeta_{n_0}}$ from the application of Theorem 4.24 to ζ_{n_0-1}. Further, we choose $N \in \mathbb{N}$ such that $2^{-N-1} < \varepsilon \leq 2^{-N}$ and, thus, $N \geq N_{\zeta_{n_0}}$. Also, we choose $t \leq T_K$ and $|x| \leq \varepsilon^\alpha \leq 2^{-N\alpha}$ such that, by assumption of Theorem 4.24, for some $|\hat{x} - x| \leq \varepsilon^\alpha \leq 2^{-N\alpha}$,

$$|\tilde{X}(t,\hat{x})| \leq \varepsilon^\zeta \leq 2^{-N\zeta} \leq 2^{-N\zeta_{n_0-1}}.$$

Hence, $(t,x) \in Z_{K,N,\zeta_{n_0-1}}$. For any y such that $|y - x| \leq \varepsilon^\alpha$, we get, by (4.48),

$$\begin{aligned}|\tilde{X}(t,y)| &\leq |\tilde{X}(t,\hat{x})| + |\tilde{X}(t,\hat{x}) - \tilde{X}(t,x)| + |\tilde{X}(t,x) - \tilde{X}(t,y)| \\ &\leq 2^{-N\zeta} + 2^{-N\zeta_{n_0}} + 2^{-N\zeta_{n_0}} \leq 3 \cdot 2^{-N\zeta} \leq 6\varepsilon^\zeta.\end{aligned}$$

Now, we consider $\varepsilon \in (2^{-N_{\zeta_{n_0}}}, 1]$. Then, for (t,x) and (t,y) as in the assumption, we get

$$\begin{aligned}|\tilde{X}(t,y)| &\leq |\tilde{X}(t,x)| + |\tilde{X}(t,y) - \tilde{X}(t,x)| \\ &\leq K + 2^{-N\zeta} \leq (K+1)2^{N_{\zeta_{n_0}}\zeta}\varepsilon^\zeta\end{aligned}$$

by $\varepsilon 2^{N_{\zeta_{n_0}}} > 1$ and, therefore, we have shown (4.78) with $C_{\zeta,K} = (K+1)2^{N_{\zeta_{n_0}}\zeta} + 6$.
It remains to show the estimate (4.79). Therefore, we use (4.49) to conclude that

$$\begin{aligned}\mathbb{Q}^{X,K}\Big(C_{\zeta,K} \geq r\Big) &= \mathbb{Q}^{X,K}\left(2^{N_{\zeta_{n_0}}\zeta} \geq \frac{r-6}{K+1}\right) = \mathbb{Q}^{X,K}\left(N_{\zeta_{n_0}} \geq \frac{1}{\zeta}\log_2\left(\frac{r-6}{K+1}\right)\right) \\ &\leq C\left(\mathbb{Q}^{X,K}\left(N_{\zeta_{n_0-1}} \geq \frac{1}{R\zeta}\log_2\left(\frac{r-6}{K+1}\right)\right) + K\exp\left(-c_2 2^{\frac{\delta}{\zeta}\log_2\left(\frac{r-6}{K+1}\right)}\right)\right).\end{aligned}$$

Applying (4.49) n_0-times, we end up with

$$\begin{aligned}&\mathbb{Q}^{X,K}v(C_{\zeta,K} \geq r) \\ &\leq C^{n_0}\mathbb{Q}^{X,K}\left(N_{\frac{\alpha}{2}(\frac{1}{2}-\alpha)} \geq \frac{1}{\zeta R^{n_0}}\log_2\left(\frac{r-6}{K+1}\right)\right) + \sum_{i=0}^{n_0} C^i K e^{-c_2 2^{R^{-i-1}\frac{\delta}{\zeta}\log_2\left(\frac{r-6}{K+1}\right)}} \\ &\leq C^{n_0}n_0\left(\mathbb{Q}^{X,K}\left(N_{\frac{\alpha}{2}(\frac{1}{2}-\alpha)} \geq \frac{1}{\tilde{R}}\log_2\left(\frac{r-6}{K+1}\right)\right) + Ke^{-c_2\left(\frac{r-6}{K+1}\right)^{\frac{\delta}{\zeta R^{n_0}}}}\right) \\ &=: C_1\left(\mathbb{Q}^{X,K}\left(N_{\frac{\alpha}{2}(\frac{1}{2}-\alpha)} \geq \frac{1}{\tilde{R}}\log_2\left(\frac{r-6}{K+1}\right)\right) + Ke^{-c_2\left(\frac{r-6}{K+1}\right)^{\tilde{\delta}}}\right),\end{aligned}$$

where $C_1, \tilde{\delta}, \tilde{R} > 0$ depend on ζ but not on K. □

We will handle the event on the right-hand side of (4.79) under the measure $\mathbb{P}$ again.

Proposition 4.26. *In the setup and notation of Corollary 4.25, one has*

$$\mathbb{P}\Big(N_{\frac{\alpha}{2}(\frac{1}{2}-\alpha)} \geq \frac{1}{\tilde{R}} \log_2\left(\frac{r-6}{K+1}\right)\Big) \lesssim \left(\frac{r-6}{K+1}\right)^{-\varepsilon},$$

for some $\varepsilon > 0$.

Proof. We show that, for every $M \in \mathbb{R}_+$,

$$\mathbb{P}\big(N_{\frac{\alpha}{2}(\frac{1}{2}-\alpha)} \geq M\big) \lesssim 2^{-M\varepsilon}$$

for some $\varepsilon > 0$, which then yields the statement.
Indeed, from Proposition 4.11 (ii), we have that

$$\mathbb{E}[|\tilde{X}(t,x) - \tilde{X}(t',x')|^p] \lesssim |t-t'|^{(\frac{1}{2}-\alpha)p} + |x-x'|^{(\frac{1}{2}-\alpha)p},$$

for all $p \geq 2$, $t,t' \in [0,T]$ and $|x|,|x'| \leq 1$. By choosing $(t,x) \in Z_{K,N,\zeta}$, (t',x') from the definition of $Z_{K,N,\zeta}$ and $p > 2$ such that $\alpha p(\frac{1}{2}-\alpha) = 1+\beta$ for some $\beta > 0$, it holds that

$$\mathbb{E}[|\tilde{X}(t,x) - \tilde{X}(t',x')|^p] \lesssim 2^{-N(1+\beta)} + 2^{-N(1+\beta)} \lesssim 2^{-N(1+\beta)}.$$

We discretize $[0,T]\times[-1,1]$ on the dyadic rational numbers. For simplicity, we assume $T = 1$. First, for some $n \in \mathbb{N}$, we keep some space variable $x \in \{k2^{-n}, k \in -2^n, \dots, 0, 1, \dots, 2^n\}$ fixed and apply Markov's inequality to get

$$\mathbb{P}\Big(|\tilde{X}(k2^{-n},x) - \tilde{X}((k-1)2^{-n},x)| \geq 2^{-\zeta n}\Big) \lesssim 2^{\zeta np}2^{-n(1+\beta)} = 2^{-n(1+\beta-\zeta p)}$$

for any $k \in 1,\dots,2^n$. Next, we define the following events:

$$\begin{aligned}
A_n = A_n(\zeta) &:= \Big\{\max_{k\in\{-2^n+1,\dots,2^n\}} |\tilde{X}(k2^{-n},x) - \tilde{X}((k-1)2^{-n},x)| \geq 2^{-\zeta n-1}\Big\},\\
B_n &:= \bigcup_{m=n}^{\infty} A_m, \quad N := \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} B_n.
\end{aligned}$$

Then, for every $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{P}(A_n) &\leq \sum_{k=-2^n+1}^{2^n} \mathbb{P}\Big(|\tilde{X}(k2^{-n},x) - \tilde{X}((k-1)2^{-n},x)| \geq 2^{-\zeta n-1}\Big)\\
&\lesssim 2^{n+2}2^{-n(1+\beta-\zeta p)+p} = 2^{2+p}2^{-n(\beta-\zeta p)}. \qquad (4.80)
\end{aligned}$$

We choose, for $\zeta = \frac{\alpha}{2}(\frac{1}{2}-\alpha)$,

$$p > \max\Big\{\frac{1+\beta}{\alpha(\frac{1}{2}-\alpha)}, \frac{1}{\frac{\alpha}{2}-\zeta-\alpha^2}\Big\}.$$

Note that $\frac{\alpha}{2} - \zeta - \alpha^2 = \frac{\alpha}{2} - \frac{\alpha}{2}(\frac{1}{2} - \alpha) - \alpha^2 = \frac{\alpha}{4} - \frac{\alpha^2}{2} > 0$ as $\alpha < \frac{1}{2}$. Then, we have that

$$0 < p\left(\frac{\alpha}{2} - \zeta - \alpha^2\right) - 1 = \alpha p\left(\frac{1}{2} - \alpha\right) - 1 - \zeta p = \beta - \zeta p \tag{4.81}$$

and from (4.80) it follows by the geometric series that

$$\mathbb{P}(B_n) \leq \sum_{m=n}^{\infty} \mathbb{P}(A_m) \lesssim 2^{2+p} \frac{2^{-n(\beta-\zeta p)}}{1 - 2^{\zeta p - \beta}} \to 0 \quad \text{as } n \to \infty,$$

where $2^{\zeta p - \beta} < 1$ because of (4.81).
Analogously, we fix some time variable t and get an analogue version of inequality (4.80). Now, we fix an event $\omega \in \Omega$ and some

$$N \geq N_{\frac{\alpha}{2}(\frac{1}{2}-\alpha)}(\omega),$$

where $N_{\frac{\alpha}{2}(\frac{1}{2}-\alpha)}(\omega)$ is such that

$$\omega \notin \bigcup_{n=N_{\frac{\alpha}{2}(\frac{1}{2}-\alpha)}}^{\infty} A_n,$$

and this should also hold for the union of the analogue sets for fixed t, denote those by $A_n^{(2)}$.
Let $t, t', x, x' \in D_N$ with $|t - t'| \leq 2^{-N}$ and $|x - x'| \leq 2^{-\alpha N}$. Then, we have

$$\begin{aligned} |\tilde{X}(t,x,\omega) - \tilde{X}(t',x',\omega)| &\leq |\tilde{X}(t,x,\omega) - \tilde{X}(t',x,\omega)| + |\tilde{X}(t',x,\omega) - \tilde{X}(t',x',\omega)| \\ &\leq 2 \cdot 2^{-\zeta N - 1} = 2^{-\zeta N}. \end{aligned}$$

Then, we get from (4.80) that

$$\mathbb{P}(N_\zeta \geq M) \leq \sum_{m=M}^{\infty} \mathbb{P}(A_m) + \sum_{m=M}^{\infty} \mathbb{P}(A_m^{(2)}) \lesssim \sum_{m=M}^{\infty} 2^{-m(\beta-\zeta p)} = \frac{2^{-M(\beta-\zeta p)}}{1 - 2^{\zeta p - \beta}} \lesssim 2^{-M\varepsilon}$$

with $\varepsilon := \beta - \zeta p$, by the geometric series with $\beta - \zeta p > 0$.
By the density of the dyadic rational numbers in the reals and the continuity of $\tilde{X}$, the regularity extends to the whole $[0,T] \times [-1,1]$ and, thus, the statement holds. □

We want to fix $\zeta \in (0,1)$, that fulfills the requirements of the previous corollary.

Lemma 4.27. *With fixed $\alpha \in (0, \frac{1}{2})$ and $\xi \in (\frac{1}{2}, 1)$ satisfying*

$$1 > \xi > \frac{1}{2(1-\alpha)} > \frac{1}{2},$$

we can choose $\zeta \in (0,1)$ such that

$$\frac{\alpha}{2\xi - 1} < \zeta < \left(\frac{\frac{1}{2} - \alpha}{1 - \xi} \wedge 1\right). \tag{4.82}$$

Especially, we get

$$\eta := \frac{\zeta}{\alpha} > \frac{1}{2\xi - 1}.$$

Proof. First, we consider $\frac{\frac{1}{2}-\alpha}{1-\xi} < 1$. In this case, we have that

$$\begin{aligned}\frac{\frac{1}{2} - \alpha}{1 - \xi} - \frac{\alpha}{2\xi - 1} &= \frac{(\frac{1}{2} - \alpha)(2\xi - 1) - \alpha(1 - \xi)}{(1 - \xi)(2\xi - 1)} \\ &= \frac{\xi - \frac{1}{2} - 2\alpha\xi + \alpha - \alpha + \alpha\xi}{(1 - \xi)(2\xi - 1)} = \frac{\xi(1 - \alpha) - \frac{1}{2}}{(1 - \xi)(2\xi - 1)} > 0,\end{aligned}$$

by the assumption on ξ.
On the other hand, if $\frac{\frac{1}{2}-\alpha}{1-\xi} \geq 1$, then $\alpha \leq \xi - \frac{1}{2}$, i.e. $\frac{\alpha}{2\xi-1} \leq \frac{1}{2}$, and we can fix ζ such that (4.82) holds. □

Let us finally introduce the following stopping time, that plays a central role for the following Lemma 4.29, and is the reason, why we needed Corollary 4.25 and Proposition 4.26:

$$T_{\zeta,K} := \inf_{t \geq 0} \left\{ \begin{array}{l} t \leq T_K \text{ and there exist } \varepsilon \in (0,1], \hat{x}, x, y \in \mathbb{R} \text{ with} \\ |x| \leq \varepsilon^{\alpha}, |\tilde{X}(t,\hat{x})| \leq \varepsilon^{\zeta}, |x - \hat{x}| \leq \varepsilon^{\alpha}, |x - y| \leq \varepsilon^{\alpha} \\ \text{ such that } |\tilde{X}(t,y)| > c_0(K)\varepsilon^{\zeta} \end{array} \right\} \wedge T_K \wedge T, \tag{4.83}$$

where $c_0(K) := r_0(K) \vee K^2 > 0$ with $r_0(k)$ from Corollary 4.25.

Corollary 4.28. *The stopping time $T_{\zeta,K}$ fulfills $T_{\zeta,K} \to T$ as $K \to \infty$ a.s.*

Proof. We fix arbitrary $K, \tilde{K} > 0$ such that $\tilde{K} \leq K$. We can bound for any $t \in [0,T)$,

$$\begin{aligned}\mathbb{P}\big(T_{\zeta,K} \leq t\big) &\leq \mathbb{P}\Big(\{T_{\zeta,K} \leq t\} \cap \{T_{\tilde{K}} \geq T\}\Big) + \mathbb{P}\big(T_{\tilde{K}} < T\big) \\ &=: P_1^{K,\tilde{K}} + P_2^{\tilde{K}}.\end{aligned} \tag{4.84}$$

We show that $\lim_{K\to\infty} P_1^{K,\tilde{K}} = 0$. For this purpose, we consider the probability measure $\mathbb{Q}^{X,\tilde{K}}$ from Corollary 4.25. By the definition of $T_{\zeta,K}$ and Corollary 4.25, we obtain that

$$\begin{aligned}&\mathbb{Q}^{X,\tilde{K}}\Big(\{T_{\zeta,K} \leq t\} \cap \{T_{\tilde{K}} \geq T\}\Big) \\ &\quad \leq \mathbb{Q}^{X,\tilde{K}}\big(T_K \leq t\big) + \mathbb{Q}^{X,\tilde{K}}\big(C_{\zeta,K} > c_0(K)\big) \\ &\quad \leq \mathbb{Q}^{X,\tilde{K}}\big(T_K \leq t\big) + C_1\bigg[\mathbb{Q}^{X,\tilde{K}}\Big(N_{\frac{\alpha}{2}(\frac{1}{2}-\alpha)} \geq \frac{1}{\tilde{R}}\log_2\Big(\frac{K^2 - 6}{\tilde{K} + 1}\Big)\Big) + \tilde{K}e^{-c_2\left(\frac{K^2-6}{\tilde{K}+1}\right)\delta}\bigg].\end{aligned} \tag{4.85}$$

By Proposition 4.26 we know that the respective of the second probability on the right-hand side of (4.85) with $\mathbb{P}$ instead of $\mathbb{Q}^{X,\tilde{K}}$ tends to zero as $K \to \infty$. Since $\mathbb{Q}^{X,\tilde{K}} \ll \mathbb{P}$ holds on $(\Omega, \mathscr{F})$, $\lim_{K\to\infty} \mathbb{P}(A_K) = 0$ implies $\lim_{K\to\infty} \mathbb{Q}^{X,\tilde{K}}(A_K) = 0$ for any sequence $(A_K)_{K\in\mathbb{N}}$ of events in Ω (see e.g. [Rud87, Theorem 6.11]) and, since $T_K \to \infty$ as $K \to \infty$ a.s., by the continuity of the solutions X^1 and X^2, we conclude that the whole right-hand side of (4.85) tends to zero as $K \to \infty$. Hence, since $\mathbb{P} \ll \mathbb{Q}^{X,\tilde{K}}$ on $(\Omega, \mathscr{F}^{\tilde{K}})$ and the event inside $P_1^{K,\tilde{K}}$ is trivially in $\mathscr{F}^{\tilde{K}}$, this implies also tending to zero for the respective $\mathbb{P}$-probability and we obtain $\lim\limits_{K\to\infty} P_1^{K,\tilde{K}} = 0$.
Therefore, using the continuity of X^1 and X^2 again, we can for every $\varepsilon > 0$ find some $\tilde{K} > 0$ such that (4.84) yields

$$\lim_{K\to\infty} \mathbb{P}\big(T_{\zeta,K} \leq t\big) \leq \mathbb{P}\big(T_{\tilde{K}} < T\big) < \varepsilon$$

and we obtain $\lim\limits_{K\to\infty} \mathbb{P}\big(T_{\zeta,K} \leq t\big) = 0$, which yields the statement. □

Recall that we have a fixed constant $\eta > \frac{1}{2\xi - 1}$, determined by Lemma 4.27. We use this to fix the sequence $(m^{(n)})_{n\in\mathbb{N}}$ by defining

$$m^{(n)} := a_{n-1}^{-\frac{1}{\eta}} > 1,$$

where a_n is the Yamada–Watanabe sequence, defined in (4.29). With this, we get the following crucial lemma, that regularizes $\tilde{X}$ based on regularity of the approximation $|\langle \tilde{X}, \Phi^n \rangle|$.

Lemma 4.29. *For all $x \in B(0, \frac{1}{m})$ and $s \in [0, T_{\zeta,K}]$, if $|\langle \tilde{X}_s, \Phi_x^n \rangle| \leq a_{n-1}$, then*

$$\sup_{y \in B(x, \frac{1}{m})} |\tilde{X}(s, y)| \leq \tilde{C}_K a_{n-1},$$

for some $\tilde{C}_K > 0$ only dependent on K.

Proof. By the assumption $|\langle \tilde{X}_s, \Phi_x^n \rangle| \leq a_{n-1}$, we can apply Proposition 4.21 (v) to get that there exists $\hat{x} \in B(x, \frac{1}{m})$ with $|\tilde{X}(s, \hat{x})| \leq C_K a_{n-1}$.
For fixed $n \geq 1$, we define $\varepsilon_n > 0$ such that

$$\varepsilon_n^{\alpha} = \frac{1}{m^{(n)}} C_K^{\frac{1}{\eta}}$$

holds and, thus, by the choice $\eta = \frac{\zeta}{\alpha}$,

$$C_K a_{n-1} = C_K \left(\frac{1}{m}\right)^{\eta} = \left(\frac{C_K^{\frac{1}{\eta}}}{m}\right)^{\eta} = \varepsilon_n^{\zeta}.$$

We use this and the definition of $T_{\zeta,K}$ in (4.83) to get the desired result with $\tilde{C}_K = C_K c_0(K)$. □

Finally, we can handle the term $I_4^{m,n}$ from (4.33).

Lemma 4.30. *With $I_4^{m,n}$ from (4.33) and $T_{\zeta,K}$ defined in (4.83), one has*

$$\lim_{n\to\infty} \mathbb{E}[|I_4^{m,n}(t\wedge T_{\zeta,K})|] = 0.$$

Proof. We use the Hölder continuity of σ as well as the bounded support of ψ_n, the inequality $\psi_n(x) \le \frac{2}{nx}\mathbb{1}_{\{a_n\le x\le a_{n-1}\}}$, the boundedness of Ψ, Lemma 4.29 and Proposition 4.21 (ii) to get

$$\begin{aligned}
|I_4^{m,n}(t\wedge T_{\zeta,K})| &\lesssim \left|\int_0^{t\wedge T_{\zeta,K}}\int_{\mathbb{R}} \psi_n(|\langle \tilde{X}_s,\Phi_x^n\rangle|)\Phi_x^n(0)^2\Psi_s(x)\,\mathrm{d}x|\tilde{X}(s,0)|^{2\xi}\,\mathrm{d}s\right| \\
&\lesssim \int_0^{t\wedge T_{\zeta,K}}\int_{\mathbb{R}} \mathbb{1}_{\{a_n\le|\langle\tilde{X}_s,\Phi_x^n\rangle|\le a_{n-1}\}}\frac{2}{na_n}\Phi_x^n(0)^2\Psi_s(x)\,\mathrm{d}x|\tilde{X}(s,0)|^{2\xi}\,\mathrm{d}s \\
&\le \frac{\|\Psi\|_\infty}{na_n}\int_0^{t\wedge T_{\zeta,K}}\int_{\mathbb{R}}\Phi_x^n(0)^2\,\mathrm{d}x(\tilde{C}_K a_{n-1})^{2\xi}\,\mathrm{d}s \\
&\lesssim \frac{a_{n-1}^{2\xi}}{na_n}\int_0^{t\wedge T_{\zeta,K}}\int_{\mathbb{R}}\Phi_x^n(0)^2\,\mathrm{d}x\,\mathrm{d}s \\
&\lesssim \frac{a_{n-1}^{2\xi}}{na_n}m^{(n)} \lesssim \frac{a_{n-1}^{2\xi}}{na_n}a_{n-1}^{-\frac{1}{\eta}} = \frac{1}{n}\frac{a_{n-1}^{2\xi-\frac{1}{\eta}}}{a_n}.
\end{aligned} \tag{4.86}$$

We know that $\frac{a_{n-1}}{a_n} = e^n$, $a_0 = 1$ and, thus, get inductively that $a_n = e^{-\frac{n(n+1)}{2}}$. Therefore, (4.86) tends to zero as $n\to\infty$ if

$$n(n+1) - (2\xi-\eta^{-1})(n-1)n < 0$$

for n large, which holds if and only if $1-(2\xi-\eta^{-1}) < 0$, i.e., $\xi > \frac{1}{2}+\frac{1}{2\eta}$, which holds by Lemma 4.27. □

We summarize the essential findings for the proof of Theorem 4.3 in the next proposition.

Proposition 4.31. *With Ψ that fulfills Assumption 4.19 and $T_{\zeta,K}$ defined in (4.83) for $K>0$, one has, for $t\in[0,T]$, that*

$$\begin{aligned}
\int_{\mathbb{R}} \mathbb{E}[|\tilde{X}(t\wedge T_{\zeta,K},x)|]\Psi_{t\wedge T_{\zeta,K}}(x)\,\mathrm{d}x &\lesssim \int_0^{t\wedge T_{\zeta,K}}\int_{\mathbb{R}}\mathbb{E}[|\tilde{X}(s,x)|]|\Delta_\theta\Psi_s(x)+\dot{\Psi}_s(x)|\,\mathrm{d}x\,\mathrm{d}s \\
&\quad + \int_0^{t\wedge T_{\zeta,K}}\Psi_s(0)\mathbb{E}[|\tilde{X}(s,0)|]\,\mathrm{d}s.
\end{aligned} \tag{4.87}$$

Proof. By Proposition 4.20, Lemma 4.23, Lemma 4.30 and sending $n\to\infty$ after applying

Fatou's lemma to exchange limiting and the integral, we get

$$
\begin{aligned}
&\int_{\mathbb{R}} \mathbb{E}[|\tilde{X}(t \wedge T_{\zeta,K}, x)|]\Psi_{t\wedge T_{K,\zeta}}(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}} \liminf_{n\to\infty} \mathbb{E}[\phi_n(\langle \tilde{X}_{t\wedge T_{\zeta,K}}, \Phi_x^n\rangle)]\Psi_{t\wedge T_{K,\zeta}}(x)\,\mathrm{d}x \\
&\le \liminf_{n\to\infty} \int_{\mathbb{R}} \mathbb{E}[\phi_n(\langle \tilde{X}_{t\wedge T_{\zeta,K}}, \Phi_x^n\rangle)]\Psi_{t\wedge T_{K,\zeta}}(x)\,\mathrm{d}x \\
&\lesssim \mathbb{E}\Big[\int_0^{t\wedge T_{\zeta,K}} \int_{\mathbb{R}} |\tilde{X}(s,x)|\big(\Delta_\theta \Psi_s(x) + \dot{\Psi}_s(x)\big)\,\mathrm{d}x\,\mathrm{d}s\Big] \\
&\quad + \mathbb{E}\Big[\int_0^{t\wedge T_{\zeta,K}} \Psi_s(0)|\tilde{X}(s,0)|\,\mathrm{d}s\Big].
\end{aligned}
\tag{4.88}
$$

Applying Fubini's theorem then yields (4.87). ☐

4.6 Step 5: Removing the auxiliary localizations

We want to construct appropriate test functions $\Psi \in C_0^\infty([0,t],\mathbb{R})$ for some fixed $t \in [0,T]$. They will be of the form

$$
\Psi_{N,M}(s,x) := (S_{t-s}\phi_M(x))g_N(x) \tag{4.89}
$$

for $N, M \in \mathbb{N}$, where $(S_u)_{u\in[0,T]}$ denotes the semigroup generated by Δ_θ and we specify the sequences of functions $\phi_M, g_N \in C_0^\infty(\mathbb{R})$ in the following.
With the sequence $(\phi_M)_{M\in\mathbb{N}}$ we want to approximate the Dirac distribution around 0. To that end, we define

$$
\phi_M(x) := Me^{-M^2x^2}\mathbb{1}_{\{|x|\le \frac{1}{M}\}} + s_M(x), \quad M \ge 2,
$$

where the function $s_M(x)$ extends smoothly to zero outside the ball $B(1, \frac{1}{M-1})$ such that $\lim_{M\to\infty} \phi_M(x) = \delta_0(x)$ pointwise.
Moreover, let $(g_N)_{N\in\mathbb{N}}$ be a sequence of functions in $C_0^\infty(\mathbb{R})$ such that $g_N\colon \mathbb{R} \to [0,1]$,

$$
B(0,N) \subset \{x \in \mathbb{R} : g_N(x) = 1\}, \quad B(0,N+1)^C \subset \{x \in \mathbb{R} : g_N(x) = 0\},
$$

and

$$
\sup_{N\in\mathbb{N}} \Big[\||x|^{-\theta} g_N'(x)\|_\infty + \|\Delta_\theta g_N(x)\|_\infty\Big] =: C_g < \infty. \tag{4.90}
$$

We simplify the term on the right-hand side of (4.88) in the next corollary.

Corollary 4.32. *With $\Psi_{N,M}$ constructed in* (4.89), *one has that*

$$
\begin{aligned}
&\Delta_\theta \Psi_{N,M}(s,x) + \dot{\Psi}_{N,M}(s,x) \\
&\quad = 4\alpha^2|x|^{-\theta}\Big(\frac{\partial}{\partial x}S_{t-s}\phi_M(x)\Big)\Big(\frac{\partial}{\partial x}g_N(x)\Big) + S_{t-s}\phi_M(x)\Delta_\theta g_N(x).
\end{aligned}
\tag{4.91}
$$

Proof. Recall, that, by the definition of the semigroup $(S_t)_{t\in[0,T]}$ in (4.15) and using the fundamental solution of (4.10), we get

$$\Delta_\theta S_t\phi(x) = \frac{\partial}{\partial t} S_t\phi(x), \quad t \in [0,T],$$

for all $\phi \in C_0^\infty(\mathbb{R})$. Therefore, the second term on the left-hand side of (4.91) equals

$$\begin{aligned}
\dot{\Psi}_{N,M}(s,x) &= g_N(x)\frac{\partial}{\partial s}\big(S_{t-s}\phi_M(x)\big) \\
&= -g_N(x)\Delta_\theta\big(S_{t-s}\phi_M(x)\big) \\
&= -2\alpha^2 g_N(x)\frac{\partial}{\partial x}\Big(|x|^{-\theta}\frac{\partial}{\partial x}\big(S_{t-s}\phi_M(x)\big)\Big) \\
&= -2\alpha^2 g_N(x)\Big(\frac{\partial}{\partial x}|x|^{-\theta}\Big)\Big(\frac{\partial}{\partial x}S_{t-s}\phi_M(x)\Big) - 2\alpha^2 g_N(x)|x|^{-\theta}\Big(\frac{\partial^2}{\partial x^2}S_{t-s}\phi_M(x)\Big).
\end{aligned} \tag{4.92}$$

For the first term on the left-hand side of (4.91), we calculate

$$\begin{aligned}
&\Delta_\theta\Psi_{N,M}(s,x) \\
&\quad= 2\alpha^2\frac{\partial}{\partial x}\Big(|x|^{-\theta}\frac{\partial}{\partial x}\Psi_{N,M}(s,x)\Big) \\
&\quad= 2\alpha^2|x|^{-\theta}\frac{\partial^2}{\partial x^2}\Big(S_{t-s}\phi_M(x)g_N(x)\Big) + 2\alpha^2\Big(\frac{\partial}{\partial x}|x|^{-\theta}\Big)\Big(\frac{\partial}{\partial x}S_{t-s}\phi_M(x)g_N(x)\Big) \\
&\quad= 4\alpha^2|x|^{-\theta}\Big(\frac{\partial}{\partial x}S_{t-s}\phi_M(x)\Big)\Big(\frac{\partial}{\partial x}g_N(x)\Big) + 2\alpha^2|x|^{-\theta}g_N(x)\Big(\frac{\partial^2}{\partial x^2}S_{t-s}\phi_M(x)\Big) \\
&\qquad+ 2\alpha^2|x|^{-\theta}\Big(S_{t-s}\phi_M(x)\Big)\Big(\frac{\partial^2}{\partial x^2}g_N(x)\Big) \\
&\qquad+ 2\alpha^2\Big(\frac{\partial}{\partial x}|x|^{-\theta}\Big)\Big(\frac{\partial}{\partial x}S_{t-s}\phi_M(x)\Big)g_N(x) \\
&\qquad+ 2\alpha^2\Big(\frac{\partial}{\partial x}|x|^{-\theta}\Big)\big(S_{t-s}\phi_M(x)\big)\Big(\frac{\partial}{\partial x}g_N(x)\Big).
\end{aligned} \tag{4.93}$$

Hence, adding up (4.92) and (4.93), we obtain

$$\begin{aligned}
&\Delta_\theta\Psi_{N,M}(s,x) + \dot{\Psi}_{N,M}(s,x) \\
&\quad= 4\alpha^2|x|^{-\theta}\Big(\frac{\partial}{\partial x}S_{t-s}\phi_M(x)\Big)\Big(\frac{\partial}{\partial x}g_N(x)\Big) + 2\alpha^2|x|^{-\theta}\Big(S_{t-s}\phi_M(x)\Big)\Big(\frac{\partial^2}{\partial x^2}g_N(x)\Big) \\
&\qquad+ 2\alpha^2\Big(\frac{\partial}{\partial x}|x|^{-\theta}\Big)\big(S_{t-s}\phi_M(x)\big)\Big(\frac{\partial}{\partial x}g_N(x)\Big) \\
&\quad= 4\alpha^2|x|^{-\theta}\Big(\frac{\partial}{\partial x}S_{t-s}\phi_M(x)\Big)\Big(\frac{\partial}{\partial x}g_N(x)\Big) + S_{t-s}\phi_M(x)\Delta_\theta g_N(x).
\end{aligned}$$

□

With these observations, we want to show that the semigroup $(S_t)_{t\in[0,T]}$ can be exponentially bounded in the following way.

Lemma 4.33. *For any $\phi \in C_0^\infty(\mathbb{R})$, $t \in [0,T]$ and for any $\lambda > 0$, there is a constant $C_{\lambda,\phi,t} > 0$ such that*

$$\left| S_t\phi(x) + \frac{\partial}{\partial x}(S_t\phi(x)) \right| \mathbb{1}_{\{N+1>|x|>N\}} \le C_{\lambda,\phi,t} e^{-\lambda|x|} \mathbb{1}_{\{N+1>|x|>N\}}$$

for any $N \ge 1$ and $x \in \mathbb{R}$.

Proof. For $t = 0$, the statement is trivial due to $S_0\phi(x) + \frac{\partial}{\partial x}(S_0\phi(x)) = \phi(x) + \phi'(x)$, which is bounded with compact support. Thus, we fix $t > 0$ and consider the first summand without the derivative. We use the inequality

$$I_\nu(b) < \left(\frac{b}{a}\right)^\nu e^{b-a} \left(\frac{a+\nu+\frac{1}{2}}{b+\nu+\frac{1}{2}}\right)^{\nu+\frac{1}{2}} I_\nu(a), \quad 0 < a < b, \nu > -1, \tag{4.94}$$

from [IS91, Theorem 2.1 (ii)], with $a = \frac{|y|^{1+\frac{\theta}{2}}}{t}$ and $b = \frac{|xy|^{1+\frac{\theta}{2}}}{t}$ such that $b > a$ due to $|x| > N \ge 1$. By the explicit form of $p_t^\theta(x,y)$ from Lemma 4.13, due to the compact support of ϕ, which we denote by S_ϕ, and using (4.94), we get

$$\begin{aligned}
S_t\phi(x) &= \int_{\mathbb{R}} \frac{(2+\theta)}{2t} |xy|^{\frac{(1+\theta)}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}} I_\nu\left(\frac{|xy|^{1+\frac{\theta}{2}}}{t}\right) \phi(y)\,\mathrm{d}y \\
&\le C_\phi \int_{S_\phi} \frac{(2+\theta)}{2t} |xy|^{\frac{(1+\theta)}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}} |x|^{\nu(1+\frac{\theta}{2})} e^{\frac{|xy|^{1+\frac{\theta}{2}}}{t} - \frac{|y|^{1+\frac{\theta}{2}}}{t}} I_\nu\left(\frac{|y|^{1+\frac{\theta}{2}}}{t}\right) \mathrm{d}y \\
&\le C_\phi \left(\int_{\mathbb{R}} \frac{(2+\theta)}{2t} |y|^{\frac{(1+\theta)}{2}} e^{-\frac{1^{2+\theta}+|y|^{2+\theta}}{2t}} I_\nu\left(\frac{|y|^{1+\frac{\theta}{2}}}{t}\right) \mathrm{d}y \right) |x|^{(\nu+1)(1+\frac{\theta}{2})} e^{-\frac{|x-1|^{2+\theta}}{2t}} \\
&\quad \times e^{c_\phi(|x|^{1+\frac{\theta}{2}}-1)} \\
&= C_\phi \left(\int_{\mathbb{R}} p_t^\theta(1,y)\,\mathrm{d}y \right) |x|^{(\nu+1)(1+\frac{\theta}{2})} e^{-\frac{|x-1|^{2+\theta}}{2t} + c_\phi(|x|^{1+\frac{\theta}{2}}-1) + \lambda|x|} e^{-\lambda|x|} \\
&\le C_{\lambda,\phi,t} e^{-\lambda|x|},
\end{aligned} \tag{4.95}$$

since the function $x \mapsto |x|^{(\nu+1)(1+\frac{\theta}{2})} e^{-\frac{|x-1|^{2+\theta}}{2t} + c_\phi(|x|^{1+\frac{\theta}{2}}-1) + \lambda|x|}$ attains a maximum on $\mathbb{R}$ for all $c_\phi > 0$.

For the second summand, we substitute $z = \frac{|xy|^{1+\frac{\theta}{2}}}{t}$ such that $\frac{1}{\partial x} = \frac{1+\frac{\theta}{2}}{t} y |xy|^{\frac{\theta}{2}} \frac{1}{\partial z}$, apply the product rule and $\frac{\partial}{\partial z} I_\nu(z) = \frac{\nu}{z} I_\nu(z) + I_{\nu+1}(z)$ (see [MOS66, page 67]) to get, for $|x| > 1$,

$$
\begin{aligned}
\frac{\partial}{\partial x}(S_t\phi(x)) &= \frac{\partial}{\partial x}\int_{\mathbb{R}} \frac{(2+\theta)}{2t}|xy|^{\frac{(1+\theta)}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}} I_\nu\left(\frac{|xy|^{1+\frac{\theta}{2}}}{t}\right)\phi(y)\,\mathrm{d}y \\
&= \frac{(2+\theta)}{2t}\int_{\mathbb{R}} \frac{\partial}{\partial z}\left(|xy|^{\frac{(1+\theta)}{2}} \frac{1+\frac{\theta}{2}}{t} y|xy|^{\frac{\theta}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}} I_\nu(z)\right)\phi(y)\,\mathrm{d}y \\
&= \frac{(2+\theta)}{2t}\int_{\mathbb{R}} \left(\frac{\partial}{\partial z}\left(|xy|^{\frac{(1+\theta)}{2}} \frac{1+\frac{\theta}{2}}{t} y|xy|^{\frac{\theta}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}}\right) I_\nu(z)\right. \\
&\qquad\qquad \left. + |xy|^{\frac{(1+\theta)}{2}} \frac{1+\frac{\theta}{2}}{t} y|xy|^{\frac{\theta}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}} \frac{\partial}{\partial z}(I_\nu(z))\right)\phi(y)\,\mathrm{d}y \\
&= \frac{(2+\theta)}{2t}\int_{\mathbb{R}} \left(\frac{1+\theta}{2} y|xy|^{\frac{(\theta-1)}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}}\right. \\
&\qquad\qquad \left. - \frac{2+\theta}{2t}|x|^{1+\theta}|xy|^{\frac{1+\theta}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}}\right) I_\nu\left(\frac{|xy|^{1+\frac{\theta}{2}}}{t}\right)\phi(y)\,\mathrm{d}y \\
&\quad + \frac{(2+\theta)}{2t}\int_{\mathbb{R}} \left(|xy|^{\frac{(1+\theta)}{2}} \frac{1+\frac{\theta}{2}}{t} y|xy|^{\frac{\theta}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}}\right. \\
&\qquad\qquad \left.\left(\nu \frac{t}{|xy|^{1+\frac{\theta}{2}}} I_\nu\left(\frac{|xy|^{1+\frac{\theta}{2}}}{t}\right) + I_{\nu+1}\left(\frac{|xy|^{1+\frac{\theta}{2}}}{t}\right)\right)\right)\phi(y)\,\mathrm{d}y \\
&\leq C_{t,\phi}\int_{S_\phi} \left(|xy|^{\frac{(1+\theta)}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}}\right. \\
&\qquad\qquad \left. + |x|^{1+\theta}|xy|^{\frac{(1+\theta)}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}}\right) I_\nu\left(\frac{|xy|^{1+\frac{\theta}{2}}}{t}\right)\mathrm{d}y \\
&\quad + \int_{S_\phi} \left(|xy|^{\frac{(1+\theta)}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}} \left(I_\nu\left(\frac{|xy|^{1+\frac{\theta}{2}}}{t}\right) + I_{\nu+1}\left(\frac{|xy|^{1+\frac{\theta}{2}}}{t}\right)\right)\right)\mathrm{d}y \\
&\leq C_{t,\phi}\int_{S_\phi} |x|^{1+\theta}|xy|^{\frac{(1+\theta)}{2}} e^{-\frac{|x|^{2+\theta}+|y|^{2+\theta}}{2t}} \left(I_\nu\left(\frac{|xy|^{1+\frac{\theta}{2}}}{t}\right) + I_{\nu+1}\left(\frac{|xy|^{1+\frac{\theta}{2}}}{t}\right)\right)\mathrm{d}y,
\end{aligned}
\tag{4.96}
$$

where $S_\phi := \{y \in \mathbb{R} \colon \phi(y) \neq 0\}$. The integrands in (4.96) vanish for $y = 0$ by the definition of I_ν in (4.21) with $\nu = \frac{1}{2+\theta} - 1 < \frac{1+\theta}{2}$. If we thus show that, for any $\nu > -1$, there is a constant $C_\nu > 0$ such that

$$
I_\nu(z) + I_{\nu+1}(z) \leq C_\nu\big(z^{\nu+1} + z^{\nu+2}\big)e^z \tag{4.97}
$$

holds for all $z > 0$, then the statement will follow, since, similar as in (4.95), all the x-polynomials in (4.96) and the Bessel function terms are dominated by the term $e^{-\frac{|x|^{2+\theta}}{2t}}$ and the y terms can be bounded using the compact support of ϕ.
To get (4.97), we use the equality (see [LS72, (5.7.9), page 110])

$$
I_\nu(z) = 2(\nu+1)I_{\nu+1}(z) + I_{\nu+2}(z), \tag{4.98}
$$

and, since $\nu+1, \nu+2 > -\frac{1}{2}$, we can then apply the following inequality from [Luk72, (6.25), page 63], for $x > 0$:

$$I_\nu(x) < \frac{e^x + e^{-x}}{2\Gamma(\nu+1)}\left(\frac{x}{2}\right)^\nu < \frac{e^x}{\Gamma(\nu+1)}\left(\frac{x}{2}\right)^\nu. \tag{4.99}$$

(4.98) and (4.99) yield, as $\Gamma(x) > 0$ for $x > 0$, that

$$\begin{aligned} I_\nu(z) + I_{\nu+1}(z) &= 2\left(\nu + \frac{3}{2}\right) I_{\nu+1}(z) + I_{\nu+2}(z) \\ &< 2\left(\nu + \frac{3}{2}\right)\frac{e^z}{\Gamma(\nu+2)}\left(\frac{z}{2}\right)^{\nu+1} + \frac{e^z}{\Gamma(\nu+3)}\left(\frac{z}{2}\right)^{\nu+2} \\ &\le C_\nu\big(z^{\nu+1} + z^{\nu+2}\big)e^z, \end{aligned}$$

which proves (4.97). $\square$

Proposition 4.34. *It holds that*

$$\mathbb{E}[|\tilde{X}(t,0)|] \lesssim \int_0^t (t-s)^{-\alpha}\mathbb{E}[|\tilde{X}(s,0)|]\,\mathrm{d}s, \qquad t \in [0,T].$$

Proof. First, to apply Proposition 4.31, we need to show that $\Psi_{N,M}$ defined in (4.89) fulfills Assumption 4.19. $\Psi_{N,M} \in C^2([0,T]\times\mathbb{R})$ and the conditions $\Psi_{N,M}(s,0) > 0$ and $\Gamma(t) \in B(0, J(t))$ for some $J(t) > 0$ follow by construction. Moreover, Lemma 4.33 directly yields that the last property holds:

$$\sup_{s\le t}\left|\int_{\mathbb{R}} |x|^{-\theta}\left(\frac{\partial}{\partial x}\Psi_{N,M}(s,x)\right)^2 \mathrm{d}x\right| \le C\int_{\mathbb{R}} |x|^{-\theta} e^{-2\lambda|x|}\,\mathrm{d}x,$$

which is clearly finite as $\theta < 1$. Hence, Assumption 4.19 holds.
Thus, Proposition 4.31 holds and plugging (4.89) into (4.87), sending $K \to \infty$ such that $T_{\zeta,K} \to T$ by Corollary 4.28 and using Corollary 4.32, (4.90) and Lemma 4.33, we get

$$\begin{aligned} &\int_{\mathbb{R}} \mathbb{E}[|\tilde{X}(t,x)|]\phi_M(x)g_N(x)\,\mathrm{d}x \\ &\quad\lesssim \int_0^t\int_{\mathbb{R}} \mathbb{E}[|\tilde{X}(s,x)|]\left|4\alpha^2|x|^{-\theta}\Big(\frac{\partial}{\partial x}S_{t-s}\phi_M(x)\Big)\Big(\frac{\partial}{\partial x}g_N(x)\Big) + S_{t-s}\phi_M(x)\Delta_\theta g_N(x)\right|\mathrm{d}x\,\mathrm{d}s \\ &\qquad + \int_0^t \Psi_{N,M}(s,0)\mathbb{E}[|\tilde{X}(s,0)|]\,\mathrm{d}s \\ &\quad\lesssim \int_0^t\int_{\mathbb{R}} \mathbb{E}[|\tilde{X}(s,x)|]\Big(\frac{\partial}{\partial x}S_{t-s}\phi_M(x)\Big) + S_{t-s}\phi_M(x)|\mathbb{1}_{\{N+1>|x|>N\}}\,\mathrm{d}x\,\mathrm{d}s \\ &\qquad + \int_0^t \Psi_{N,M}(s,0)\mathbb{E}[|\tilde{X}(s,0)|]\,\mathrm{d}s \\ &\quad\lesssim \int_0^t\int_{\mathbb{R}} \mathbb{E}[|\tilde{X}(s,x)|]e^{-\lambda|x|}\mathbb{1}_{\{N+1>|x|>N\}}\,\mathrm{d}x\,\mathrm{d}s + \int_0^t \Psi_{N,M}(s,0)\mathbb{E}[|\tilde{X}(s,0)|]\,\mathrm{d}s. \end{aligned} \tag{4.100}$$

We want to send $N, M \to \infty$. By Proposition 4.11 (i) we get that

$$\int_0^t \int_{\mathbb{R}} \mathbb{E}[|\tilde{X}(s,x)|] e^{-\lambda|x|} \mathbb{1}_{\{N+1>|x|>N\}} \,\mathrm{d}x \,\mathrm{d}s \lesssim t \int_N^{N+1} e^{-\lambda x} \,\mathrm{d}x \to 0 \quad \text{as } N \to \infty.$$

Moreover, we get

$$\begin{aligned}
\int_0^t \Psi_{N,M}(s,0) \mathbb{E}[|\tilde{X}(s,0)|] \,\mathrm{d}s &= \int_0^t (S_{t-s}\phi_M(0)) g_N(x) \mathbb{E}[|\tilde{X}(s,0)|] \,\mathrm{d}s \\
&= \int_0^t \left(\int_{\mathbb{R}} p^\theta_{t-s}(y,0)\phi_M(y) \,\mathrm{d}y \right) \mathbb{E}[|\tilde{X}(s,0)|] \,\mathrm{d}s \\
&\overset{M\to\infty}{\to} \int_0^t p^\theta_{t-s}(0) \mathbb{E}[|\tilde{X}(s,0)|] \,\mathrm{d}s \quad \text{as } M \to \infty,
\end{aligned}$$

which gives

$$\int_0^t \Psi_{N,M}(s,0) \mathbb{E}[|\tilde{X}(s,0)|] \,\mathrm{d}s = c_\theta \int_0^t (t-s)^{-\alpha} \mathbb{E}[|\tilde{X}(s,0)|] \,\mathrm{d}s.$$

Hence, sending $N, M \to \infty$ in (4.100) yields

$$\mathbb{E}[|\tilde{X}(t,0)| \lesssim \int_0^t (t-s)^{-\alpha} \mathbb{E}[|\tilde{X}(s,0)|] \,\mathrm{d}s.$$

$\square$

Chapter 5

Mean-field SVEs

The content of this chapter is published in [PS23b].

Introduction

Mean-field stochastic differential equations (mean-field SDEs), also known as McKean–Vlasov stochastic differential equations, provide mathematical descriptions of random systems of interacting particles, whose time evolutions depend, in some manner, on the probability distribution of the entire systems. A crucial reason for the frequent use of mean-field SDEs in applied mathematics is the fact that they allow for modelling the phenomena of "propagation of chaos" of large interacting particle systems. Recall, on a microscopic scale the trajectory of each individual particle can often be appropriately modelled by a stochastic process. However, when the number of particles becomes very large, the microscopic scale usually contains too much information, making the interaction of individual particles intractable. Fortunately, sending the number of particles to infinity, propagation of chaos states that the behavior of an individual particle depends only on the probability distribution of the entire system, i.e., on the macroscopic scale the interaction of individual particles becomes negligible.

Mean-field SDEs as well as propagation of chaos originated in statistical physics and were first studied by Kac [Kac56], McKean [McK66] and Vlasov [Vla68]. By now, these concepts have found a wide range of applications in a variety of fields such as physics, finance and data science. We refer, e.g., to [Szn91, JW17, CD18a, CD18b, CD22a, CD22b] for comprehensive introductions to mean-field SDEs and their numerous applications. Except a very small number of publications, like the rough path based approaches to mean-field SDEs [BCD20, CDFM20, BCD21], the vast majority of literature on mean-field SDEs and propagation of chaos is restricted to Markovian systems of interacting particles, i.e. the behavior of each particle has to be independent of all past states of the systems. On the contrary, it is well observed that many real-world dynamical systems do have memory effects and, thus, do indeed depend on past states of the underlying systems. Well-known examples of such systems are the growth of populations, the spread of epidemics and

turbulence flows.
Classical mathematical models for random dynamical systems with memory effects are given by stochastic Volterra equations (SVEs), as introduced in the seminal works of Berger and Mizel [BM80a, BM80b], see also e.g. [Pro85, PP90]. While SVEs allow for generating non-Markovian stochastic processes, the solutions of SVEs, in contrast to mean-field SDEs, do not depend directly on the probability distributions of the generated random systems.
In this chapter we aim to unify the theory of mean-field stochastic differential equations and stochastic Volterra equations, which enables to combine the desirable modelling advantages of both classes of equations. More precisely, we introduce mean-field stochastic Volterra equations (mean-field SVEs)

$$X_t = X_0 + \int_0^t K_\mu(s,t)\mu(s,X_s,\mathcal{L}(X_s))\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s,X_s,\mathcal{L}(X_s))\,\mathrm{d}B_s, \quad t \in [0,T], \tag{5.1}$$

where X_0 is a random variable, B is a Brownian motion, and the coefficients μ, σ as well as the kernels K_μ, K_σ are measurable functions. Here, $\mathcal{L}(X_s)$ denotes the law of the random variable X_s. In words, mean-field SVEs are a class of stochastic integral equations that describe the dynamics of random systems with both nonlinear interactions and memory effects. They constitute a generalization of mean-field SDEs and of classical SVEs. Notice that a solution to the mean-field SVE (5.1) is, in general, neither a Markov process nor a semimartingale.
Our first contribution is to establish the (strong) well-posedness of the mean-field SVE (5.1), meaning that there exists a unique strong solution to (5.1), under two sets of assumptions. On the one hand, we show the existence of a unique solution to the mean-field SVE (5.1) in a multi-dimensional setting with standard assumptions on the kernels and coefficients, i.e. we assume some integrability on the kernels as well as Lipschitz continuity and a linear growth condition on the coefficients, cf. e.g. [Wan08, Car16]. The proof is based on a classical fixed point argument in combination with techniques from the theories of mean-field SDEs and SVEs. On the other hand, we show the existence of a unique solution to the mean-field SVE (5.1) in a one-dimensional setting, assuming sufficiently smooth kernels and Hölder continuous diffusion coefficients which are independent of the law of the solution. To that end, we rely on a Yamada–Watanabe approach [YW71], as recently generalized in [AJLP19] or Chapter 2 to SVEs with sufficiently smooth kernels. As comparison, for well-posedness results in case of mean-field SDEs we refer to [BMM20, KP21] such as [HW23] in the Hölder continuous diffusion coefficient case, and in case of SVEs to [Wan08, AJLP19], or Chapter 2. Furthermore, let us remark that a specific type of mean-field SVEs was studied in [SWY13].
Our second contribution is to establish quantitative, pointwise propagation of chaos results of Volterra-type systems of interacting particles. In words, sending the number of Volterra-type interacting particles to infinity, we obtain a macroscopic description of the systems based on a mean-field stochastic Volterra equation. The developed approach is based on a synchronous coupling method, as it was initiated by McKean [McK67] and extended

by Sznitman [Szn91]. In the case of mean-field SDEs, synchronous coupling methods are widely used for systems that are described by systems of McKean–Vlasov diffusions, and often lead to pathwise propagation of chaos, see e.g. [CD22a, Theorem 3.20] or [Car16, Theorem 1.10]. In the present case of mean-field SVEs, implementing a synchronous coupling method becomes more challenging as the underlying McKean–Vlasov processes are of Volterra type and, thus, in general, lack the semimartingale and Markov property. As for the presented well-posedness theory of mean-field SVEs, we distinguish between the aforementioned multi- and one-dimensional setting. The pointwise natures of the presented propagation of chaos results for mean-field SVEs is caused by the non-availability of a Burkholder–Davis–Gundy inequality in the multi-dimensional setting and by the Hölder continuity of the diffusion coefficients in the one-dimensional setting. The latter setting requires to combine the synchronous coupling method with a Yamada–Watanabe approach like in the SDE case in [HW23].

Organization of the chapter: In Section 5.1 we present the main results regarding the well-posedness and propagation of chaos for mean-field stochastic Volterra equations. Section 5.2 provides some necessary well-posedness results for ordinary stochastic Volterra equations. The proofs of the main results are contained in Section 5.3, 5.4 and 5.5.

5.1 Main results: well-posedness and propagation of chaos

Let $T \in (0, \infty)$, $d, m \in \mathbb{N}$, and let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space, which satisfies the usual conditions. Suppose $B = (B_t)_{t\in[0,T]}$ is an m-dimensional Brownian motion with respect to $(\mathcal{F}_t)_{t\in[0,T]}$. The law of a random variable X is denoted by $\mathcal{L}(X)$ and, for $p \geq 1$, the space of probability measures on $\mathbb{R}^d$ with finite p-th moments by $\mathcal{P}_p(\mathbb{R}^d)$. For $\rho, \tilde{\rho} \in \mathcal{P}_p(\mathbb{R}^d)$, we write $W_p(\rho, \tilde{\rho})$ for the p-Wasserstein distance between ρ and $\tilde{\rho}$, see [CD18a, Chapter 5] for its definition. The space $\mathbb{R}^d$ is always equipped with the Euclidean norm $|\cdot|$. Moreover, we set $\Delta_T := \{(s,t) \in [0,T] \times [0,T] \colon\ 0 \leq s \leq t \leq T\}$ and use the notation $A_\eta \lesssim B_\eta$ for a generic parameter η, meaning that $A_\eta \leq C B_\eta$ for some constant $C > 0$ independent of η.

We consider the d-dimensional mean-field stochastic Volterra equation

$$X_t = X_0 + \int_0^t K_\mu(s,t)\mu(s, X_s, \mathcal{L}(X_s))\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s, X_s, \mathcal{L}(X_s))\,\mathrm{d}B_s, \quad t \in [0,T], \tag{5.2}$$

where X_0 is a d-dimensional, $\mathcal{F}_0$-measurable random variable, which is independent of B, the coefficients $\mu\colon [0,T]\times\mathbb{R}^d\times\mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$, $\sigma\colon [0,T]\times\mathbb{R}^d\times\mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^{d\times m}$ and the kernels $K_\mu, K_\sigma\colon \Delta_T \to \mathbb{R}$ are measurable functions. The integral $\int_0^t K_\sigma(s,t)\sigma(s, X_s, \mathcal{L}(X_s))\,\mathrm{d}B_s$ is defined as a stochastic Itô integral.

Let us briefly recall the concepts of well-posedness, strong solutions and pathwise uniqueness. We use, for topological spaces $\mathcal{X}, \mathcal{Y}$ and $p \geq 1$, the notation $L^p(\mathcal{X}; \mathcal{Y})$ for the space of all $\mathcal{Y}$-valued, measurable, p-integrable functions on $\mathcal{X}$ and $C(\mathcal{X}; \mathcal{Y})$ for the space of all

$\mathcal{Y}$-valued, continuous functions on $\mathcal{X}$. An $(\mathcal{F}_t)_{t\in[0,T]}$-progressively measurable stochastic process $(X_t)_{t\in[0,T]}$ in $L^p(\Omega\times[0,T];\mathbb{R}^d)$, on the given probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in[0,T]},\mathbb{P})$, is called *(strong) L^p-solution* of the mean-field SVE (5.2) if

$$\int_0^t (|K_\mu(s,t)\mu(s,X_s,\mathcal{L}(X_s))| + |K_\sigma(s,t)\sigma(s,X_s,\mathcal{L}(X_s))|^2)\,\mathrm{d}s < \infty \quad \text{for all } t\in[0,T],$$

and the integral equation (5.2) holds $\mathbb{P}$-almost surely. We say *pathwise uniqueness in L^p* holds for the mean-field SVE (5.2) if $\mathbb{P}(X_t = \tilde{X}_t, \forall t \in [0,T]) = 1$ for any two L^p-solutions $(X_t)_{t\in[0,T]}$ and $(\tilde{X}_t)_{t\in[0,T]}$ of (5.2) defined on the same probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in[0,T]},\mathbb{P})$. We say that the mean-field SVE (5.2) is *well-posed in L^p* (or that there exists a *unique L^p-solution*) for $p\geq 1$ if there exists a strong L^p-solution to (5.2) and pathwise uniqueness in L^p holds.

In the following we distinguish between a multi-dimensional and a one-dimensional setting since these settings allow to establish well-posedness of the mean-field SVE (5.2) with different regularity assumptions on the kernels and coefficients. The main existence and uniqueness results regarding mean-field SVEs as well as propagation of chaos are stated in Subsection 5.1.1 and 5.1.2. In the multi-dimensional setting (Subsection 5.1.1) we make standard Lipschitz assumptions on the coefficients μ,σ, whereas in the one-dimensional setting (Subsection 5.1.2) we assume that μ is Lipschitz continuous but allow σ to be only Hölder continuous. We prove the corresponding results in Section 5.3, 5.4 and 5.5.

5.1.1 Mean-field SVEs with Lipschitz continuous coefficients

In this subsection we consider the multi-dimensional stochastic Volterra equation (5.2) with dimensions $d,m\in\mathbb{N}$ and coefficients μ,σ that are Lipschitz continuous in the space and distributional component, uniformly in the time component, allowing for potentially singular kernels. We start by stating the assumptions on the kernels.

Assumption 5.1. *Assume there are constants $\gamma\in(0,\frac{1}{2}]$, $\varepsilon>0$ and $L>0$, such that $K_\mu, K_\sigma\colon \Delta_T\to\mathbb{R}$ are measurable functions fulfilling*

$$\int_0^t |K_\mu(s,t')-K_\mu(s,t)|^{1+\varepsilon}\,\mathrm{d}s + \int_t^{t'} |K_\mu(s,t')|^{1+\varepsilon}\,\mathrm{d}s \leq L|t'-t|^{\gamma(1+\varepsilon)},$$
$$\int_0^t |K_\sigma(s,t')-K_\sigma(s,t)|^{2+\varepsilon}\,\mathrm{d}s + \int_t^{t'} |K_\sigma(s,t')|^{2+\varepsilon}\,\mathrm{d}s \leq L|t'-t|^{\gamma(2+\varepsilon)},$$

for all $(t,t')\in\Delta_T$.

Note that Assumption 5.1 allows for singular kernels, like the fractional convolutional kernel $K(s,t)=(t-s)^{-\alpha}$ for $\alpha\in(0,1/2)$ and the examples provided in [AJCLP21, Example 1.3]. Moreover, let for $\varepsilon>0$ given by Assumption 5.1, the fixed parameter $\delta>2$ be defined by

$$\delta := \frac{4+2\varepsilon}{\varepsilon}, \tag{5.3}$$

such that

$$\frac{2}{2+\varepsilon} + \frac{2}{\delta} = 1. \tag{5.4}$$

In the following we use the δ-Wasserstein distance on the space $\mathcal{P}_\delta(\mathbb{R}^d)$ of probability measures on $\mathbb{R}^d$ with finite δ-th moments. Relying on the δ-Wasserstein distance, we specify the assumptions on the regularity of the coefficients μ and σ, which are a classical linear growth condition and a Lipschitz assumption.

Assumption 5.2. *Let $\mu\colon [0,T]\times\mathbb{R}^d\times\mathcal{P}_\delta(\mathbb{R}^d)\to\mathbb{R}^d$ and $\sigma\colon [0,T]\times\mathbb{R}^d\times\mathcal{P}_\delta(\mathbb{R}^d)\to\mathbb{R}^{d\times m}$ be measurable functions such that:*

(i) for any bounded set $\mathcal{K}\subset\mathcal{P}_\delta(\mathbb{R}^d)$, there is a constant $C_{\mathcal{K}}>0$, such that the linear growth condition

$$|\mu(t,x,\rho)| + |\sigma(t,x,\rho)| \le C_{\mathcal{K}}(1+|x|)$$

holds for all $\rho\in\mathcal{K}$, $t\in[0,T]$ and $x\in\mathbb{R}^d$;

(ii) μ and σ are Lipschitz continuous in x and in ρ w.r.t. the δ-Wasserstein distance, uniformly in t, i.e. there is a constant $C_{\mu,\sigma}>0$ such that

$$|\mu(t,x,\rho)-\mu(t,\tilde{x},\tilde{\rho})| + |\sigma(t,x,\rho)-\sigma(t,\tilde{x},\tilde{\rho})| \le C_{\mu,\sigma}\big(|x-\tilde{x}| + W_\delta(\rho,\tilde{\rho})\big),$$

holds for all $t\in[0,T]$, $x,\tilde{x}\in\mathbb{R}^d$, and $\rho,\tilde{\rho}\in\mathcal{P}_\delta(\mathbb{R}^d)$.

Our first result is the well-posedness of the mean-field stochastic Volterra equation (5.2).

Theorem 5.3. *Suppose that the initial value X_0 is in $L^p(\Omega;\mathbb{R}^d)$, the kernels K_μ, K_σ fulfill Assumption 5.1, the coefficients μ,σ fulfill Assumption 5.2, and $p>\max\{\frac{1}{\gamma}, 1+\frac{2}{\varepsilon}\}$, where $\gamma\in(0,\frac{1}{2}]$ and $\varepsilon>0$ are given by Assumption 5.1. Then, the mean-field stochastic Volterra equation* (5.2) *is well-posed in L^p. Moreover, for any $q\ge 1$, if $X_0\in L^q(\Omega;\mathbb{R}^d)$, the unique L^p-solution X of* (5.2) *satisfies*

$$\sup_{t\in[0,T]} \mathbb{E}[|X_t|^q] < \infty. \tag{5.5}$$

Our second result is propagation of chaos for mean-field stochastic Volterra equations, i.e. we show that the unique L^p-solution to the mean-field stochastic Volterra equation (5.2) is the limit $N\to\infty$ of the solutions to the following symmetric system of N mean-field stochastic Volterra equations

$$X_t^{N,i} = X_0^i + \int_0^t K_\mu(s,t)\mu(s,X_s^{N,i},\bar{\rho}_s^N)\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s,X_s^{N,i},\bar{\rho}_s^N)\,\mathrm{d}B_s^i, \quad t\in[0,T], \tag{5.6}$$

for $i\in\{1,\dots,N\}$, where $\bar{\rho}_t^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_t^{N,i}}$ is the empirical distribution of $(X_t^{N,i})_{i=1,\dots,N}$, $(X_0^i)_{i\in\mathbb{N}}\subset L^q(\Omega;\mathbb{R}^d)$ is a sequence of $\mathcal{F}_0$-measurable, independent and identically distributed random variables for some $q>4$, and $(B^i)_{i\in\mathbb{N}}$ is a sequence of independent

m-dimensional Brownian motions, which are all defined on the given probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$. Strong L^p-solutions, pathwise uniqueness in L^p and well-posedness in L^p for the system (5.6) of mean-field SVEs is defined analogously to (5.2) and δ_x denotes the Dirac measure at x for $x \in \mathbb{R}^d$. Moreover, for $i \in \mathbb{N}$, let $\underline{X}^i$ be the solution of the mean-field SVE (5.2) with the initial condition X_0^i and driving Brownian motion B^i. In the present multi-dimensional setting, we obtain the following convergence result.

Theorem 5.4 (Volterra propagation of chaos). *Suppose Assumption 5.1 and 5.2, and that the sequence of initial conditions $(X_0^i)_{i\in\mathbb{N}} \subset L^q(\Omega;\mathbb{R}^d)$ for some $q > \max\{p, 2\delta\}$ and $p > \max\{\frac{1}{\gamma}, 1 + \frac{2}{\varepsilon}\}$, where δ is defined in (5.3). Then, the system (5.6) of mean-field SVEs is well-posed in L^p for every $N \geq 1$, where the unique L^p-solution is denoted by $(X_t^{N,i})_{i=1,\dots,N}$. Moreover, it holds*

$$\lim_{N\to\infty} \left(\max_{1\leq i\leq N} \Big(\sup_{t\in[0,T]} \mathbb{E}[|X_t^{N,i} - \underline{X}_t^i|^\delta] \Big) + \sup_{t\in[0,T]} \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{i=1}^N \delta_{X_t^{N,i}}, \mathcal{L}(\underline{X}_t^1)\Big)^\delta \Big] \right) = 0. \quad (5.7)$$

The rate of convergence in (5.7) is explicitly stated in the next lemma.

Lemma 5.5. *Supposing the assumptions and notation of Theorem 5.4, it holds that*

$$\max_{1\leq i\leq N} \Big(\sup_{t\in[0,T]} \mathbb{E}[|X_t^{N,i} - \underline{X}_t^i|^\delta] \Big) + \sup_{t\in[0,T]} \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{i=1}^N \delta_{X_t^{N,i}}, \mathcal{L}(\underline{X}_t^1)\Big)^\delta \Big] \lesssim \varepsilon_N, \quad (5.8)$$

where $(\varepsilon_N)_{N\in\mathbb{N}}$ is given by

$$\varepsilon_N = \begin{cases} N^{-1/2}, & \text{if } d < 2\delta \\ N^{-1/2}\log_2(1+N), & \text{if } d = 2\delta \\ N^{-\delta/d}, & \text{if } d > 2\delta \end{cases}. \quad (5.9)$$

Remark 5.6. *The rates of convergence obtained in* (5.9) *are analogue to the classical rates for ordinary mean-field SDEs with Lipschitz coefficients (see [CD22a, Theorem 3.20]), using $W_\delta(\cdots)^\delta$ instead of $W_2(\cdots)^2$ and, consequently, replacing the exponent $2/d$ by δ/d in* (5.9). *Note that in the case of ordinary mean-field SDEs one obtains a pathwise propagation of chaos result (meaning that the* sup *in* (5.8) *is inside the expectation operators), which is a stronger type of convergence than the pointwise convergence presented in Theorem 5.4. This weaker type of convergence is caused by the missing availability of a Burkholder–Davis–Gundy inequality for stochastic Volterra processes. However, the rates of convergence provided in Lemma 5.5 seem to be optimal for synchronous coupling methods, since it is shown in [FG15, Theorem 1 and there after] that for terms of the form $\mathbb{E}[W_\delta(\bar{\rho}_N, \rho)^\delta]$ the rates in* (5.9) *are sharp. Consequently, optimality could be only lost in the inequalities* (5.48) *or* (5.49), *which, at least in general, appears not to be the case.*

5.1.2 Mean-field SVEs with Hölder continuous diffusion coefficients

In this subsection we consider mean-field SVEs in a one-dimensional setting, i.e. we assume $d = m = 1$. This allows to relax the Lipschitz assumption on the diffusion coefficient σ to Hölder continuity in the space variable, provided that σ is independent of the distribution of the solution and that the kernels are sufficiently regular. More precisely, we consider the one-dimensional mean-field stochastic Volterra equation

$$X_t = X_0 + \int_0^t K_\mu(s,t)\mu(s, X_s, \mathcal{L}(X_s))\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s, X_s)\,\mathrm{d}B_s, \quad t \in [0,T], \quad (5.10)$$

where $(B_t)_{t\in[0,T]}$ is a one-dimensional Brownian motion, X_0 is an $\mathcal{F}_0$-measurable random variable, the coefficients $\mu\colon [0,T] \times \mathbb{R} \times \mathcal{P}_p(\mathbb{R}) \to \mathbb{R}$, $\sigma\colon [0,T] \times \mathbb{R} \to \mathbb{R}$ and the kernels $K_\mu, K_\sigma\colon \Delta_T \to \mathbb{R}$ are measurable functions. We consider two different sets of assumptions on the kernels and on the initial condition.

Assumption 5.7. *Let $\gamma \in (0, \frac{1}{2}]$ and $\varepsilon > 0$. Let X_0 be an $\mathcal{F}_0$-measurable random variable and $K_\mu, K_\sigma\colon \Delta_T \to \mathbb{R}$ be continuous functions such that:*

(i) $K_\mu(s,\cdot)$ is absolutely continuous for every $s \in [0,T]$ and $\partial_2 K_\mu$ is bounded on Δ_T;

(ii) $K_\sigma(\cdot,t)$ is absolutely continuous for every $t \in [0,T]$, $K_\sigma(s,\cdot)$ is absolutely continuous for every $s \in [0,T]$ with $\partial_2 K_\sigma \in L^2(\Delta_T)$, and $\partial_2 K_\sigma(\cdot,t)$ is absolutely continuous for every $t \in [0,T]$. Furthermore, there is a constant $C_1 > 0$ such that $|K_\sigma(t,t)| \geq C_1$ for any $t \in [0,T]$, and there exists $C_2 > 0$ such that

$$\int_0^s |K_\sigma(u,t) - K_\sigma(u,s)|^{2+\varepsilon}\,\mathrm{d}u \leq C_2 |t-s|^{\gamma(2+\varepsilon)} \text{ and}$$

$$|\partial_1 K_\sigma(s,t)| + |\partial_2 K_\sigma(s,s)| + \int_s^t |\partial_{21} K_\sigma(s,u)|\,\mathrm{d}u \leq C_2$$

hold for any $(s,t) \in \Delta_T$;

(iii) $X_0 \in L^p(\Omega;\mathbb{R})$ for $p > \max\{\frac{1}{\gamma}, 1 + \frac{2}{\varepsilon}\}$.

Instead of Assumption 5.7, we can alternatively require K_μ, K_σ and X_0 to fulfill the following assumption, where the kernels are supposed to be convolutional.

Assumption 5.8. *Let X_0 be an $\mathcal{F}_0$-measurable random variable and $K_\mu, K_\sigma\colon \Delta_T \to \mathbb{R}$ be continuous functions such that:*

(i) $K_\mu(s,t) = K_\sigma(s,t) = \tilde{K}(t-s)$ for some $\tilde{K} \in C^1([0,T];\mathbb{R})$;

(ii) $X_0 \in L^p(\Omega;\mathbb{R})$ for $p > 2$.

Next, we formulate the assumptions on the coefficients.

Assumption 5.9. *Let $\mu\colon [0,T] \times \mathbb{R} \times \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ and $\sigma\colon [0,T] \times \mathbb{R} \to \mathbb{R}$ be measurable functions such that:*

(i) for any bounded set $\mathcal{K} \subset \mathcal{P}_1(\mathbb{R})$, there is a constant $C_{\mathcal{K}} > 0$, such that the linear growth condition

$$|\mu(t,x,\rho)| + |\sigma(t,x)| \leq C_{\mathcal{K}}\rho(1+|x|)$$

holds for all $\rho \in \mathcal{K}$, $t \in [0,T]$ and $x \in \mathbb{R}$;

(ii) μ is Lipschitz continuous in x and ρ w.r.t. the 1-Wasserstein distance, uniformly in t, i.e. there is a constant $C_\mu > 0$ such that

$$|\mu(t,x,\rho) - \mu(t,\tilde{x},\tilde{\rho})| \leq C_\mu\big(|x-\tilde{x}| + W_1(\rho,\tilde{\rho})\big),$$

holds for all $t \in [0,T]$, $x,\tilde{x} \in \mathbb{R}$ and $\rho,\tilde{\rho} \in \mathcal{P}_1(\mathbb{R})$, and σ is Hölder continuous of order $\frac{1}{2}+\xi$ for some $\xi \in [0,\frac{1}{2}]$ in x uniformly in t, i.e. there is a constant $C_\sigma > 0$ such that

$$|\sigma(t,x) - \sigma(t,\tilde{x})| \leq C_\sigma |x-\tilde{x}|^{\frac{1}{2}+\xi},$$

holds for all $t \in [0,T]$ and $x,\tilde{x} \in \mathbb{R}$.

First, we establish the well-posedness of the mean-field stochastic Volterra equation (5.10) with Hölder continuous diffusion coefficients. Its proof is based on a Yamada–Watanabe type approach [YW71], which requires essentially a one-dimensional setting and leads to the stronger assumptions on the kernels.

Theorem 5.10. *Suppose Assumption 5.9, and the kernels K_μ, K_σ and the initial condition X_0 satisfy Assumption 5.7 or Assumption 5.8 with p given as therein. Then, the mean-field stochastic Volterra equation* (5.10) *is well-posed in L^p. Moreover, for any $q \geq 1$, if $X_0 \in L^q(\Omega;\mathbb{R}^d)$, the unique solution X of* (5.10) *satisfies*

$$\sup_{t\in[0,T]} \mathbb{E}[|X_t|^q] < \infty.$$

Secondly, we establish propagation of chaos for one-dimensional stochastic mean-field SVEs with Hölder continuous diffusion coefficients. To that end, we consider the symmetric system of N mean-field stochastic Volterra equations

$$X_t^{N,i} = X_0^i + \int_0^t K_\mu(s,t)\mu(s,X_s^{N,i},\bar{\rho}_s^N)\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s,X_s^{N,i})\,\mathrm{d}B_s^i, \quad t \in [0,T], \tag{5.11}$$

for $i \in \{1,\dots,N\}$, where $(X_0^i)_{i\in\mathbb{N}} \subset L^p(\Omega;\mathbb{R})$ is an i.i.d. sequence of initial conditions, and $(B^i)_{i\in\mathbb{N}}$ is a sequence of independent one-dimensional Brownian motions. Moreover, for $i \in \mathbb{N}$, $\underline{X}^i$ denotes the solution of the mean-field SVE (5.10) with initial condition X_0^i and driving Brownian motion B^i. In the present one-dimensional setting, we obtain the following convergence result.

Theorem 5.11 (Volterra propagation of chaos)**.** *Suppose Assumption 5.9, and the kernels K_μ, K_σ and the initial conditions X_0^i, for $i \in \mathbb{N}$, satisfy Assumption 5.7 or Assumption 5.8 with p given as therein. Then, the system* (5.11) *of mean-field SVEs is well-posed in L^p,* *where the unique L^p-solution is denoted by $(X_t^{N,i})_{i=1,\dots,N}$ for every $N \geq 1$. Moreover, it holds*

$$\lim_{N\to\infty}\bigg(\max_{1\leq i\leq N}\Big(\sup_{t\in[0,T]}\mathbb{E}[|X_t^{N,i}-\underline{X}_t^i|]\Big)+\sup_{t\in[0,T]}\mathbb{E}\Big[W_1\Big(\frac{1}{N}\sum_{i=1}^N\delta_{X_t^{N,i}},\mathcal{L}(\underline{X}_t^1)\Big)\Big]\bigg)=0. \quad (5.12)$$

The rate of convergence in (5.12) is explicitly stated in the next lemma.

Lemma 5.12. *Supposing the assumptions and notation of Theorem 5.11, it holds that*

$$\max_{1\leq i\leq N}\Big(\sup_{t\in[0,T]}\mathbb{E}[|X_t^{N,i}-\underline{X}_t^i|]\Big)+\sup_{t\in[0,T]}\mathbb{E}\Big[W_1\Big(\frac{1}{N}\sum_{i=1}^N\delta_{X_t^{N,i}},\mathcal{L}(\underline{X}_t^1)\Big)\Big]\lesssim N^{-1/2}. \quad (5.13)$$

Remark 5.13. *The rate of convergence in* (5.13) *is expected to be optimal for synchronous coupling methods, cf. Remark 5.6, since it is shown in [FG15, Theorem 1 and there after] that for terms of the form $\mathbb{E}[W_1(\bar{\rho}_N,\rho)]$ the rate is sharp. Consequently, optimality could be only lost in the inequalities* (5.36) *or* (5.47)*.*

5.2 On the well-posedness of ordinary stochastic Volterra equations

In this section, we provide various well-posedness results for ordinary stochastic Volterra equations with random initial conditions that are needed to prove the well-posedness results for mean-field stochastic Volterra equations presented in Section 5.1. We start with SVEs with Lipschitz continuous coefficients, which is a slight modification of [Wan08, Theorem 1.1].

Lemma 5.14. *Let the kernels K_μ, K_σ fulfill Assumption 5.1, $p > \max\{\frac{1}{\gamma}, 1+\frac{2}{\varepsilon}\}$ with $\gamma \in (0,\frac{1}{2}]$ and $\varepsilon > 0$ from Assumption 5.1, the initial value $X_0 \in L^p(\Omega;\mathbb{R}^d)$, and the measurable coefficients $\mu\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon [0,T]\times\mathbb{R}^d\to\mathbb{R}^{d\times m}$ for some $d,m\in\mathbb{N}$ fulfill the linear growth condition*

$$|\mu(t,x)|+|\sigma(t,x)|\leq C_{\mu,\sigma}(1+|x|),$$

for some $C_{\mu,\sigma}>0$ and all $t\in[0,T]$, $x\in\mathbb{R}^d$, and the Lipschitz condition

$$|\mu(t,x)-\mu(t,y)|+|\sigma(t,x)-\sigma(t,y)|\leq C_{\mu,\sigma}|x-y|,$$

for some $C_{\mu,\sigma}>0$ and all $t\in[0,T]$, $x,y\in\mathbb{R}^d$. Then, the d-dimensional stochastic Volterra equation

$$X_t = X_0+\int_0^t K_\mu(s,t)\mu(s,X_s)\,\mathrm{d}s+\int_0^t K_\sigma(s,t)\sigma(s,X_s)\,\mathrm{d}B_s,\quad t\in[0,T],$$

is well-posed in L^p, where $(B_t)_{t\in[0,T]}$ is an m-dimensional Brownian motion.

Proof. With the assumed integrability on X_0, it is straightforward to adapt the Picard iteration and the Grönwall type estimates in proof of [Wan08, Theorem 1.1] to allow for random initial conditions X_0, as stated in Lemma 5.14. $\square$

For one-dimensional ordinary stochastic Volterra equations the Lipschitz assumption on the diffusion coefficients can be relaxed to Hölder continuity, provided the kernels are sufficiently regular or have a convolutional structure. The next results is a slight modification of Theorem 2.3, allowing for SVEs with random initial conditions.

Lemma 5.15. *Let the kernels K_μ, K_σ fulfill Assumption 5.7, $p > \max\{\frac{1}{\gamma}, 1 + \frac{2}{\varepsilon}\}$ with $\gamma \in (0, \frac{1}{2}]$ and $\varepsilon > 0$ from Assumption 5.7, the initial value $X_0 \in L^p(\Omega; \mathbb{R})$, and the measurable coefficients $\mu\colon [0,T] \times \mathbb{R} \to \mathbb{R}$ and $\sigma\colon [0,T] \times \mathbb{R} \to \mathbb{R}$ fulfill the linear growth condition*

$$|\mu(t,x)| + |\sigma(t,x)| \leq C_{\mu,\sigma}(1 + |x|),$$

for some $C_{\mu,\sigma} > 0$ and all $t \in [0,T]$, $x \in \mathbb{R}$, μ the Lipschitz condition

$$|\mu(t,x) - \mu(t,y)| \leq C_\mu |x - y|,$$

for some $C_\mu > 0$ and all $t \in [0,T]$, $x, y \in \mathbb{R}$, and σ the Hölder condition

$$|\sigma(t,x) - \sigma(t,y)| \leq C_\sigma |x - y|^{\frac{1}{2}+\xi},$$

for $\xi \in [0, \frac{1}{2}]$, some $C_\sigma > 0$ and all $t \in [0,T]$, $x, y \in \mathbb{R}$. Then, the stochastic Volterra equation

$$X_t = X_0 + \int_0^t K_\mu(s,t)\mu(s,X_s)\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s,X_s)\,\mathrm{d}B_s, \quad t \in [0,T],$$

is well-posed in L^p, where $(B_t)_{t\in[0,T]}$ is a one-dimensional Brownian motion.

Proof. With the assumed integrability on X_0, it is straightforward to adapt the proof of Theorem 2.3 to the case that X_0 is a random variable. $\square$

The next lemma is a slight generalization of [AJEE19b, Proposition B.3], providing the well-posedness of one-dimensional SVEs with convolutional kernels and random initial conditions.

Lemma 5.16. *Suppose that $X_0 \in L^p(\Omega; \mathbb{R})$ for some $p > 2$, the kernels are of the form $K_\mu(s,t) = K_\sigma(s,t) = \tilde{K}(t-s)$ for some $\tilde{K} \in C^1([0,T]; \mathbb{R})$, and the measurable coefficients $\mu\colon [0,T] \times \mathbb{R} \to \mathbb{R}$ and $\sigma\colon [0,T] \times \mathbb{R} \to \mathbb{R}$ fulfill the linear growth condition*

$$|\mu(t,x)| + |\sigma(t,x)| \leq C_{\mu,\sigma}(1 + |x|),$$

for some $C_{\mu,\sigma} > 0$ and all $t \in [0,T]$, $x \in \mathbb{R}$, μ the Lipschitz condition

$$|\mu(t,x) - \mu(t,y)| \leq C_\mu |x - y|,$$

for some $C_\mu > 0$ *and all* $t \in [0,T]$*,* $x, y \in \mathbb{R}$*, and* σ *the Hölder condition*

$$|\sigma(t,x) - \sigma(t,y)| \leq C_\sigma |x-y|^{\frac{1}{2}+\xi},$$

for $\xi \in [0, \frac{1}{2}]$*, some* $C_\sigma > 0$ *and all* $t \in [0,T]$*,* $x, y \in \mathbb{R}$*. Then, the stochastic Volterra equation*

$$X_t = X_0 + \int_0^t \tilde{K}(t-s)\mu(s, X_s)\,\mathrm{d}s + \int_0^t \tilde{K}(t-s)\sigma(s, X_s)\,\mathrm{d}B_s, \quad t \in [0,T], \tag{5.14}$$

is well-posed in L^p*, where* $(B_t)_{t\in[0,T]}$ *is a one-dimensional Brownian motion.*

Proof. The weak existence of some L^p-solution to the SVE (5.14) follows from Theorem 3.10 with the straightforward adaptation to random initial conditions X_0. For the pathwise uniqueness, one can adapt the proof from [AJEE19b, Proposition B.3] using the Lipschitz and Hölder continuity of μ, σ uniformly in t. $\square$

Moreover, for the well-posedness results of mean-field SVEs we need a multi-dimensional well-posedness result for stochastic Volterra equations where the Hölder continuous coefficient σ is a diagonal matrix, where each entry only depends on the component of the solution of the respective dimension, as provided in the next remark.

Remark 5.17. *For* $N \in \mathbb{N}$ *let us consider the* N*-dimensional stochastic Volterra equation*

$$X_t = X_0 + \int_0^t K_\mu(s,t)\mu(s, X_s)\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s, X_s)\,\mathrm{d}B_s, \quad t \in [0,T], \tag{5.15}$$

where $(B_t)_{t\in[0,T]}$ *is an* N*-dimensional Brownian motion,*

$$X_t = \begin{pmatrix} X_t^1 \\ \vdots \\ X_t^N \end{pmatrix}, \quad X_0 = \begin{pmatrix} X_0^1 \\ \vdots \\ X_0^N \end{pmatrix}, \quad \mu(s, X_s) = \begin{pmatrix} \mu_1(s, X_s) \\ \vdots \\ \mu_N(s, X_s) \end{pmatrix}$$

and

$$\sigma(s, X_s) = \begin{pmatrix} \sigma_1(s, X_s^1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N(s, X_s^N) \end{pmatrix}.$$

Suppose that the kernels K_μ, K_σ *and the initial value* X_0 *fulfill Assumption 5.7 or Assumption 5.8 with* p *given from there, that* $\mu\colon [0,T] \times \mathbb{R}^N \to \mathbb{R}^N$ *is Lipschitz continuous in the space variable, uniformly in the time variable, and each* $\sigma_i\colon [0,T] \times \mathbb{R} \to \mathbb{R}$ *for* $i \in \{1, \dots, N\}$ *is* $1/2 + \xi$*-Hölder continuous in the space variable, uniformly in the time variable for some* $\xi \in [0, 1/2]$*. By considering each dimension separately, as e.g. done for SDEs in [YW71, Theorem 1], it is straightforward to conclude the well-posedness in* L^p *of the SVE* (5.15) *from the corresponding one-dimensional results in Lemma 5.15 and Lemma 5.16.*

We conclude this section with a remark on the path regularity of solutions and one on the notion of L^p-well-posedness.

Remark 5.18 (Path regularity)**.** *Let X be the unique (d-, 1- or N-dimensional) solution to the stochastic Volterra equation in either of the settings in Lemma 5.14, Lemma 5.15, Lemma 5.16 or Remark 5.17 with $p > \max\{\frac{1}{\gamma}, 1 + \frac{2}{\varepsilon}\}$. In the case of Assumption 5.8, we can set $\gamma = \frac{1}{2}$ and $p > 2$ given from there. By adapting Lemma 2.7 and Lemma 2.10 to the multi-dimensional setting, it follows that*

$$\sup_{t\in[0,T]} \mathbb{E}[|X_t|^q] < \infty,$$

and

$$\mathbb{E}[|X_t - X_s|^q] \lesssim |t-s|^{\beta q}$$

for any $q \geq 1$, $\beta \in (0, \gamma - \frac{1}{p})$, $s,t \in [0,T]$, and, hence, that the solution X has a modification with β-Hölder continuous sample paths.

Remark 5.19. *The notion of L^p-well-posedness, as used Lemma 5.14, Lemma 5.15, Lemma 5.16 and Remark 5.17, appears to be necessary to prove the existence of a strong solution and pathwise uniqueness. First, one needs to assume that a solution X is in $L^p(\Omega \times [0,T]; \mathbb{R}^d)$ to conclude continuity of its sample paths with standard estimates, as in Lemma 2.7. Secondly, in order to be able to apply Grönwall's Lemma to an inequality of the form*

$$\mathbb{E}[|X_t - Y_t|^p] \lesssim \int_0^t \mathbb{E}[|X_s - Y_s|^p] \,\mathrm{d}s,$$

one needs to assume that both solutions X, Y are in $L^p(\Omega \times [0,T]; \mathbb{R}^d)$ to guarantee finiteness of the expectations $\sup_{s\in[0,t]} \mathbb{E}[|X_s|^p]$ and $\sup_{s\in[0,t]} \mathbb{E}[|Y_s|^p]$ by standard estimates, as in Lemma 2.10.

5.3 Well-posedness: Proof of Theorem 5.3 and 5.10

This section is devoted to the proofs of Theorem 5.3 and of Theorem 5.10.

Proof of Theorem 5.3. We define the solution map Φ by

$$\Phi\colon C\big([0,T]; \mathcal{P}_\delta(\mathbb{R}^d)\big) \to C\big([0,T]; \mathcal{P}_\delta(\mathbb{R}^d)\big), \quad \rho \mapsto \Phi(\rho) := \big(\mathcal{L}(X_t^\rho)\big)_{t\in[0,T]}, \tag{5.16}$$

where X^ρ is the unique L^p-solution to the stochastic Volterra equation

$$X_t = X_0 + \int_0^t K_\mu(s,t)\mu(s, X_s, \rho_s) \,\mathrm{d}s + \int_0^t K_\sigma(s,t)\sigma(s, X_s, \rho_s) \,\mathrm{d}B_s, \quad t \in [0,T]. \tag{5.17}$$

Note that a unique fixed point of the solution map Φ implies the existence of a unique L^p-solution $X = (X_t)_{t\in[0,T]}$ to the mean-field SVE (5.2) satisfying $\sup_{t\in[0,T]} \mathbb{E}[|X_t|^q] < \infty$

for every $q \geq 1$, c.f. Step 1 below. Hence, it is sufficient to prove that the solution map Φ has a unique fixed point.

Step 1: We show the well-definedness of the solution map Φ.

For a fixed $\rho = (\rho_t)_{t\in[0,T]} \in C([0,T];\mathcal{P}_\delta(\mathbb{R}^d))$, the integral equation (5.17) is an ordinary stochastic Volterra equation. Due to Assumption 5.2, the linear growth and Lipschitz condition of Lemma 5.14 are satisfied. Hence, there exists a unique strong L^p-solution $X^\rho = (X^\rho_t)_{t\in[0,T]}$ to the SVE (5.17) and, by Remark 5.18, we get that $\sup_{t\in[0,T]} \mathbb{E}[|X^\rho_t|^q] < \infty$ for all $q \geq 1$ and that the sample paths of X^ρ are almost surely continuous. Moreover, note that $\big(\mathcal{L}(X^\rho_t)\big)_{t\in[0,T]} \in C\big([0,T];\mathcal{P}_\delta(\mathbb{R}^d)\big)$, since, by the representation of the Wasserstein distance in terms of random variables (see [CD18a, (5.14)]) and by Remark 5.18, we have

$$W_\delta\big(\mathcal{L}(X^\rho_t), \mathcal{L}(X^\rho_s)\big) \leq \mathbb{E}\big[|X^\rho_t - X^\rho_s|^\delta\big]^{\frac{1}{\delta}} \lesssim |t-s|^\beta, \quad s,t \in [0,T],$$

for any $\beta \in (0, \gamma - 1/p)$ with $\gamma \in (0,\frac{1}{2}]$, where the parameters are given in Assumption 5.1.

Step 2: For $\rho, \tilde{\rho} \in C\big([0,T];\mathcal{P}_\delta(\mathbb{R}^d)\big)$, we show that

$$\sup_{s\in[0,t]} W_\delta(\Phi(\rho)_s, \Phi(\tilde{\rho})_s)^\delta \lesssim \int_0^t W_\delta(\rho_s, \tilde{\rho}_s)^\delta \,\mathrm{d}s, \quad t \in [0,T]. \tag{5.18}$$

We get that

$$\begin{aligned}
\mathbb{E}\big[|X^\rho_t - X^{\tilde{\rho}}_t|^\delta\big] &\lesssim \mathbb{E}\bigg[\Big|\int_0^t K_\mu(s,t)\big(\mu(s,X^\rho_s,\rho_s) - \mu(s,X^{\tilde{\rho}}_s,\tilde{\rho}_s)\big)\,\mathrm{d}s\Big|^\delta\bigg] \\
&\quad + \mathbb{E}\bigg[\Big|\int_0^t K_\sigma(s,t)\big(\sigma(s,X^\rho_s,\rho_s) - \sigma(s,X^{\tilde{\rho}}_s,\tilde{\rho}_s)\big)\,\mathrm{d}B_s\Big|^\delta\bigg] \\
&\lesssim \Big(\int_0^t |K_\mu(s,t)|^{\frac{4+2\varepsilon}{4+\varepsilon}}\,\mathrm{d}s\Big)^{\frac{4+\varepsilon}{1+\varepsilon}} \int_0^t \mathbb{E}\big[\big|\mu(s,X^\rho_s,\rho_s) - \mu(s,X^{\tilde{\rho}}_s,\tilde{\rho}_s)\big|^\delta\big]\,\mathrm{d}s \\
&\quad + \mathbb{E}\bigg[\Big(\int_0^t \big|K_\sigma(s,t)\big(\sigma(s,X^\rho_s,\rho_s) - \sigma(s,X^{\tilde{\rho}}_s,\tilde{\rho}_s)\big)\big|^2\,\mathrm{d}s\Big)^{\frac{\delta}{2}}\bigg] \\
&\lesssim \int_0^t \mathbb{E}\big[\big|\mu(s,X^\rho_s,\rho_s) - \mu(s,X^{\tilde{\rho}}_s,\tilde{\rho}_s)\big|^\delta\big]\,\mathrm{d}s \\
&\quad + \Big(\int_0^t |K_\sigma(s,t)|^{2+\varepsilon}\,\mathrm{d}s\Big)^{\frac{4+2\varepsilon}{\varepsilon(2+\varepsilon)}} \int_0^t \mathbb{E}\big[\big|\sigma(s,X^\rho_s,\rho_s) - \sigma(s,X^{\tilde{\rho}}_s,\tilde{\rho}_s)\big|^\delta\big]\,\mathrm{d}s \\
&\lesssim \int_0^t \Big(\mathbb{E}\big[\big|X^\rho_s - X^{\tilde{\rho}}_s\big|^\delta\big] + W_\delta(\rho_s,\tilde{\rho}_s)^\delta\Big)\,\mathrm{d}s.
\end{aligned} \tag{5.19}$$

for $t \in [0,T]$, where we used Hölder's inequality in the drift integral with $\frac{4+2\varepsilon}{4+\varepsilon} < 1+\varepsilon$ such that by the choice of δ in (5.3), $\frac{4+\varepsilon}{4+2\varepsilon} + \frac{1}{\delta} = 1$ and in the diffusion integral with $\frac{2+\varepsilon}{2}$ such that (5.4) holds, Burkholder–Davis–Gundy's inequality, Fubini's theorem, the integrability of the kernels from Assumption 5.1 and the Lipschitz continuity of μ and σ

from Assumption 5.2. Since we have that

$$\sup_{s\in[0,T]} \mathbb{E}[|X_s^{\rho} - X_s^{\tilde{\rho}}|^{\delta}] < \infty,$$

we can apply Grönwall's inequality to conclude that

$$\mathbb{E}\big[|X_t^{\rho} - X_t^{\tilde{\rho}}|^{\delta}\big] \lesssim \int_0^t W_\delta(\rho_s, \tilde{\rho}_s)^{\delta} \,\mathrm{d}s. \tag{5.20}$$

Since by assumption $\rho, \tilde{\rho} \in C\big([0,T]; \mathcal{P}_\delta(\mathbb{R}^d)\big)$, we can bound the Wasserstein distance by

$$W_\delta(\Phi(\rho)_t, \Phi(\tilde{\rho})_t) = W_\delta(\mathcal{L}(X_t^{\rho}), \mathcal{L}(X_t^{\tilde{\rho}})) \le \mathbb{E}[|X_t^{\rho} - X_t^{\tilde{\rho}}|^{\delta}]^{\frac{1}{\delta}},$$

c.f. [CD18a, (5.14)], and plugging this into (5.20) and taking the supremum, we obtain (5.18).

Step 3: We show that the solution map Φ has a unique fixed point.

First note that it is sufficient to show that Φ^k is a contraction, see [Bry68, Theorem], since the Wasserstein space $C([0,T]; \mathcal{P}_\delta(\mathbb{R}^d))$ is a complete metric space, see e.g. [PZ20, Proposition 2.2.8], where Φ^k denotes the k-th composition of Φ with itself. Let $C > 0$ denote the generic constant in (5.18). Then, we get iteratively for $k \in \mathbb{N}$,

$$\begin{aligned}
\sup_{s\in[0,T]} W_\delta(\Phi^k(\rho)_s, \Phi^k(\tilde{\rho})_s)^{\delta} &\le C^k \int_0^T \frac{(T-s)^{k-1}}{(k-1)!} W_\delta(\rho_s, \tilde{\rho}_s)^{\delta} \,\mathrm{d}s \\
&\le \frac{C^k T^k}{k!} \sup_{s\in[0,T]} W_\delta(\rho_s, \tilde{\rho}_s)^{\delta}.
\end{aligned}$$

Thus, choosing k large enough such that $\frac{C^kT^k}{k!} < 1$, we see that the mapping Φ^k is a contraction and, hence, Φ admits a unique fixed point, which completes the proof. $\square$

Next, we provide the proof of Theorem 5.10. We keep its presentation fairly short since it is in parts similar to the proof of Theorem 5.3.

Proof of Theorem 5.10. We again consider the solution map Φ as defined in (5.16) but choose $\delta = 1$ and $d = 1$, that is,

$$\Phi\colon C\big([0,T]; \mathcal{P}_1(\mathbb{R})\big) \to C\big([0,T]; \mathcal{P}_1(\mathbb{R})\big), \quad \rho \mapsto \Phi(\rho) := \big(\mathcal{L}(X_t^{\rho})\big)_{t\in[0,T]}.$$

In the following we show that the solution map Φ possesses a unique fixed point. We proceed as in the proof of Theorem 5.3. *Step 1* works exactly the same, using Lemma 5.15 and Lemma 5.16, respectively, instead of Lemma 5.14, and *Step 3* works exactly the same. That means we only need to show *Step 2*, or more precisely, estimate (5.20) with $\delta = 1$. To do that, we treat the cases that Assumption 5.7 or that Assumption 5.8 holds separately.

Case (i): Suppose the kernels K_μ, K_σ and initial condition X_0 satisfy Assumption 5.7.

To get an analogue estimate as (5.20), we use the semimartingale property of a solution $(X_t^\rho)_{t\in[0,T]}$ to (5.2) with fixed $\rho \in \mathcal{P}_1(\mathbb{R})$ (cf. Lemma 2.12 or [Pro85, Theorem 3.3]),

$$\begin{aligned} X_t^\rho - X_0 &= \int_0^t K_\sigma(s,s)\sigma(s,X_s^\rho,\rho_s)\,\mathrm{d}B_s + \int_0^t K_\mu(s,s)\mu(s,X_s^\rho,\rho_s)\,\mathrm{d}s \\ &\quad + \int_0^t \left(\int_0^s \partial_2 K_\mu(u,s)\mu(u,X_u^\rho,\rho_u)\,\mathrm{d}u + \int_0^s \partial_2 K_\sigma(u,s)\sigma(u,X_u^\rho,\rho_u)\,\mathrm{d}B_u \right) \mathrm{d}s, \end{aligned}$$

and the Yamada–Watanabe functions ϕ_n for $n \in \mathbb{N}$ (cf. the proof of Theorem 2.22 or the original work [YW71]) that approximate the absolute value function in the following way: Let $(a_n)_{n\in\mathbb{N}}$ be a strictly decreasing sequence with $a_0 = 1$ such that $a_n \to 0$ as $n \to \infty$ and

$$\int_{a_n}^{a_{n-1}} \frac{1}{|x|^{1+2\xi}}\,\mathrm{d}x = n,$$

where $\frac{1}{2} + \xi$ is the Hölder regularity of σ. Furthermore, we define a sequence of mollifiers: let $(\psi_n)_{n\in\mathbb{N}} \in C_0^\infty(\mathbb{R})$ be smooth functions with compact support such that $\operatorname{supp}(\psi_n) \subset (a_n, a_{n-1})$, and with the properties

$$0 \le \psi_n(x) \le \frac{2}{n|x|^{1+2\xi}}, \quad \forall x \in \mathbb{R}, \quad \text{and} \quad \int_{a_n}^{a_{n-1}} \psi_n(x)\,\mathrm{d}x = 1. \tag{5.21}$$

We set

$$\phi_n(x) := \int_0^{|x|} \left(\int_0^y \psi_n(z)\,\mathrm{d}z \right) \mathrm{d}y, \quad x \in \mathbb{R}.$$

By (5.21) and the compact support of ψ_n, it follows that $\phi_n(\cdot) \to |\cdot|$ uniformly as $n \to \infty$. Since every ψ_n and, thus, every ϕ_n is zero in a neighborhood around zero, the functions ϕ_n are smooth with

$$\|\phi_n'\|_\infty \le 1, \quad \phi_n'(x) = \operatorname{sgn}(x) \int_0^{|x|} \psi_n(y)\,\mathrm{d}y, \quad \text{and} \quad \phi_n''(x) = \psi_n(|x|) \quad \text{for } x \in \mathbb{R},$$

where $\|\cdot\|_\infty$ denotes the sup-norm on $\mathbb{R}$.
Using ϕ_n, we apply Itô's formula to $\tilde{X}_t := X_t^\rho - X_t^{\tilde{\rho}}$, with the notation

$$\tilde{Z}_t := \int_0^t \big(\mu(s,X_s^\rho,\rho_s) - \mu(s,X_s^{\tilde{\rho}},\tilde{\rho}_s)\big)\,\mathrm{d}s, \;\; Y_t^\rho := \int_0^t \sigma(s,X_s^\rho)\,\mathrm{d}B_s, \;\; H_t^\rho := \int_0^t \partial_2 K_\sigma(s,t)\,\mathrm{d}Y_s^\rho,$$

and $Y_t^{\tilde{\rho}}$ and $H_t^{\tilde{\rho}}$ analogue, as well as $\tilde{Y}_t := Y_t^\rho - Y_t^{\tilde{\rho}}$, and $\tilde{H}_t := H_t^\rho - H_t^{\tilde{\rho}}$, for $t \in [0,T]$, to obtain

$$
\begin{aligned}
\phi_n(\tilde{X}_t) &= \int_0^t \phi_n'(\tilde{X}_s)\,\mathrm{d}\tilde{X}_s + \frac{1}{2}\int_0^t \phi_n''(\tilde{X}_s)\,\mathrm{d}\langle \tilde{X}\rangle_s \\
&= \int_0^t \phi_n'(\tilde{X}_s)K_\mu(s,s)(\mu(s,X_s^\rho,\rho_s) - \mu(s,X_s^{\tilde{\rho}},\tilde{\rho}_s))\,\mathrm{d}s \\
&\quad + \int_0^t \phi_n'(\tilde{X}_s)\left(\int_0^s \partial_2 K_\mu(u,s)\,\mathrm{d}\tilde{Z}_u\right)\mathrm{d}s \\
&\quad + \int_0^t \phi_n'(\tilde{X}_s)\tilde{H}_s\,\mathrm{d}s + \int_0^t \phi_n'(\tilde{X}_s)K_\sigma(s,s)\,\mathrm{d}\tilde{Y}_s \\
&\quad + \frac{1}{2}\int_0^t \phi_n''(\tilde{X}_s)K_\sigma(s,s)^2\big(\sigma(s,X_s^\rho) - \sigma(s,X_s^{\tilde{\rho}})\big)^2\,\mathrm{d}s \\
&=: I_{1,t}^n + I_{2,t}^n + I_{3,t}^n + I_{4,t}^n + I_{5,t}^n.
\end{aligned} \tag{5.22}
$$

For $I_{1,t}^n$, the bound $\|\phi_n'\|_\infty \le 1$, boundedness of K_μ, Lipschitz continuity of μ, and Jensen's inequality yield

$$
\mathbb{E}[I_{1,t}^n] \lesssim \int_0^t \big(\mathbb{E}[|\tilde{X}_s|] + W_1(\rho_s,\tilde{\rho}_s)\big)\,\mathrm{d}s. \tag{5.23}
$$

For $I_{2,t}^n$, we additionally use the boundedness of $\partial_2 K_\mu(u,s)$ on Δ_T to obtain

$$
\mathbb{E}[I_{2,t}^n] \lesssim \int_0^t \big(\mathbb{E}[|\tilde{X}_s|] + W_1(\rho_s,\tilde{\rho}_s)\big)\,\mathrm{d}s. \tag{5.24}
$$

For $I_{3,t}^n$, we use $\|\phi_n'\|_\infty \le 1$ and the integration by parts formula to estimate

$$
\begin{aligned}
\mathbb{E}[I_{3,t}^n] &\le \int_0^t \mathbb{E}[|\tilde{H}_s|]\,\mathrm{d}s \\
&\le \int_0^t |\partial_2 K_\sigma(s,s)|\mathbb{E}[|\tilde{Y}_s|]\,\mathrm{d}s + \int_0^t\int_0^s |\partial_{21}K_\sigma(u,s)|\mathbb{E}[|\tilde{Y}_u|]\,\mathrm{d}u\,\mathrm{d}s \\
&\le \int_0^t \mathbb{E}[|\tilde{Y}_s|]\left(\partial_2 K_\sigma(s,s) + \int_s^t |\partial_{21}K_\sigma(s,u)|\,\mathrm{d}u\right)\mathrm{d}s \\
&\lesssim \int_0^t \mathbb{E}[|\tilde{Y}_s|]\,\mathrm{d}s,
\end{aligned} \tag{5.25}
$$

with the boundedness of $\partial_2 K_\sigma(s,s)$ and $\int_s^t \partial_{21}K_\sigma(s,u)\,\mathrm{d}u$ from Assumption 5.7. For $I_{4,t}^n$, since $I_{4,t}^n$ is a martingale by [Pro04, p. 73, Corollary 3] due to the boundedness of K_σ, the growth bound on σ and the finiteness of the moments of X^ρ and $X^{\tilde{\rho}}$ (cf. Lemma 2.10), we get

$$
\mathbb{E}[I_{4,t}^n] = \mathbb{E}\left[\int_0^t \phi_n'(\tilde{X}_s)K_\sigma(s,s)(\sigma(s,X_s^\rho) - \sigma(s,X_s^{\tilde{\rho}}))\,\mathrm{d}B_s\right] = 0. \tag{5.26}
$$

For $I^n_{5,t}$, we get by using the boundedness of K_σ, the Hölder continuity of σ, and the inequality $\phi_n''(x) \le \frac{2}{n|x|^{1+2\xi}}$ that

$$\mathbb{E}[I^n_{5,t}] \lesssim \mathbb{E}\bigg[\int_0^t \phi_n''(\tilde{X}_s)|\tilde{X}_s|^{1+2\xi}\,\mathrm{d}s\bigg] \le \mathbb{E}\bigg[\int_0^t \frac{2}{n|\tilde{X}_s|^{1+2\xi}}|\tilde{X}_s|^{1+2\xi}\,\mathrm{d}s\bigg] \lesssim \frac{1}{n}. \tag{5.27}$$

Sending $n \to \infty$ and combining the five previous estimates (5.23), (5.24), (5.25), (5.26) and (5.27) with (5.22) yields

$$\mathbb{E}[|\tilde{X}_t|] \lesssim \int_0^t \Big(\mathbb{E}[|\tilde{X}_s|] + \mathbb{E}[|\tilde{Y}_s|] + W_1(\rho_s, \tilde{\rho}_s)\Big)\,\mathrm{d}s. \tag{5.28}$$

To apply Grönwall's lemma, we set $M(t) := \mathbb{E}[|\tilde{X}_t|] + \mathbb{E}[|\tilde{Y}_t|]$ for $t \in [0,T]$. To find a bound for $\mathbb{E}[|\tilde{Y}_t|]$, we apply integration by part formula to obtain

$$\begin{aligned}\tilde{X}_t &= \int_0^t K_\mu(s,t)(\mu(s,X_s^\rho,\rho_s) - \mu(s,X_s^{\tilde{\rho}},\tilde{\rho}_s))\,\mathrm{d}s + \int_0^t K_\sigma(s,t)\,\mathrm{d}\tilde{Y}_s \\ &= \int_0^t K_\mu(s,t)(\mu(s,X_s^\rho,\rho_s) - \mu(s,X_s^{\tilde{\rho}},\tilde{\rho}_s))\,\mathrm{d}s + K_\sigma(t,t)\tilde{Y}_t - \int_0^t \partial_1 K_\sigma(s,t)\tilde{Y}_s\,\mathrm{d}s\end{aligned} \tag{5.29}$$

keeping in mind that that $K_\sigma(\cdot,t)$ is absolutely continuous for every $t \in [0,T]$. Due to $|K_\sigma(t,t)| > C$ for some constant $C > 0$, we can rearrange (5.29) and use (5.28) to get

$$\begin{aligned}\mathbb{E}\Big[|\tilde{Y}_t|\Big] \le &C\bigg(\int_0^t \mathbb{E}\big[|\mu(s,X_s^\rho,\rho_s) - \mu(s,X_s^{\tilde{\rho}},\tilde{\rho}_s)|\big]\,\mathrm{d}s \\ &+ \mathbb{E}\big[|\tilde{X}_t|\big] + \int_0^t |\partial_1 K_\sigma(s,t)|\mathbb{E}\big[|\tilde{Y}_s|\big]\,\mathrm{d}s\bigg) \\ \lesssim &\int_0^t \Big(\mathbb{E}\big[|\tilde{X}_s|\big] + \mathbb{E}[|\tilde{Y}_s|] + W_1(\rho_s,\tilde{\rho}_s)\Big)\,\mathrm{d}s.\end{aligned} \tag{5.30}$$

Now, Grönwall's Lemma applied to (5.28) and (5.30) yields $M(t) \lesssim \int_0^t W_1(\rho_s,\tilde{\rho}_s)\,\mathrm{d}s$ and hence $\mathbb{E}[|X_t^\rho - X_t^{\tilde{\rho}}|] \lesssim \int_0^t W_1(\rho_s,\tilde{\rho}_s)\,\mathrm{d}s$, which is the analogue estimate of (5.20).

Case (ii): Suppose the kernels K_μ, K_σ and initial condition X_0 to satisfy Assumption 5.8. We need to find an analogue to estimate (5.20). By using the notation $\tilde{X}_t := X_t^\rho - X_t^{\tilde{\rho}}$ and $Y_t^\rho := \int_0^t \mu(s,X_s^\rho,\rho_s)\,\mathrm{d}s + \int_0^t \sigma(s,X_s^\rho)\,\mathrm{d}B_s$, $Y_t^{\tilde{\rho}}$ analogue, $\tilde{Y}_t := Y_t^\rho - Y_t^{\tilde{\rho}}$ and the semimartingale property

$$X_t^\rho - X_0 = \int_0^t \tilde{K}(0)\,\mathrm{d}Y_s^\rho + \int_0^t \int_0^s \tilde{K}'(s-u)\,\mathrm{d}Y_u^\rho\,\mathrm{d}s,$$

we can implement the Yamada–Watanabe approach with

$$\begin{aligned}\phi_n(\tilde{X}_t) &= \int_0^t \phi_n'(\tilde{X}_s)\tilde{K}(0)\,\mathrm{d}\tilde{Y}_s + \int_0^t \phi_n'(\tilde{X}_s)\int_0^s \tilde{K}'(s-u)\,\mathrm{d}\tilde{Y}_u\,\mathrm{d}s \\ &\quad + \frac{1}{2}\int_0^t \phi_n''(\tilde{X}_s)\tilde{K}(0)^2\big(\sigma(s,X_s^{\rho}) - \sigma(s,X_s^{\tilde{\rho}})\big)^2\,\mathrm{d}s \\ &=: I_{1,t}^n + I_{2,t}^n + I_{3,t}^n. \end{aligned} \tag{5.31}$$

Now, the Lipschitz assumption on μ applied to $I_{1,t}^n$ and $I_{2,t}^n$ such as the Hölder assumption on σ applied to $I_{3,t}^n$, the boundedness of $\tilde{K}$ and $\tilde{K}'$, the inequalities $\|\phi_n\|_\infty \leq 1$ and $\phi_n''(x) \leq \frac{2}{n|x|^{1+2\xi}}$, and sending $n \to \infty$ yields as in *Case (i)* with Grönwall's Lemma the inequality $\mathbb{E}[|X_t^{\rho} - X_t^{\tilde{\rho}}|] \lesssim \int_0^t W_1(\rho_s, \tilde{\rho}_s)\,\mathrm{d}s$, which implies the estimate (5.20) and, hence, yields the claimed well-posedness of the mean-field SVE (5.10). □

Remark 5.20. *The well-posedness from Theorems 5.3 and 5.10 implies together with a general version of the classical Yamada–Watanabe result (see e.g. [Kur14, Theorem 1.5], see also [Kur14, Example 2.14]) that there is some measurable map $G\colon \mathbb{R}^d \times C([0,T];\mathbb{R}^m) \to C([0,T];\mathbb{R}^d)$ such that any solution X of* (5.2) *and* (5.10)*, respectively, given some initial value X_0 and Brownian motion B can be represented as $X = G(X_0, B)$. Hence, if $X, \tilde{X}$ are solutions of* (5.2) *and* (5.10)*, respectively, for initial values $X_0, \tilde{X}_0$ with the same law, and Brownian motions $B, \tilde{B}$, it is straightforward that $\mathcal{L}(X_t) = \mathcal{L}(\tilde{X}_t)$ a.s. for all $t \in [0,T]$.*

5.4 Propagation of chaos: Proof of Theorem 5.4 and 5.11

An important argument in the proofs of the propagation of chaos results will be to show that the coupled processes $((X^{N,i}, \underline{X}^i))_{1\leq i\leq N}$ are identically distributed. To that end, the following lemma plays a crucial role. Recall that a sequence of random variables $(\zeta^1, \zeta^2, \dots)$ is called exchangeable if for any $N \in \mathbb{N}$ the vectors $(\zeta^1, \dots, \zeta^N)$ and $(\zeta^{\sigma(1)}, \dots, \zeta^{\sigma(N)})$ have the same joint distribution, where $\{\sigma(1), \dots, \sigma(N)\}$ is an arbitrary permutation of $\{1, \dots, N\}$.

Lemma 5.21. *Let $(A, \mathcal{F}_A)$ and $(B, \mathcal{F}_B)$ be measurable spaces and let for some fixed $N \in \mathbb{N}$, $(\zeta^1, \dots, \zeta^N)$ be an exchangeable family of A-valued random variables. Let $F\colon A \to B$ be a measurable function and define the family of random variables $(X^1, \dots, X^N)$ by $X^i := F(\zeta^i)$ for $i \in \{1, \dots, N\}$. Further, let $G\colon A^N \to B^N$ be a measurable function which fulfills the following exchangeability property:*

$$(y_1, \dots, y_N) = G((x_1, \dots, x_N)) \Rightarrow (y_{\sigma(1)}, \dots, y_{\sigma(N)}) = G((x_{\sigma(1)}, \dots, x_{\sigma(N)})), \tag{5.32}$$

for arbitrary $x_1, \dots, x_N \in A$ and any permutation $\{\sigma(1), \dots, \sigma(N)\}$ of $\{1, \dots, N\}$. Define the family of random variables $(Y^1, \dots, Y^N)$ by

$$(Y^1, \dots, Y^N) := G\big((\zeta^1, \dots, \zeta^N)\big).$$

Then, the coupled family of random variables $((X^i, Y^i))_{1\leq i\leq N}$ is exchangeable.

Proof. Let $\{\sigma(1),\dots,\sigma(N)\}$ be an arbitrary permutation of $\{1,\dots,N\}$. By (5.32), we have that

$$\begin{aligned} Y^{\sigma(1)} &= G_1\big((\zeta^{\sigma(1)},\zeta^{\sigma(2)},\dots,\zeta^{\sigma(N-1)},\zeta^{\sigma(N)})\big),\\ Y^{\sigma(2)} &= G_1\big((\zeta^{\sigma(2)},\zeta^{\sigma(3)},\dots,\zeta^{\sigma(N)},\zeta^{\sigma(1)})\big),\\ &\dots\\ Y^{\sigma(N)} &= G_1\big((\zeta^{\sigma(N)},\zeta^{\sigma(1)},\dots,\zeta^{\sigma(N-2)},\zeta^{\sigma(N-1)})\big), \end{aligned} \tag{5.33}$$

where G_1 denotes the first component of the N-dimensional mapping G. Define $W^i := (X^i, Y^i)$ for $i\in\{1,\dots,N\}$. Then, by the definition of X^i and (5.33),

$$\begin{aligned} W^{\sigma(1)} &= \Big(F(\zeta^{\sigma(1)}), G_1\big((\zeta^{\sigma(1)},\zeta^{\sigma(2)},\dots,\zeta^{\sigma(N-1)},\zeta^{\sigma(N)})\big)\Big),\\ W^{\sigma(2)} &= \Big(F(\zeta^{\sigma(2)}), G_1\big((\zeta^{\sigma(2)},\zeta^{\sigma(3)},\dots,\zeta^{\sigma(N)},\zeta^{\sigma(1)})\big)\Big),\\ &\dots\\ W^{\sigma(N)} &= \Big(F(\zeta^{\sigma(N)}), G_1\big((\zeta^{\sigma(N)},\zeta^{\sigma(1)},\dots,\zeta^{\sigma(N-2)},\zeta^{\sigma(N-1)})\big)\Big). \end{aligned} \tag{5.34}$$

Analogous, we have

$$\begin{aligned} W^{1} &= \Big(F(\zeta^{1}), G_1\big((\zeta^{1},\zeta^{2},\dots,\zeta^{N-1},\zeta^{N})\big)\Big),\\ W^{2} &= \Big(F(\zeta^{2}), G_1\big((\zeta^{2},\zeta^{3},\dots,\zeta^{N},\zeta^{1})\big)\Big),\\ &\dots\\ W^{N} &= \Big(F(\zeta^{N}), G_1\big((\zeta^{N},\zeta^{1},\dots,\zeta^{N-2},\zeta^{N-1})\big)\Big). \end{aligned} \tag{5.35}$$

Now, since by assumption $(\zeta^1,\dots,\zeta^N)$ and $(\zeta^{\sigma(1)},\dots,\zeta^{\sigma(N)})$ have the same joint distribution, (5.34) and (5.35) yield that also $(W^1,\dots,W^N)$ and $(W^{\sigma(1)},\dots,W^{\sigma(N)})$ have the same joint distribution which proves the claimed exchangeability. □

We start with the proof of Theorem 5.4.

Proof of Theorem 5.4. Let us briefly outline the main steps of the proof:

Step 1: We show the existence of the system of processes $(X^{N,i})_{i=1,\dots,N}$ uniquely solving (5.6), for every $N\in\mathbb{N}$.

Step 2: We prove the inequality

$$\mathbb{E}[|X_t^{N,i}-\underline{X}_t^i|^\delta] \lesssim \int_0^t \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j},\mathcal{L}(\underline{X}_s^i)\Big)^\delta\Big]\,\mathrm{d}s,\quad t\in[0,T], \tag{5.36}$$

for any $1\le i\le N$. Recall that $X^{N,i}$ is defined in (5.6) and $\underline{X}^i$ is defined as the solution of the mean-field SVE (5.2) with initial condition X_0^i and driving Brownian motion B^i.

Step 3: We prove that the right-hand side of (5.36) tends to zero.

Step 4: We show that *Step 2* and *Step 3* imply the statement.

Step 1: By the Lipschitz continuity of μ and σ, and the observation that $W_\delta(\bar\rho_x^N,\bar\rho_y^N)^\delta \le \frac{1}{N}\sum_{j=1}^N |x_j-y_j|^\delta$ for $x,y\in\mathbb{R}^{N\times d}$ with the notation $\bar\rho_x^N = \frac{1}{N}\sum_{j=1}^N \delta_{x_j} \in \mathcal{P}_\delta(\mathbb{R}^d)$, we obtain for every $i\in\{1,\dots,N\}$ the Lipschitz condition

$$\begin{aligned}\big|\mu(t,x_i,\bar\rho_x^N)-\mu(t,y_i,\bar\rho_y^N)\big|^\delta + \big|\sigma(t,x_i,\bar\rho_x^N)-\sigma(t,y_i,\bar\rho_y^N)\big|^\delta &\lesssim |x_i-y_i|^\delta + \frac{1}{N}\sum_{j=1}^N |x_j-y_j|^\delta \\ &\lesssim \|x-y\|_{N\times d}^\delta,\end{aligned}$$

where $\|\cdot\|_{N\times d}$ denotes the row sum norm on $\mathbb{R}^{N\times d}$. With the notation $\tilde\mu_i(t,x) := \mu(t,x_i,\bar\rho_x^N)$ and $\tilde\sigma_i(t,x)$ analogue for any $1\le i\le N$, we directly conclude that the growth condition is fulfilled by

$$\begin{aligned}\big|\tilde\mu_i(t,x)\big| + \big|\tilde\sigma_i(t,x)\big| &\le \big|\tilde\mu_i(t,x)-\tilde\mu_i(t,0)\big| + \big|\tilde\sigma_i(t,x)-\tilde\sigma_i(t,0)\big| + \big|\tilde\mu_i(t,0)\big| + \big|\tilde\sigma_i(t,0)\big| \\ &\lesssim \|x\|_{N\times d} + \big|\mu(t,0,\delta_0)\big| + \big|\sigma(t,0,\delta_0)\big| \\ &\lesssim \|x\|_{N\times d} + C_{\delta_0} \\ &\lesssim 1+\|x\|_{N\times d},\end{aligned}$$

for all $t\in[0,T]$, $x\in\mathbb{R}^{N\times d}$. Thus, due to the equivalence of all norms on the finite dimensional vector space $\mathbb{R}^{N\times d}$, we can apply the standard Volterra well-posedness result for Lipschitz coefficients from Lemma 5.14 to obtain the system of processes $(X^{N,i})_{i=1,\dots,N}$, which uniquely solves (5.6), for every $N\in\mathbb{N}$.

Step 2: We consider the first summand on the left-hand side of (5.7), i.e. $\mathbb{E}[|X_t^{N,i}-\underline{X}_t^i|^\delta]$. By using Hölder's inequality like in (5.19), Fubini's theorem and the Burkholder–Davis–Gundy inequality such as the Lipschitz continuity of μ and σ, we can bound, for $1\le i\le N$,

$$\begin{aligned}&\mathbb{E}[|X_t^{N,i}-\underline{X}_t^i|^\delta] \\ &\quad= \mathbb{E}\Big[\Big|\int_0^t K_\mu(s,t)\Big(\mu(s,X_s^{N,i},\bar\rho_s^N)-\mu\big(s,\underline{X}_s^i,\mathcal{L}(\underline{X}_s^i)\big)\Big)\,\mathrm{d}s \\ &\qquad\qquad + \int_0^t K_\sigma(s,t)\Big(\sigma(s,X_s^{N,i},\bar\rho_s^N)-\sigma\big(s,\underline{X}_s^i,\mathcal{L}(\underline{X}_s^i)\big)\Big)\,\mathrm{d}B_s^i\Big|^\delta\Big] \\ &\quad\lesssim \Big(\int_0^t \big|K_\mu(s,t)\big|^{\frac{4+2\varepsilon}{4+\varepsilon}}\,\mathrm{d}s\Big)^{\frac{4+\varepsilon}{1+\varepsilon}} \int_0^t \mathbb{E}\Big[\Big|\mu(s,X_s^{N,i},\bar\rho_s^N)-\mu\big(s,\underline{X}_s^i,\mathcal{L}(\underline{X}_s^i)\big)\Big|^\delta\Big]\,\mathrm{d}s \\ &\qquad\qquad + \mathbb{E}\Big[\Big(\int_0^t \big|K_\sigma(s,t)\big(\sigma(s,X_s^{N,i},\bar\rho_s^N)-\sigma\big(s,\underline{X}_s^i,\mathcal{L}(\underline{X}_s^i)\big)\big|^2\,\mathrm{d}s\Big)^{\frac{\delta}{2}}\Big] \\ &\quad\lesssim \int_0^t \mathbb{E}\Big[\Big|\mu(s,X_s^{N,i},\bar\rho_s^N)-\mu\big(s,\underline{X}_s^i,\mathcal{L}(\underline{X}_s^i)\big)\Big|^\delta\Big]\,\mathrm{d}s\end{aligned}$$

$$+\Big(\int_0^t |K_\sigma(s,t)|^{2+\varepsilon}\,\mathrm{d}s\Big)^{\frac{4+2\varepsilon}{\varepsilon(2+\varepsilon)}} \int_0^t \mathbb{E}\Big[\big|\sigma(s,X_s^{N,i},\bar\rho_s^N)-\sigma\big(s,\underline{\mathrm{X}}_s^i,\mathcal{L}(\underline{\mathrm{X}}_s^i)\big)\big|^\delta\Big]\,\mathrm{d}s$$

$$\lesssim \int_0^t \mathbb{E}\Big[|X_s^{N,i}-\underline{\mathrm{X}}_s^i|^\delta + W_\delta(\bar\rho_s^N,\mathcal{L}(\underline{\mathrm{X}}_s^i))^\delta\Big]\,\mathrm{d}s, \tag{5.37}$$

for any $t\in[0,T]$. By Remark 5.20, we obtain that $\mathcal{L}(\underline{\mathrm{X}}_s^i)=\mathcal{L}(\underline{\mathrm{X}}_s^1)$. Hence, we get that

$$\begin{aligned} W_\delta\big(\bar\rho_s^N,\mathcal{L}(\underline{\mathrm{X}}_s^i)\big)^\delta &= W_\delta\Big(\bar\rho_s^N,\mathcal{L}(\underline{\mathrm{X}}_s^1)\Big)^\delta \\ &\le 2^\delta W_\delta\Big(\bar\rho_s^N,\frac{1}{N}\sum_{j=1}^N \delta_{\underline{\mathrm{X}}_s^j}\Big)^\delta + 2^\delta W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{\mathrm{X}}_s^j},\mathcal{L}(\underline{\mathrm{X}}_s^1)\Big)^\delta \\ &\lesssim \frac{1}{N}\sum_{j=1}^N \big|X_s^{N,j}-\underline{\mathrm{X}}_s^j\big|^\delta + W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{\mathrm{X}}_s^j},\mathcal{L}(\underline{\mathrm{X}}_s^1)\Big)^\delta. \end{aligned} \tag{5.38}$$

Moreover, by Remark 5.20, we can find a measurable map $G\colon \mathbb{R}^d\times C([0,T];\mathbb{R}^m)\to C([0,T];\mathbb{R}^d)$ such that, for any $1\le i\le N$,

$$\underline{\mathrm{X}}^i = G(X_0^i,B^i).$$

In the same way, there is a measurable map $G_N\colon \mathbb{R}^{N\times d}\times C([0,T];\mathbb{R}^m)^N\to C([0,T];\mathbb{R}^d)^N$, such that

$$(X^{N,1},\dots,X^{N,N}) = G_N\big((X_0^1,\dots,X_0^N),(B^1,\dots,B^N)\big).$$

More generally, by the symmetry of the system (5.6), for any permutation ς of $\{1,\dots,N\}$, it is

$$(X^{N,\varsigma(1)},\dots,X^{N,\varsigma(N)}) = G_N\big((X_0^{\varsigma(1)},\dots,X_0^{\varsigma(N)}),(B^{\varsigma(1)},\dots,B^{\varsigma(N)})\big).$$

Hence, since the random variables $((X_0^i,B^i))_{1\le i\le N}$ are i.i.d. and, in particular, exchangeable, we can apply Lemma 5.21 to obtain that the coupled processes $\big((X^{N,i},\underline{\mathrm{X}}^i)\big)_{1\le i\le N}$ are exchangeable and hence, in particular, are identically distributed. We can for $i=1$ insert (5.38) into (5.37) and conclude by Jensen's inequality that

$$\begin{aligned} &\mathbb{E}[|X_t^{N,1}-\underline{\mathrm{X}}_t^1|^\delta] \\ &\lesssim \int_0^t \mathbb{E}\Big[|X_s^{N,1}-\underline{\mathrm{X}}_s^1|^\delta + \frac{1}{N}\sum_{j=1}^N \big|X_s^{N,j}-\underline{\mathrm{X}}_s^j\big|^\delta + W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{\mathrm{X}}_s^j},\mathcal{L}(\underline{\mathrm{X}}_s^1)\Big)^\delta\Big]\,\mathrm{d}s \\ &= \int_0^t \mathbb{E}\Big[2|X_s^{N,1}-\underline{\mathrm{X}}_s^1|^\delta + W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{\mathrm{X}}_s^j},\mathcal{L}(\underline{\mathrm{X}}_s^1)\Big)^\delta\Big]\,\mathrm{d}s. \end{aligned}$$

Using Grönwall's lemma, we deduce

$$\mathbb{E}[|X_t^{N,1}-\underline{\mathrm{X}}_t^1|^\delta] \lesssim \int_0^t \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{\mathrm{X}}_s^j},\mathcal{L}(\underline{\mathrm{X}}_s^1)\Big)^\delta\Big]\,\mathrm{d}s,$$

and since the processes $\big((X^{N,i}, \underline{X}^i)\big)_{1 \le i \le N}$ are identically distributed, this completes *Step 2.*

Step 3: First, we show that

$$\lim_{N \to \infty} \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)^\delta\Big] = 0, \tag{5.39}$$

for any $s \in [0,T]$ by showing convergence in probability and uniform integrability. By the Glivenko–Cantelli theorem (see [SW09, Chapter 26, Theorem 1] for a general version) and since the $\underline{X}^j$ are i.i.d., we get the convergence

$$\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j} \to \mathcal{L}(\underline{X}_s^1), \quad \text{as } N \to \infty,$$

almost surely and hence in probability. Furthermore, using again the notation $\bar{\rho}_s^N = \frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}$ we can bound using Hölder's inequality and the boundedness of all moments of $\underline{X}_s^i$, $1 \le i \le N$, in (5.5), that

$$\begin{aligned}
&\sup_{N \in \mathbb{N}} \mathbb{E}\big[W_\delta\big(\bar{\rho}_s^N, \mathcal{L}(\underline{X}_s^1)\big)^\delta \mathbb{1}_{\{W_\delta(\bar{\rho}_s^N, \mathcal{L}(\underline{X}_s^1)) > K\}}\big] \\
&\qquad \le K^{-1} \sup_{N \in \mathbb{N}} \mathbb{E}\big[W_\delta\big(\bar{\rho}_s^N, \mathcal{L}(\underline{X}_s^1)\big)^{\delta+1}\big] \\
&\qquad \le K^{-1} \sup_{N \in \mathbb{N}} \mathbb{E}\big[W_{\delta+1}\big(\bar{\rho}_s^N, \mathcal{L}(\underline{X}_s^1)\big)^{\delta+1}\big] \\
&\qquad \le K^{-1} \sup_{N \in \mathbb{N}} \mathbb{E}\big[W_{\delta+1}\big(\bar{\rho}_s^N, \delta_0\big)^{\delta+1} + W_{\delta+1}\big(\delta_0, \mathcal{L}(\underline{X}_s^1)\big)^{\delta+1}\big] \\
&\qquad = K^{-1} \sup_{N \in \mathbb{N}} \mathbb{E}\Big[\frac{1}{N}\big(\sum_{i=1}^N |\underline{X}_s^i|^{\delta+1}\big) + |\underline{X}_s^1|^{\delta+1}\Big] \\
&\qquad = 2K^{-1}\mathbb{E}[|\underline{X}_s^1|^{\delta+1}] \to 0,
\end{aligned} \tag{5.40}$$

as $K \to \infty$, which shows uniform δ-integrability of the family of random variables

$$\Big(W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)\Big)_{N \in \mathbb{N}}.$$

Hence, Vitali's convergence theorem (see [Bog07, Theorem 4.5.4]) reveals the L^δ-convergence as claimed in (5.39).
To conclude *Step 3*, it remains to show that the convergence (5.39) is uniform in s. There-

fore, we first notice that for any $p \geq \delta$,

$$\begin{aligned}\mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)^p\Big] &\leq \mathbb{E}\Big[W_p\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)^p\Big] \\ &\lesssim \mathbb{E}\Big[W_p\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \delta_0\Big)^p\Big] + W_p\big(\delta_0, \mathcal{L}(\underline{X}_s^1)\big)^p \\ &= \frac{1}{N}\mathbb{E}\Big[\sum_{j=1}^N |\underline{X}_s^j|^p\Big] + \mathbb{E}[|\underline{X}_s^1|^p] = 2\mathbb{E}[|\underline{X}_s^1|^p] < \infty, \end{aligned} \tag{5.41}$$

by (5.5). With Jensen's inequality (5.41) also follows for $1 \leq p < \delta$.
Let $k := \lceil \delta \rceil \geq \delta$ denote the smallest integer greater or equal to δ. Notice that with the same argumentation as in (5.40) by substituting the exponent δ by k and then bounding from above using the $k+1$-Wasserstein distance and again by Vitali's convergence theorem also the L^k-convergence of the δ-Wasserstein distance in (5.39) follows. Once we show that this L^k-convergence is uniform in s, then it will follow that

$$\begin{aligned}&\lim_{N\to\infty} \sup_{s\in[0,T]} \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)^\delta\Big] \\ &\quad\leq \lim_{N\to\infty} \sup_{s\in[0,T]} \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)^k\Big]^{\frac{\delta}{k}} = 0. \end{aligned} \tag{5.42}$$

To prove (5.42), using the factorization

$$a^k - b^k = (a-b)\sum_{r=0}^{k-1} a^{k-1-r} b^r, \tag{5.43}$$

and Hölder's inequality with δ and $q = \frac{4+2\varepsilon}{4+\varepsilon}$ such that $\frac{1}{\delta} + \frac{1}{q} = 1$, we get

$$\begin{aligned}&\Big|\mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \mathcal{L}(\underline{X}_t^1)\Big)^k\Big] - \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)^k\Big]\Big| \\ &\quad= \Big|\mathbb{E}\Big[\Big(W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \mathcal{L}(\underline{X}_t^1)\Big) - W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)\Big) \\ &\qquad\sum_{r=0}^{k-1} W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \mathcal{L}(\underline{X}_t^1)\Big)^{k-1-r} W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)^r\Big]\Big| \\ &\quad\leq \Big|\mathbb{E}\Big[\Big(W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \mathcal{L}(\underline{X}_t^1)\Big) - W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)\Big)^\delta\Big]^{\frac{1}{\delta}} \\ &\qquad\mathbb{E}\Big[\Big(\sum_{r=0}^{k-1} W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \mathcal{L}(\underline{X}_t^1)\Big)^{k-1-r} W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)^r\Big)^q\Big]^{\frac{1}{q}}\Big|. \end{aligned} \tag{5.44}$$

Using again Hölder's inequality such as (5.41), we can bound the second expectation by

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{r=0}^{k-1} W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \mathcal{L}(\underline{X}_t^1)\Big)^{k-1-r} W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)^r\Big)^q\Big]^{\frac{1}{q}} \\
&\lesssim \Big(\sum_{r=0}^{k-1} \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \mathcal{L}(\underline{X}_t^1)\Big)^{q(k-1-r)} W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)^{qr}\Big]\Big)^{\frac{1}{q}} \\
&\lesssim \Big(\sum_{r=0}^{k-1} \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \mathcal{L}(\underline{X}_t^1)\Big)^{2q(k-1-r)}\Big]^{\frac{1}{2}} \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)^{2qr}\Big]^{\frac{1}{2}}\Big)^{\frac{1}{q}} < \infty,
\end{aligned} \tag{5.45}$$

such that inserting (5.45) into (5.44) and using the triangle inequality

$$\begin{aligned}
&W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \mathcal{L}(\underline{X}_t^1)\Big) \\
&\quad \le W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}\Big) + W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big) + W_\delta\Big(\mathcal{L}(\underline{X}_s^1), \mathcal{L}(\underline{X}_t^1)\Big),
\end{aligned}$$

which also holds if we switch s and t, we continue with

$$\begin{aligned}
&\Big|\mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \mathcal{L}(\underline{X}_t^1)\Big)^k\Big] - \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)^k\Big]\Big| \\
&\quad \lesssim \mathbb{E}\Big[\Big|\Big(W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \mathcal{L}(\underline{X}_t^1)\Big) - W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1)\Big)\Big)\Big|^\delta\Big]^{\frac{1}{\delta}} \\
&\quad \lesssim \mathbb{E}\Big[\Big(W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}\Big) + W_\delta\Big(\mathcal{L}(\underline{X}_t^1), \mathcal{L}(\underline{X}_s^1)\Big)\Big)^\delta\Big]^{\frac{1}{\delta}} \\
&\quad \lesssim \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_t^j}, \frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}\Big)^\delta\Big]^{\frac{1}{\delta}} + \mathbb{E}\Big[W_\delta\Big(\mathcal{L}(\underline{X}_t^1), \mathcal{L}(\underline{X}_s^1)\Big)^\delta\Big]^{\frac{1}{\delta}} \\
&\quad \lesssim \mathbb{E}[|\underline{X}_t^1 - \underline{X}_s^1|^\delta]^{\frac{1}{\delta}} \\
&\quad \lesssim |t-s|^\beta,
\end{aligned}$$

where the last line holds by Remark 5.18 for any $\beta \in (0, \gamma - 1/p)$ with $\gamma \in (0, \frac{1}{2}]$ from Assumption 5.1. Hence, we obtain that (5.42) holds which shows together with (5.36) that

$$\lim_{N\to\infty} \sup_{t\in[0,T]} \mathbb{E}[|X_t^{N,1} - \underline{X}_t^1|^\delta] = 0, \tag{5.46}$$

and knowing that $((X^{N,i}, \underline{X}^i))_{1\le i\le N}$ are identically distributed, this completes *Step 3.*

Step 4: We already know from *Step 3* that the first summand in (5.7) converges to zero. For the second summand, we use the triangle inequality and Jensen's inequality, to obtain

$$
\begin{aligned}
&\sup_{t\in[0,T]} \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^{N,i}}, \mathcal{L}(\underline{\mathrm{X}}_t^1)\Big)^\delta\Big] \\
&\quad\lesssim \sup_{t\in[0,T]} \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^{N,i}}, \frac{1}{N}\sum_{i=1}^{N}\delta_{\underline{\mathrm{X}}_t^i}\Big)^\delta\Big] + \sup_{t\in[0,T]} \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{i=1}^{N}\delta_{\underline{\mathrm{X}}_t^i}, \mathcal{L}(\underline{\mathrm{X}}_t^1)\Big)^\delta\Big] \\
&\quad\lesssim \sup_{t\in[0,T]} \mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^{N}|X_t^{N,i}-\underline{\mathrm{X}}_t^i|^\delta\Big] + \sup_{t\in[0,T]} \mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{i=1}^{N}\delta_{\underline{\mathrm{X}}_t^i}, \mathcal{L}(\underline{\mathrm{X}}_t^1)\Big)^\delta\Big],
\end{aligned} \tag{5.47}
$$

which also tends to 0 as $N \to \infty$, by (5.42) and (5.46). □

We continue with the proof of Theorem 5.11. Since the proof is similar to the proof of Theorem 5.11, we focus, for the sake of brevity, on the main differences.

Proof of Theorem 5.11. We prove the statement by using the same *Step 1-Step 4* as in the proof of Theorem 5.4, but with $\delta = 1$. Only for *Step 2*, we need to differ between Assumption 5.7, referred to as *Case (i)*, and Assumption 5.8, referred to as *Case (ii)*, to hold.

Step 1: By Remark 5.17, we obtain as in the proof of Theorem 5.4 the unique system of stochastic processes $(X^{N,i})_{i=1,\dots,N}$ that solves (5.6).

Step 2:
Case (i): Suppose the kernels K_μ, K_σ and initial condition X_0 satisfy Assumption 5.7. To mimic inequality (5.36), we use the semimartingale property

$$
\begin{aligned}
X_t^{N,i} - \underline{\mathrm{X}}_t^i = &\int_0^t K_\sigma(s,s)\Big(\sigma(s,X_s^{N,i}) - \sigma(s,\underline{\mathrm{X}}_s^i)\Big)\,\mathrm{d}B_s \\
&+ \int_0^t K_\mu(s,s)\Big(\mu(s,X_s^{N,i},\bar{\rho}_s^N) - \mu(s,\underline{\mathrm{X}}_s^i,\mathcal{L}(\underline{\mathrm{X}}_s^i))\Big)\,\mathrm{d}s \\
&+ \int_0^t \Big(\int_0^s \partial_2 K_\mu(u,s)\Big(\mu(u,X_u^{N,i},\bar{\rho}_u^N) - \mu(u,\underline{\mathrm{X}}_u^i,\mathcal{L}(\underline{\mathrm{X}}_u^i))\Big)\,\mathrm{d}u \\
&+ \int_0^s \partial_2 K_\sigma(u,s)\Big(\sigma(u,X_u^{N,i}) - \sigma(u,\underline{\mathrm{X}}_u^i)\Big)\,\mathrm{d}B_u\Big)\,\mathrm{d}s,
\end{aligned}
$$

to perform a Yamada–Watanabe approach exactly as we did around equality (5.22), and obtain for fixed $i \in \{1,\dots,N\}$ with the notation $M^{N,i}(t) := \mathbb{E}[|X_t^{N,i} - \underline{\mathrm{X}}_t^i|] + \mathbb{E}[|\tilde{Y}_t|]$, where $\tilde{Y}_t := \int_0^t \sigma(s,X_s^{N,i})\,\mathrm{d}B_s^i - \int_0^t \sigma(s,\underline{\mathrm{X}}_s^n)\,\mathrm{d}B_s^i$, that

$$
M^{N,i}(t) \lesssim \int_0^t \Big(M^{N,i}(s) + \mathbb{E}[W_1(\bar{\rho}_s^N, \mathcal{L}(\underline{\mathrm{X}}_s^i))]\Big)\,\mathrm{d}s,
$$

such that, proceeding as in the proof of Theorem 5.4 including applying Grönwall's inequality, we obtain

$$\mathbb{E}[|X_t^{N,i} - \underline{X}_t^i|] \lesssim \int_0^t \mathbb{E}\big[W_1(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^i))\big] \,\mathrm{d}s. \tag{5.48}$$

Case (ii): Suppose the kernels K_μ, K_σ and initial condition X_0 satisfy Assumption 5.8. As in *Case (i)* to mimic inequality (5.36), we use the semimartingale property

$$\begin{aligned}
X_t^{N,i} - \underline{X}_t^i = &\int_0^t \tilde{K}(0)\Big(\sigma(s, X_s^{N,i}) - \sigma(s, \underline{X}_s^i)\Big) \,\mathrm{d}B_s \\
&+ \int_0^t \tilde{K}(0)\Big(\mu(s, X_s^{N,i}, \bar{\rho}_s^N) - \mu(s, \underline{X}_s^i, \mathcal{L}(\underline{X}_s^i))\Big) \,\mathrm{d}s \\
&+ \int_0^t \bigg(\int_0^s \tilde{K}'(s-u)\Big(\mu(u, X_u^{N,i}, \bar{\rho}_u^N) - \mu(u, \underline{X}_u^i, \mathcal{L}(\underline{X}_u^i))\Big) \,\mathrm{d}u \\
&+ \int_0^s \tilde{K}'(s-u)\Big(\sigma(u, X_u^{N,i}) - \sigma(u, \underline{X}_u^i)\Big) \,\mathrm{d}B_u\bigg) \,\mathrm{d}s,
\end{aligned}$$

perform a Yamada–Watanabe approach and apply Grönwall's inequality as in (5.31) which yields

$$\mathbb{E}[|X_t^{N,i} - \underline{X}_t^i|] \lesssim \int_0^t \mathbb{E}[W_1(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^i))]\Big) \,\mathrm{d}s.$$

Step 3: Obtaining the convergence to zero uniformly in s of the right-hand side of (5.48) follows now easily by using

$$\mathbb{E}\big[W_1(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1))\big] \le \mathbb{E}\big[W_2(\frac{1}{N}\sum_{j=1}^N \delta_{\underline{X}_s^j}, \mathcal{L}(\underline{X}_s^1))^2\big]^{\frac{1}{2}},$$

and then using [CD18a, (5.19)], and proceeding as in [CD18b, Proof of Theorem 2.12].

Step 4: As in (5.47), we obtain

$$\begin{aligned}
&\sup_{0\le t\le T} \mathbb{E}\Big[W_1\Big(\frac{1}{N}\sum_{i=1}^N \delta_{X_t^{N,i}}, \mathcal{L}(\underline{X}_t^1)\Big)\Big] \\
&\quad\lesssim \sup_{0\le t\le T} \mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^N |X_t^{N,i} - \underline{X}_t^i|\Big] + \sup_{0\le t\le T} \mathbb{E}\Big[W_1\Big(\frac{1}{N}\sum_{i=1}^N \delta_{\underline{X}_t^i}, \mathcal{L}(\underline{X}_t^1)\Big)\Big],
\end{aligned} \tag{5.49}$$

which tends to zero by the uniform convergence to zero of the right-hand side of (5.48), and finishes the proof. □

5.5 Rate of convergence: Proof of Lemma 5.5 and 5.12

The proofs of Lemma 5.5 and Lemma 5.12 rely on a quantitative Glivenko–Cantelli theorem due to Fournier and Guillin [FG15], which provides a sharp estimate of the δ-Wasserstein distance. For the sake of completeness, we recall [FG15, Theorem 1] in the following lemma.

Lemma 5.22. *Let $\bar{\rho}^N := \frac{1}{N}\sum_{i=1}^{N}\delta_{X^i}$ be the empirical distribution of i.i.d. random variables $(X^i)_{i=1,\dots,N}$ with common distribution ρ such that $\rho \in \mathcal{P}_p(\mathbb{R}^d)$ for every $p \geq 1$. Then, we have*

$$\mathbb{E}[W_\delta(\bar{\rho}^N, \rho)^\delta] \lesssim \varepsilon_N,$$

where $(\varepsilon_N)_{N\in\mathbb{N}}$ is given by (5.9), *i.e.*

$$\varepsilon_N = \begin{cases} N^{-1/2}, & \text{if } d < 2\delta, \\ N^{-1/2}\log_2(1+N), & \text{if } d = 2\delta, \\ N^{-\delta/d}, & \text{if } d > 2\delta, \end{cases}$$

and

$$\mathbb{E}[W_1(\bar{\rho}^N, \rho)] \lesssim N^{-1/2}.$$

With this lemma at hand, we can prove Lemma 5.5 and 5.12.

Proof of Lemma 5.5. By Lemma 5.22 we obtain that for any $t \in [0,T]$,

$$\mathbb{E}\Big[W_\delta\Big(\frac{1}{N}\sum_{i=1}^{N}\delta_{\underline{X}_t^i}, \mathcal{L}(\underline{X}_t^1)\Big)^\delta\Big] \lesssim \varepsilon_N, \tag{5.50}$$

where $(\varepsilon_N)_{N\in\mathbb{N}}$ is given by (5.9) and the right-hand side does not depend on t. Plugging (5.50) into (5.36) and taking the supremum over $[0,T]$ and maximum over $1,\dots,N$ shows the desired convergence rate of the first term in (5.8). Then, using this and plugging (5.50) into (5.47) gives the desired rate for the second term. □

Proof of Lemma 5.12. Case (i): Suppose the kernels K_μ, K_σ and initial condition X_0 satisfy Assumption 5.7. By Lemma 5.22 we obtain that

$$\mathbb{E}\Big[W_1\Big(\frac{1}{N}\sum_{i=1}^{N}\delta_{\underline{X}_t^i}, \mathcal{L}(\underline{X}_t^1)\Big)\Big] \lesssim N^{-1/2}, \tag{5.51}$$

independently from $t \in [0,T]$. Plugging (5.51) into (5.48) and (5.49) yields the statement.

Case (ii): Suppose the kernels K_μ, K_σ and initial condition X_0 satisfy Assumption 5.8. Plugging (5.51) into the analogues of (5.48) and (5.49) yields the statement. □

Chapter 6

Neural SVEs

The content of this chapter is unpublished so far.

Introduction

We introduce neural stochastic Volterra equations (neural SVEs) and consider the supervised learning problem (see [Wat20, Section 1.3.1]) for solutions to SVEs. By supervised learning of a stochastic equation we mean the setup that we have a training set consisting of paths of the true process and of the underlying driving stochastic noise and stochastic initial value, and build a model that tries to reproduce the paths, given the noise path and initial value, as best as possible. For a supervised learning problem in the case of stochastic partial differential equations, see e.g. [SLG22].
SVEs offer a wide range of applications, e.g. in finance or biology, see e.g. [AJEE19b], [EER19], [MS15], [AJCLP21], and are a natural generalization of ordinary stochastic differential equations (SDEs). For SDEs, the concept of neural SDEs was introduced in [Kid22] where the unsupervised problem was considered, i.e. the setting where the model receives only the paths of the process as training input without any knowledge of the underlying noise, and tries to reproduce paths that are as similar as possible to those original paths in some sense in a generative way. In an unsupervised learning setting, the model can be seen as a distribution over paths, where some randomness must come into play to generate paths, while in the supervised setting the randomness must already happen before and given as input into our model which then outputs the path of the belonging process in a deterministic way.

6.1 Background on stochastic Volterra equations

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space, which satisfies the usual conditions, $(B_t)_{t\in[0,T]}$ be a standard Brownian motion and $T \in (0,\infty)$. We consider for $d, m \in \mathbb{N}$ the d-dimensional stochastic Volterra equation (SVE) of convolution type driven by an m-

dimensional Brownian motion,

$$X_t = \xi\, g(t) + \int_0^t K_\mu(t-s)\mu(s,X_s)\,\mathrm{d}s + \int_0^t K_\sigma(t-s)\sigma(s,X_s)\,\mathrm{d}B_s, \quad t \in [0,T], \tag{6.1}$$

where $\xi \in \mathbb{R}^d$ is the (stochastic) initial value, $g\colon [0,T] \to \mathbb{R}$ is a (deterministic) continuous function (where we usually scale $g(0) = 1$), the coefficients $\mu\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$ and the kernels of convolutional type $K_\mu, K_\sigma\colon [0,T] \to \mathbb{R}$ are measurable functions, and $(B_t)_{t\in[0,T]}$ is an m-dimensional standard Brownian motion. Typically, we choose $m = d$. Furthermore, $\int_0^t K_\sigma(t-s)\sigma(s,X_s)\,\mathrm{d}B_s$ is defined as an Itô integral.

We aim to learn strong (L^p-)solutions of the stochastic Volterra equation (6.1) given the Brownian path. To define the notion of a strong (L^p-)solution, let $L^p(\Omega \times [0,T])$ be the space of all real-valued, p-integrable functions on $\Omega \times [0,T]$. We call an $(\mathcal{F}_t)_{t\in[0,T]}$-progressively measurable stochastic process $(X_t)_{t\in[0,T]}$ in $L^p(\Omega\times[0,T])$ on the given probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, a *(strong) L^p-solution* of the SVE (6.1) if $\int_0^t (|K_\mu(t-s)\mu(s,X_s)| + |K_\sigma(t-s)\sigma(s,X_s)|^2)\,\mathrm{d}s < \infty$ for all $t \in [0,T]$ and the integral equation (6.1) hold $\mathbb{P}$-almost surely. As usual, a strong L^1-solution $(X_t)_{t\in[0,T]}$ of the SVE (6.1) is often just called *solution* of the SVE (6.1).

6.2 Neural SVEs

We want to learn the dynamics of the SVE (6.1), i.e. of the operators ξ, g, K_μ, K_σ, μ and σ, by some neural network structure. Let therefore for some latent dimension $d_h > d$,

$$\begin{gathered} L_\theta\colon \mathbb{R}^d \to \mathbb{R}^{d_h}, \quad g_\theta\colon [0,T] \to \mathbb{R}, \quad K_{\mu,\theta}\colon [0,T] \to \mathbb{R}, \quad K_{\sigma,\theta}\colon [0,T] \to \mathbb{R}, \\ \mu_\theta\colon [0,T]\times\mathbb{R}^{d_h} \to \mathbb{R}^{d_h}, \quad \sigma_\theta\colon [0,T]\times\mathbb{R}^{d_h} \to \mathbb{R}^{d_h\times m}, \quad \Pi_\theta\colon \mathbb{R}^{d_h} \to \mathbb{R}^d \end{gathered}$$

be six feedforward neural networks (see [YYK15, Section 3.6.1]) that are parameterized by some common parameter θ. Note that L_θ lifts the given initial value to the latent space $\mathbb{R}^{d_h}$, Π_θ is the readout back from the latent space to the process space $\mathbb{R}^d$, and the other networks try to imitate their respectives in equation (6.1), on the latent d_h-dimensional space.

Given input data $\xi \in \mathbb{R}^d$ and $(B_t)_{t\in[0,T]} \in C([0,T];\mathbb{R}^m)$, a neural SVE is defined as

$$\begin{aligned} Z_0 &= L_\theta(\xi), \\ Z_t &= Z_0\, g_\theta(t) + \int_0^t K_{\mu,\theta}(t-s)\mu_\theta(s,Z_s)\,\mathrm{d}s + \int_0^t K_{\sigma,\theta}(t-s)\sigma_\theta(s,Z_s)\,\mathrm{d}B_s, \qquad (6.2) \\ X_t &= \Pi_\theta(Z_t), \quad t \in [0,T]. \end{aligned}$$

The objective is to train θ as best as possible such that the generated paths are as close as possible to the true training paths.
Given a trained supervised model $(L_\theta, K_{\mu,\theta}, K_{\sigma,\theta}, \mu_\theta, \sigma_\theta, \Pi_\theta)$, we can evaluate the neural SVE on some data (ξ, B) by solving equation (6.2) using any numerical scheme to solve a stochastic Volterra equation. We therefore use the Volterra Euler–Maruyama scheme introduced in [Zha08] for the training procedure. Note that Lipschitz conditions on μ_θ and σ_θ can be imposed by using e.g. LipSwish, ReLU or tanh activation functions.

6.2.1 Neural architecture

The structure of the neural SVE model (6.2) is analogue to the one of neural SDEs introduced in [Kid22] and to the one of neural SPDEs in [SLG22]. The d_h-dimensional process Z represents the hidden state. We impose the readout Π_θ to get back to dimension d. The model has, at least if one considers a setting where the initial condition cannot be observed like an unsupervised setting, some minimal amount of architecture. It is in such a setting necessary to induce the lift L_θ and the randomness by some additional variable $\tilde{\xi}$ to learn the randomness induced by the initial condition $X_0 = \Pi_\theta\big(L_\theta(\tilde{\xi})g_\theta(0)\big)$ (otherwise X_0 would not be random since it does not depend on B). Moreover, the structure induced by the lift L_θ and the readout Π_θ is the natural choice to lift the d-dimensional SVE (6.1) to the latent dimension $d_h > d$.

We use LipSwish activation functions in any layer of any network. These were introduced in [CBDJ19] as $\rho(z) = 0.909z\sigma(z)$, where σ is the sigmoid function. Due to the constant 0.909, LipSwish activations are Lipschitz continuous with Lipschitz constant one and smooth. Moreover, they have shown strong empirical evidence in a variety of challenging approximation tasks, see [RZL17].

For a given latent dimension $d_h > d$, the lift L_θ is modeled as a linear 1-layer network from dimension d to d_h without any additional hidden layer, and as its counterpart the readout Π_θ as a linear 1-layer network from d_h to d. The networks $K_{\mu,\theta}, K_{\sigma,\theta}$ and g_θ are all designed as linear networks from dimension 1 to 1 with two hidden layers of size d_K for some additional dimension $d_K > d$. Lastly, the network μ_θ is defined as a linear network from dimension $1 + d_h$ to d_h with one hidden layer of size d_h and the network σ_θ from $1 + d_h$ to $d_h \cdot m$ with one hidden layer of size $d_h \cdot m$.

6.3 Examples

We introduce the disturbed pendulum equation, the generalized Ornstein–Uhlenbeck process and the rough Heston model, which can all be modeled by SVEs and on which we perform the neural model (6.2). The experimental results are then presented in the following section.

6.3.1 Disturbed second-order differential systems

For the first example, we consider general deterministic second-order differential systems (without first-order terms) that are disturbed by some multiplicative noise. We derive in the following that the resulting stochastic process is a Volterra process with the smooth kernels $K_\mu(t-s) = K_\sigma(t-s) = (t-s)$.
Consider the second-order system

$$y''(t) = \mu(t, y(t)), \quad t \in [0, T]$$

that is disturbed by some multiplicative noise such that

$$y''(t) = \mu(t, y(t)) + \sigma(t, y(t))\dot{B}_t, \quad t \in [0, T],$$

where $\dot{B}_t = \frac{dB_t}{dt}$ is White noise for some standard Brownian motion $(B_t)_{t\in[0,T]}$. We can write

$$\begin{aligned} y'(t) &= y'(0) + \int_0^t y''(s)\,\mathrm{d}s \\ &= y'(0) + \int_0^t \mu(s, y(s))\,\mathrm{d}s + \int_0^t \sigma(s, y(s))\,\mathrm{d}B_s, \end{aligned}$$

hence the first derivative is the solution of an SDE. Further, by the deterministic and the stochastic Fubini theorem,

$$\begin{aligned} y(t) &= y(0) + \int_0^t y'(s)\,\mathrm{d}s \\ &= y(0) + \int_0^t \Big(y'(0) + \int_0^s \mu(u, y(u))\,\mathrm{d}u + \int_0^s \sigma(u, y(u))\,\mathrm{d}B_u\Big)\,\mathrm{d}s \\ &= y(0) + t\cdot y'(0) + \int_0^t\int_u^t \mu(u, y(u))\,\mathrm{d}s\,\mathrm{d}u + \int_0^t\int_u^t \sigma(u, y(u))\,\mathrm{d}s\,\mathrm{d}B_u \\ &= y(0) + t\cdot y'(0) + \int_0^t (t-u)\mu(u, y(u))\,\mathrm{d}u + \int_0^t (t-u)\sigma(u, y(u))\,\mathrm{d}B_u. \end{aligned}$$

Hence, every solution to a multiplicatively disturbed second-order differential equation without first-order terms is the solution to a stochastic Volterra equation with smooth kernels $K_\mu(t-s) = K_\sigma(t-s) = (t-s)$.

A concrete example from physics is the disturbed pendulum equation (see [Ok03, Exercise 5.12.]) resulting from Newton's second law. The motion of an object X with deterministic initial value x_0 under some force F can be described by the differential equation

$$m\frac{\mathrm{d}^2X(t)}{\mathrm{d}t^2} = F(X(t)), \quad X(0) = x_0, \tag{6.3}$$

which follows by Newton's second law (see e.g. [Kre99, Section 2.4]). With (small) random perturbations in its environment we consider for some $\varepsilon > 0$ the disturbed equation

$$m\frac{\mathrm{d}^2X(t)}{\mathrm{d}t^2} = F(X(t)) + \varepsilon X(t)W_t, \quad X(0) = x_0, \tag{6.4}$$

where $W_t = \frac{\mathrm{d}B_t}{\mathrm{d}t}$. Then, X solves the SVE

$$X(t) = x_0 + tX'(0) + \int_0^t (t-s)\frac{F(X(s))}{m}\,\mathrm{d}s + \int_0^t (t-s)\frac{\varepsilon X_s}{m}\,\mathrm{d}B_s. \tag{6.5}$$

6.3.2 Generalized Ornstein–Uhlenbeck process

The Ornstein–Uhlenbeck process, introduced in [UO30], is a commonly used stochastic process with applications finance, physics or biology (see e.g. [Vas12], [TE99],[Mar94]). We consider the generalized Ornstein–Uhlenbeck process that is given by the stochastic differential equation

$$\mathrm{d}X_t = \theta(\mu(t, X_t) - X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t, \quad t \in [0, T]. \tag{6.6}$$

We derive in the following that we can rewrite equation (6.6) as an SVE. Itô's formula yields

$$\begin{aligned} X_t e^{\theta t} &= X_0 + \int_0^t e^{\theta s}\,\mathrm{d}X_s + \int_0^t \theta e^{\theta s} X_s\,\mathrm{d}s \\ &= X_0 + \int_0^t e^{\theta s}\Big(\theta\big(\mu(s, X_s) - X_s\big)\Big)\,\mathrm{d}s + \int_0^t e^{\theta s}\sigma(s, X_s)\,\mathrm{d}B_s + \int_0^t \theta e^{\theta s} X_s\,\mathrm{d}s \\ &= X_0 + \theta\int_0^t e^{\theta s}\mu(s, X_s)\,\mathrm{d}s + \int_0^t e^{\theta s}\sigma(s, X_s)\,\mathrm{d}B_s. \end{aligned}$$

Hence, the generalized Ornstein–Uhlenbeck process is given by the SVE

$$X_t = X_0 e^{-\theta t} + \theta\int_0^t e^{-\theta(t-s)}\mu(s, X_s)\,\mathrm{d}s + \int_0^t e^{-\theta(t-s)}\sigma(s, X_s)\,\mathrm{d}B_s, \quad t \in [0, T]. \tag{6.7}$$

6.3.3 Rough Heston model

The rough Heston model is one of the most prominent representatives of rough volatility models (see [EER19],[AJEE19b]), where the volatility process is modeled using the singular kernels $K_\mu(t-s) = K_\sigma(t-s) = (t-s)^{-\alpha}$ for some $\alpha \in (0, 1/2)$, by the rough SVE

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{-\alpha}\lambda(\theta - V_s)\,\mathrm{d}s + \frac{\lambda\nu}{\Gamma(\alpha)}\int_0^t (t-s)^{-\alpha}\sqrt{|V_s|}\,\mathrm{d}B_s, \quad t \in [0, T], \tag{6.8}$$

where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,\mathrm{d}t$ denotes the real valued Gamma function, and $\lambda, \theta, \nu \in \mathbb{R}$.

6.4 Numerical results

For all the neural SVEs, we chose the latent dimensions $d_h = d_K = 12$ which experimentally proved to be well suited. We consider the interval $[0, T]$ for $T = 5$ and discretize it equally-sized using the grid size $\Delta t = 0.1$.

As a benchmark model, we use the Deep Operator Network (DeepONet) algorithm. DeepONet is a popular class of neural learning algorithms for general operators on function spaces that was introduced in [LJP$^+$21]. A DeepONet consists of two neural networks: the branch network which operates on the function space $C([0, T])$ (where $[0, T]$ is represented by some fixed discretization), and the so-called trunk network which operates on the evaluation point $t \in [0, T]$. Then, the output of the DeepONet is defined as

$$\mathrm{DeepONet}(f)(t) = \sum_{k=1}^{p} b_k t_k + b_0,$$

where $(b_k)_{k=1,\ldots,p}$ is the output of the branch network operating on the discretization of $f \in C([0, T])$, $(t_k)_{k=1,\ldots,p}$ is the output of the trunk network operating on $t \in [0, T]$ and $p \in \mathbb{N}$ is the dimension of the output of both networks. Following [LJP$^+$21], we model both networks as feedforward networks. We perform a grid search to optimally determine the depth and width of both networks such as the activation functions, optimizer and learning rate.

Note that one big advantage of Neural SVEs is that they are discretization invariant, i.e. that in the training procedure and also when evaluating a trained model, it does not matter if the input functions (realizations of the Brownian motion) are discretized by the same grid. In contrast, the DeepONet needs the same discretization for all functions in the training and in the evaluation data.

We perform experiments on a one-dimensional disturbed pendulum equation, a one- and a two-dimensional Ornstein–Uhlenbeck equation such as a one-dimensional rough Heston equation. We perform the experiments on low-, mid- and high-data regimes with $n = 100$, $n = 500$ and $n = 2000$, and use 80% of the data for training and 20% for testing. We compare the results of Neural SVEs to those of DeepONet and consider for both algorithms the mean relative L^2-loss. All experiments are trained for an appropriate number *epochs* of iterations until there is no improvement anymore. For Neural SVEs, we use the learning rate 0.01 and scale the learning rate by a factor 0.8 after every 25% of *epochs*.
Note that since DeepONet is not able to deal with random initial conditions, we use deterministic initial conditions $\xi = 2$ in the DeepONet experiments. For Neural SVEs, we use initial conditions $\xi \sim \mathcal{N}(2, 0.2)$.

Remark 6.1. *The results in this section show that Neural SVEs are able to outperform DeepONet significantly (see Table 6.1-Table 6.8). Especially, Neural SVEs generalize much better which can be seen in the good performance on the test sets where Neural SVEs are up to* 20 *times better than DeepONet. This can be explained by the explicit structure of the Volterra equation that is already part of the model for Neural SVEs.*

Neural SVE	Train set	Test set
$n = 100$	0.01	0.013
$n = 500$	0.008	0.008
$n = 2000$	0.006	0.006

DeepONet	Train set	Test set
$n = 100$	0.003	0.2
$n = 500$	0.003	0.06
$n = 2000$	0.003	0.02

Table 6.1: Mean relative L^2-losses after training for the disturbed pendulum equation (6.9).

Another advantage is that Neural SVEs are time-resolution invariant, meaning that they can be trained and evaluated on arbitrary, possibly different time grid discretizations which is not possible for DeepONet.

All the code is published in `https://github.com/davidscheffels/Neural_SVEs`.

Disturbed pendulum equation

We consider the one-dimensional equation

$$y_t = \xi + \int_0^t (t-s) y_s \, \mathrm{d}s + \int_0^t (t-s) y_s \, \mathrm{d}B_s, \qquad t \in [0, T], \tag{6.9}$$

and learn solutions to it by Neural SVEs and DeepONet. The results are presented in Table 6.1.

Example paths of the training and the testing sets together with their learned approximations are shown in Table 6.2. It is clearly visible that while DeepONet is not able to generalize properly to the testing set, the learned Neural SVE paths are very close to the true paths also for the test set.

Ornstein–Uhlenbeck process

We consider the one-dimensional equation

$$X_t = \xi e^{-t} + \int_0^t e^{-(t-s)} X_s \, \mathrm{d}s + \int_0^t e^{-(t-s)} \sqrt{|X_s|} \, \mathrm{d}B_s, \qquad t \in [0, T], \tag{6.10}$$

and learn solutions to it by Neural SVEs and DeepONet. The results are presented in Table 6.3.

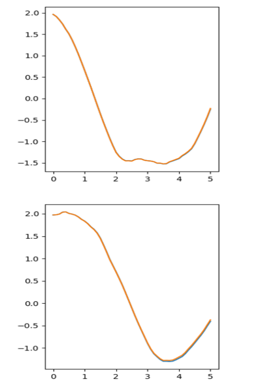
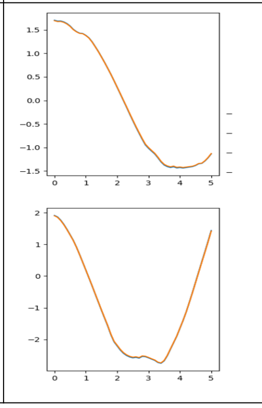
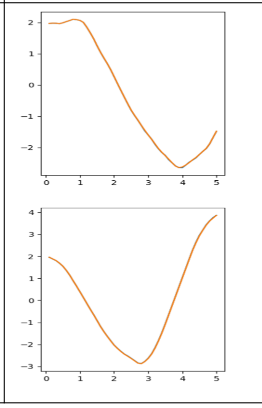
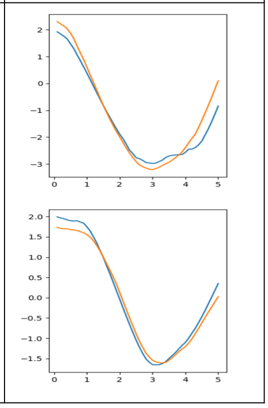
Neural SVE: Training set	Neural SVE: Test set	DeepONet: Training set	DeepONet: Test set
			

Table 6.2: Sample Neural SVE and DeepONet paths from the training and the test set for the disturbed pendulum equation and $n = 100$. Blue (barely visible) are the original paths and orange the learned approximations.

Neural SVE	Train set	Test set
$n = 100$	0.015	0.038
$n = 500$	0.014	0.036
$n = 2000$	0.014	0.02

DeepONet	Train set	Test set
$n = 100$	0.025	0.23
$n = 500$	0.018	0.15
$n = 2000$	0.028	0.12

Table 6.3: Relative L^2-losses after training for the one-dimensional Ornstein–Uhlenbeck equation (6.10).

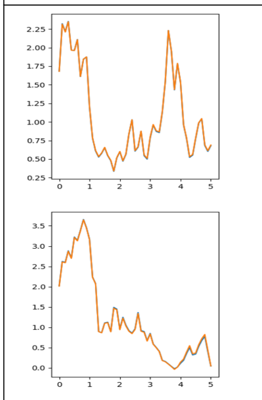
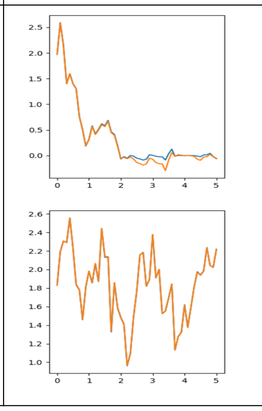
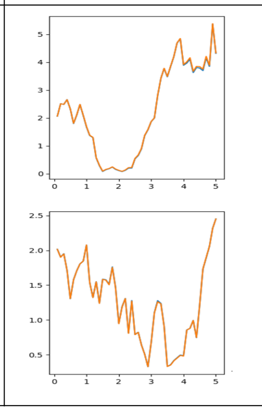
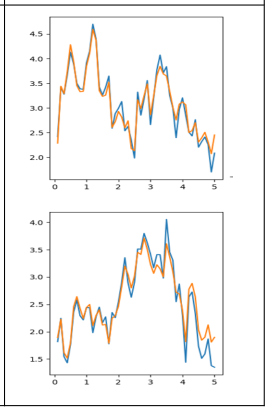
Neural SVE: Training set	Neural SVE: Test set	DeepONet: Training set	DeepONet: Test set
			

Table 6.4: Sample Neural SVE and DeepONet paths from the training and the test set for the one-dimensional Ornstein–Uhlenbeck equation and $n = 500$. Blue (barely visible) are the original paths and orange the learned approximations.

Neural SVE	Train set	Test set
$n = 100$	0.038	0.095
$n = 500$	0.04	0.085
$n = 2000$	0.038	0.04

Table 6.5: Relative L^2-losses after training for the two-dimensional Ornstein–Uhlenbeck equation (6.11).

Example paths of the training and the testing sets together with their learned approximations are shown in Table 6.4.

Moreover, Neural SVEs are also able to learn multi-dimensional SVEs. As an example, we consider the two-dimensional equation

$$\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} e^{-t} + \int_0^t e^{-(t-s)} \begin{pmatrix} X_s^1 \\ X_s^2 \end{pmatrix} \mathrm{d}s + \int_0^t e^{-(t-s)} \begin{pmatrix} \sqrt{|X_s^1|}, 0 \\ 0, \sqrt{|X_s^2|} \end{pmatrix} \mathrm{d}B_s, \qquad t \in [0, T], \tag{6.11}$$

where B is a 2-dimensional Brownian motion, and learn solutions to it by Neural SVEs. The results are presented in Table 6.5.

Example paths of the training and the testing sets together with their learned approximations are shown in Table 6.6.

Rough Heston equation

We consider the one-dimensional equation

$$V_t = \xi + \frac{1}{\Gamma(0.4)} \int_0^t (t-s)^{-0.4}(2 - V_s)\,\mathrm{d}s + \frac{1}{\Gamma(0.4)} \int_0^t (t-s)^{-0.4} \sqrt{|V_s|}\,\mathrm{d}B_s, \qquad t \in [0, T], \tag{6.12}$$

and learn solutions to it by Neural SVEs and DeepONet. The results are presented in Table 6.7. Neural SVEs outperform DeepONet here by far.

Example paths of the training and the testing sets together with their learned approximations are shown in Table 6.8.

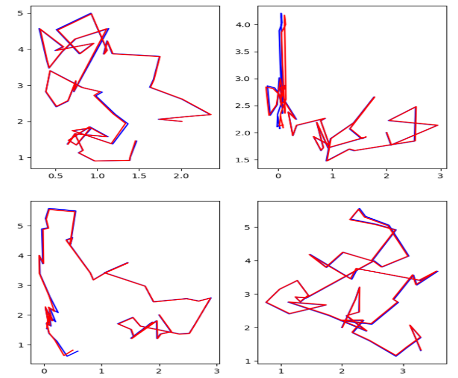
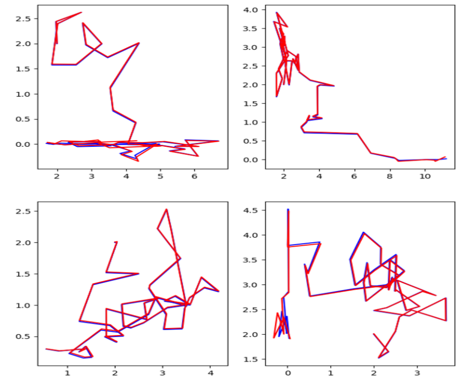
Neural SVE: Training set	Neural SVE: Test set
	

Table 6.6: Sample Neural SVE paths from the training and the test set for the two-dimensional Ornstein–Uhlenbeck equation and $n = 2000$. Blue (barely visible) are the original paths and red the learned approximations.

Neural SVE	Train set	Test set
$n = 100$	0.003	0.003
$n = 500$	0.0025	0.0028
$n = 2000$	0.0015	0.0017

DeepONet	Train set	Test set
$n = 100$	0.035	0.13
$n = 500$	0.004	0.037
$n = 2000$	0.003	0.014

Table 6.7: Relative L^2-losses after training for the rough Heston equation (6.12).

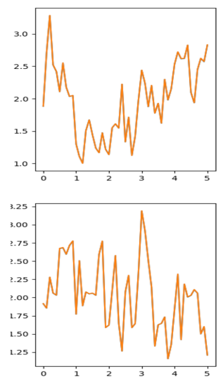
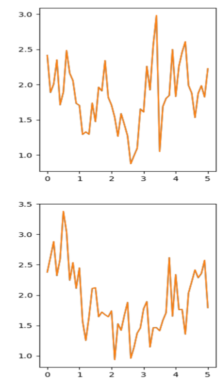
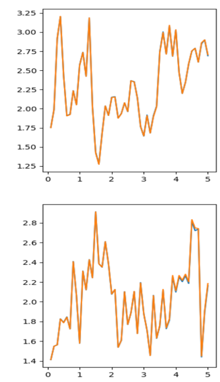
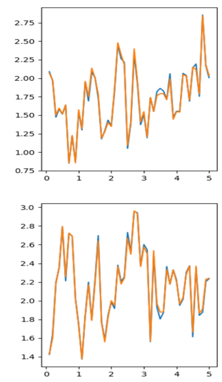
Neural SVE: Training set	Neural SVE: Test set	DeepONet: Training set	DeepONet: Test set
			

Table 6.8: Sample Neural SVE and DeepONet paths from the training and the test set for the rough Heston equation and $n = 2000$. Blue (barely visible) are the original paths and orange the learned approximations.

Comparison to neural SDEs

Introduced in [Kid22], a neural stochastic differential equation (neural SDE) is defined by

$$\begin{aligned} Z_0 &= L_\theta(\xi), \\ Z_t &= Z_0\, g_\theta(t) + \int_0^t \mu_\theta(s, Z_s)\,\mathrm{d}s + \int_0^t \sigma_\theta(s, Z_s)\,\mathrm{d}B_s, \\ X_t &= \Pi_\theta(Z_t), \quad t \in [0, T], \end{aligned} \tag{6.13}$$

where all objects are defined as in the neural SVE (6.2). Since the neural SDE is missing the kernel functions $K_{\mu,\theta}$ and $K_{\sigma,\theta}$ compared to the neural SVE (6.2), it is not able to fully capture the dynamics induced by SVEs.

Note that due to the need of discretizing the time interval when it comes to computations, some of the properties introduced by the kernels are attenuated. However, the memory structure of an SVE is a property which can be learned by a neural SVE but in general not by a neural SDE since SDEs posses the Markov property. Therefore, to see the potential capabilities of neural SVEs compared to neural SDEs, it is best to look at examples where the dependency on the whole path plays a crucial role. To construct such an example, we consider the kernels

$$K_\mu(s,t) := K_\sigma(s,t) := K(t-s) = \begin{cases} 1, & \text{if } (t-s) \leq T/4, \\ -1, & \text{if } (t-s) > T/4, \end{cases}$$

and aim to learn solutions to the one-dimensional SVE

$$X_t = \xi + \int_0^t K(t-s)(2 - X_s)\,\mathrm{d}s + \int_0^t K(t-s)\sqrt{|X_s|}\,\mathrm{d}B_s, \qquad t \in [0, T], \tag{6.14}$$

where $\xi \sim \mathscr{N}(5, 0.5)$ and $T = 5$. The process $(X_t)_{t\in[0,5]}$ is expected to decrease in the first quarter of the interval $[0, 5]$ where $K(t - s) = 1$ holds due to the mean-reverting effect of the drift coefficient $\mu(s, x) = 2 - x$, then something unpredictable will happen and finally in the last part of the interval $t \in [0, 5]$ where the kernels attain -1 for a large proportion of $s \in [0, t]$, the process might become big due to the turning sign in the drift. Hence, it is to expect that the path dependency will have a substantial impact.
We learn the dynamics of equation (6.14) simulated on an equally-sized grid with grid size $\Delta t = 0.1$ by a neural SDE and by a neural SVE for a dataset of size $n = 500$ and compare the results in Table 6.9. It can be observed that the neural SDE fails to learn the dynamics of (6.14) properly while the neural SVE performs well.

Example paths of the training and the testing sets together with their learned approximations are shown in Table 6.10.

Neural SVE	Train set	Test set
$n = 500$	0.008	0.009

Neural SDE	Train set	Test set
$n = 500$	0.19	0.21

Table 6.9: Relative L^2-losses after training for the SVE (6.14).

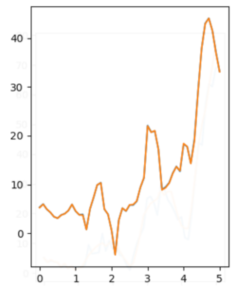
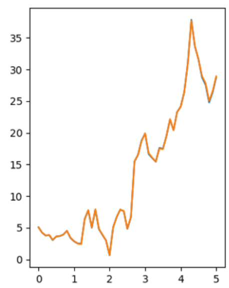
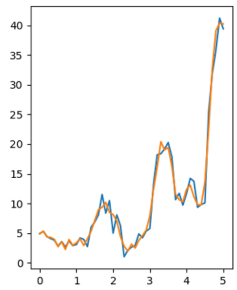
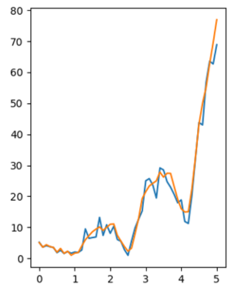
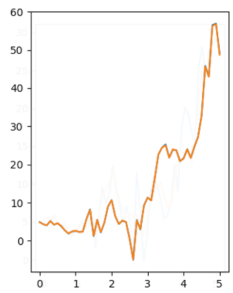
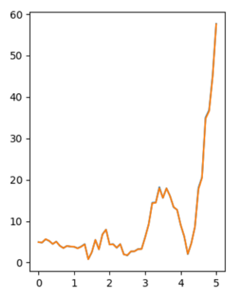
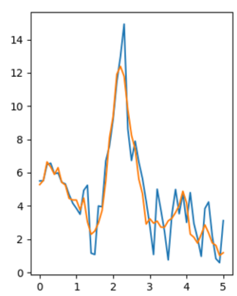
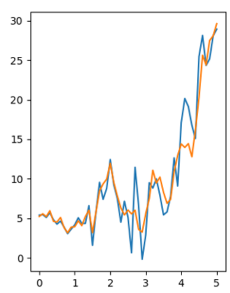
Neural SVE: Training set	Neural SVE: Test set	Neural SDE: Training set	Neural SDE: Test set
			
			

Table 6.10: Sample Neural SVE and neural SDE paths from the training and the test set for the SVE (6.14) and $n = 500$. Blue are the original paths and orange the learned approximations.

Bibliography

[ADPF18] Luisa Andreis, Paolo Dai Pra, and Markus Fischer, *McKean-Vlasov limit for interacting systems with simultaneous jumps*, Stoch. Anal. Appl. **36** (2018), no. 6, 960–995.

[AJ21] Eduardo Abi Jaber, *Weak existence and uniqueness for affine stochastic Volterra equations with L^1-kernels*, Bernoulli **27** (2021), no. 3, 1583–1615.

[AJCLP21] Eduardo Abi Jaber, Christa Cuchiero, Martin Larsson, and Sergio Pulido, *A weak solution theory for stochastic Volterra equations of convolution type*, Ann. Appl. Probab. **31** (2021), no. 6, 2924–2952.

[AJEE19a] Eduardo Abi Jaber and Omar El Euch, *Markovian structure of the Volterra Heston model*, Statist. Probab. Lett. **149** (2019), 63–72.

[AJEE19b] ———, *Multifactor approximation of rough volatility models*, SIAM J. Financial Math. **10** (2019), no. 2, 309–349.

[AJLP19] Eduardo Abi Jaber, Martin Larsson, and Sergio Pulido, *Affine Volterra processes*, Ann. Appl. Probab. **29** (2019), no. 5, 3155–3200.

[AN97] Elisa Alòs and David Nualart, *Anticipating stochastic Volterra equations*, Stochastic Process. Appl. **72** (1997), no. 1, 73–95.

[Apo67] Tom M. Apostol, *Calculus. Vol. I: One-variable calculus, with an introduction to linear algebra*, Blaisdell Publishing Co. [Ginn and Co.], Waltham, Mass.-Toronto, Ont.-London, 1967, pp. xx+666.

[BCC11] François Bolley, José A. Cañizo, and José A. Carrillo, *Stochastic mean-field limit: non-Lipschitz forces and swarming*, Math. Models Methods Appl. Sci. **21** (2011), no. 11, 2179–2210.

[BCD20] Ismaël Bailleul, Rémi Catellier, and François Delarue, *Solving mean field rough differential equations*, Electron. J. Probab. **25** (2020), Paper No. 21, 51.

[BCD21] ———, *Propagation of chaos for mean field rough differential equations*, Ann. Probab. **49** (2021), no. 2, 944–996.

[BFG16] Christian Bayer, Peter Friz, and Jim Gatheral, *Pricing under rough volatility*, Quant. Finance **16** (2016), no. 6, 887–904.

[BM80a] Marc A. Berger and Victor J. Mizel, *Volterra equations with Itô integrals. I*, J. Integral Equations **2** (1980), no. 3, 187–245.

[BM80b] ———, *Volterra equations with Itô integrals. II*, J. Integral Equations **2** (1980), no. 4, 319–337.

[BMM20] Khaled Bahlali, Mohamed Amine Mezerdi, and Brahim Mezerdi, *Stability of McKean-Vlasov stochastic differential equations and applications*, Stoch. Dyn. **20** (2020), no. 1, 2050007, 19.

[Bog07] V. I. Bogachev, *Measure theory. Vol. I*, Springer-Verlag, Berlin, 2007.

[Bry68] V. W. Bryant, *A remark on a fixed-point theorem for iterated mappings*, Amer. Math. Monthly **75** (1968), 399–400.

[BS12] Fischer Black and Myron Scholes, *The pricing of options and corporate liabilities [reprint of J. Polit. Econ.* **81** *(1973), no. 3, 637–654]*, Financial risk measurement and management, Internat. Lib. Crit. Writ. Econ., vol. 267, Edward Elgar, Cheltenham, 2012, pp. 100–117. MR 3235225

[Car16] René Carmona, *Lectures on BSDEs, stochastic control, and stochastic differential games with financial applications*, Financial Mathematics, vol. 1, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2016.

[CBDJ19] Ricky T. Q. Chen, Jens Behrmann, David K Duvenaud, and Joern-Henrik Jacobsen, *Residual flows for invertible generative modeling*, Advances in Neural Information Processing Systems (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), vol. 32, Curran Associates, Inc., 2019.

[CD01] L. Coutin and L. Decreusefond, *Stochastic Volterra equations with singular kernels*, Stochastic analysis and mathematical physics, Progr. Probab., vol. 50, Birkhäuser Boston, Boston, MA, 2001, pp. 39–50.

[CD18a] René Carmona and François Delarue, *Probabilistic theory of mean field games with applications. I*, Probability Theory and Stochastic Modelling, vol. 83, Springer, Cham, 2018, Mean field FBSDEs, control, and games.

[CD18b] ———, *Probabilistic theory of mean field games with applications. II*, Probability Theory and Stochastic Modelling, vol. 84, Springer, Cham, 2018, Mean field games with common noise and master equations.

[CD22a] Louis-Pierre Chaintron and Antoine Diez, *Propagation of chaos: a review of models, methods and applications. I. Models and methods*, Kinet. Relat. Models **15** (2022), no. 6, 895–1015.

[CD22b] ———, *Propagation of chaos: a review of models, methods and applications. II. Applications*, Kinet. Relat. Models **15** (2022), no. 6, 1017–1173.

[CDFM20] Michele Coghi, Jean-Dominique Deuschel, Peter K. Friz, and Mario Maurelli, *Pathwise McKean-Vlasov theory with additive noise*, Ann. Appl. Probab. **30** (2020), no. 5, 2355–2392.

[CLP95] W. George Cochran, Jung-Soon Lee, and Jürgen Potthoff, *Stochastic Volterra equations with singular kernels*, Stochastic Process. Appl. **56** (1995), no. 2, 337–349.

[CR98] Fabienne Comte and Eric Renault, *Long memory in continuous-time stochastic volatility models*, Math. Finance **8** (1998), no. 4, 291–323. MR 1645101

[CT20] Christa Cuchiero and Josef Teichmann, *Generalized Feller processes and Markovian lifts of stochastic Volterra processes: the affine case*, J. Evol. Equ. **20** (2020), no. 4, 1301–1348.

[DE97] Paul Dupuis and Richard S. Ellis, *A weak convergence approach to the theory of large deviations*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, 1997, A Wiley-Interscience Publication.

[DK94] Emanuel Derman and Iraj Kani, *Riding on a smile*, Risk **7** (1994).

[Dud02] R. M. Dudley, *Real analysis and probability*, 2 ed., Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2002.

[EER19] Omar El Euch and Mathieu Rosenbaum, *The characteristic function of rough Heston models*, Math. Finance **29** (2019), no. 1, 3–38.

[Ern22] Xavier Erny, *Well-posedness and propagation of chaos for McKean-Vlasov equations with jumps and locally Lipschitz coefficients*, Stochastic Process. Appl. **150** (2022), 192–214.

[FG15] Nicolas Fournier and Arnaud Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probab. Theory Related Fields **162** (2015), no. 3-4, 707–738.

[GJR18] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum, *Volatility is rough*, Quant. Finance **18** (2018), no. 6, 933–949. MR 3805308

[GR11] István Gyöngy and Miklós Rásonyi, *A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients*, Stochastic Process. Appl. **121** (2011), no. 10, 2189–2200.

[Gra92] Carl Graham, *McKean-Vlasov Itô-Skorohod equations, and nonlinear diffusions with discrete jump sets*, Stochastic Process. Appl. **40** (1992), no. 1, 69–82.

[GRR71] A. M. Garsia, E. Rodemich, and H. Rumsey, Jr., *A real variable lemma and the continuity of paths of some Gaussian processes*, Indiana Univ. Math. J. **20** (1970/71), 565–578.

[Hes93] Steven L. Heston, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, Rev. Financ. Stud. **6** (1993), no. 2, 327–343. MR 3929676

[HS12] Martina Hofmanová and Jan Seidler, *On weak solutions of stochastic differential equations*, Stoch. Anal. Appl. **30** (2012), no. 1, 100–121. MR 2870529

[HW23] Xing Huang and Xucheng Wang, *Path dependent mckean-vlasov sdes with hölder continuous diffusion*, Discrete and Continuous Dynamical Systems - S **16** (2023), no. 5, 982–998.

[IS91] E. K. Ifantis and P. D. Siafarikas, *Bounds for modified Bessel functions*, Rendiconti del Circolo Matematico di Palermo Series 2 **40** (1991), 347–356.

[JP12] Jean Jacod and Philip Protter, *Discretization of processes*, Stochastic Modelling and Applied Probability, vol. 67, Springer, Heidelberg, 2012.

[JS03] Jean Jacod and Albert N. Shiryaev, *Limit theorems for stochastic processes*, 2nd ed., Springer, 2003.

[JW17] Pierre-Emmanuel Jabin and Zhenfu Wang, *Mean field limit for stochastic particle systems*, Active particles. Vol. 1. Advances in theory, models, and applications, Model. Simul. Sci. Eng. Technol., Birkhäuser/Springer, Cham, 2017, pp. 379–402.

[Kac56] M. Kac, *Foundations of kinetic theory.*, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III,, ,, 1956, pp. 171–197. MR 84985

[Kac59] ______, *Probability and related topics in physical sciences*, Lectures in applied mathematics (American Mathematical Society) ; 1.A, American Mathematical Society, 1959.

[Kal21] Alexander Kalinin, *Support characterization for regular path-dependent stochastic Volterra integral equations*, Electron. J. Probab. **26** (2021), Paper No. 29, 29.

[Kid22] Patrick Kidger, *On neural differential equations*, 2022.

[Kle14] Achim Klenke, *Probability theory*, second ed., Universitext, Springer, London, 2014, A comprehensive course.

[KP92] Peter E. Kloeden and Eckhard Platen, *Numerical solution of stochastic differential equations*, Applications of Mathematics (New York), vol. 23, Springer-Verlag, Berlin, 1992. MR 1214374

[KP21] Meyer-Brandis Thilo Kalinin, Alexander and Frank Proske, *Stability, uniqueness and existence of solutions to McKean–Vlasov SDEs: a multidimensional Yamada–Watanabe approach*, ArXiv Preprint arXiv:2107.07838 (2021).

[Kre99] Erwin Kreyszig, *Advanced engineering mathematics*, eighth ed., John Wiley & Sons, Inc., New York, 1999. MR 1665766

[Kru14] Raphael Kruse, *Strong and weak approximation of semilinear stochastic evolution equations*, Lecture Notes in Mathematics, vol. 2093, Springer, Cham, 2014.

[KS91] Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991.

[KS14] Gopinath Kallianpur and P. Sundar, *Stochastic analysis and diffusion processes*, Oxford Graduate Texts in Mathematics, vol. 24, Oxford University Press, Oxford, 2014.

[Kur14] Thomas G. Kurtz, *Weak and strong solutions of general stochastic models*, Electron. Commun. Probab. **19** (2014), no. 58, 16.

[LJP+21] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Karniadakis, *Learning nonlinear operators via deeponet based on the universal approximation theorem of operators*, Nature Machine Intelligence **3** (2021), 218–229.

[LS72] N.N. Lebedev and R.A. Silverman, *Special functions and their applications*, Dover Books on Mathematics, Dover Publications, 1972.

[Luk72] Yudell L Luke, *Inequalities for generalized hypergeometric functions*, Journal of Approximation Theory **5** (1972), no. 1, 41–65.

[Mao94] Xuerong Mao, *Approximate solutions for stochastic differential equations with pathwise uniqueness*, Stochastic Anal. Appl. **12** (1994), no. 3, 355–367.

[Mar94] Emilia P. Martins, *Estimating the rate of phenotypic evolution from comparative data*, The American Naturalist **144** (1994), no. 2, 193–209.

[McK66] Henry P. McKean, *A class of markov processes associated with nonlinear parabolic equations*, Proceedings of the National Academy of Sciences of the United States of America **56** (1966), 1907 – 1911.

[McK67] H. P. McKean, Jr., *Propagation of chaos for a class of non-linear parabolic equations*, Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967), Air Force Office Sci. Res., Arlington, Va., 1967, pp. 41–57.

[MOS66] Wilhelm Magnus, Fritz Oberhettinger, and Raj Pal Soni, *Formulas and theorems for the special functions of mathematical physics*, enlarged ed., Die Grundlehren der mathematischen Wissenschaften, Band 52, Springer-Verlag New York, Inc., New York, 1966.

[MP11] Leonid Mytnik and Edwin Perkins, *Pathwise uniqueness for stochastic heat equations with Hölder continuous coefficients: the white noise case*, Probab. Theory Related Fields **149** (2011), no. 1-2, 1–96.

[MPS06] Leonid Mytnik, Edwin Perkins, and Anja Sturm, *On pathwise uniqueness for stochastic heat equations with non-Lipschitz coefficients*, Ann. Probab. **34** (2006), no. 5, 1910–1959.

[MS15] Leonid Mytnik and Thomas S. Salisbury, *Uniqueness for volterra-type stochastic integral equations*, ArXiv preprint arXiv:1502.05513 (2015).

[OHNO21] A. Ouaddah, J. Henderson, J. J. Nieto, and A. Ouahab, *A fractional bihari inequality and some applications to fractional differential equations and stochastic equations*, Mediterranean Journal of Mathematics **18** (2021).

[Ok03] Bernt Ø ksendal, *Stochastic differential equations*, sixth ed., Universitext, Springer-Verlag, Berlin, 2003, An introduction with applications. MR 2001996

[ØZ93] Bernt Øksendal and Tu Sheng Zhang, *The stochastic Volterra equation*, Barcelona Seminar on Stochastic Analysis (St. Feliu de Guíxols, 1991), Progr. Probab., vol. 32, Birkhäuser, Basel, 1993, pp. 168–202.

[PP90] Étienne Pardoux and Philip Protter, *Stochastic Volterra equations with anticipating coefficients*, Ann. Probab. **18** (1990), no. 4, 1635–1655.

[Pro85] Philip Protter, *Volterra equations driven by semimartingales*, Ann. Probab. **13** (1985), no. 2, 519–530.

[Pro92] ———, *Stochastic integration and differential equation*, second ed., Springer-Verlag, Berlin, Heidelberg, 1992.

[Pro04] Philip E. Protter, *Stochastic integration and differential equations*, second ed., Applications of Mathematics (New York), vol. 21, Springer-Verlag, Berlin, 2004, Stochastic Modelling and Applied Probability.

[PS22] David J. Prömel and David Scheffels, *Pathwise uniqueness for singular stochastic volterra equations with hölder coefficients*, 2022.

[PS23a] David J. Prömel and David Scheffels, *Stochastic Volterra equations with Hölder diffusion coefficients*, Stochastic Process. Appl. **161** (2023), 291–315. MR 4578147

[PS23b] David J. Prömel and David Scheffels, *Mean-field stochastic volterra equations*, 2023.

[PS23c] ———, *On the existence of weak solutions to stochastic volterra equations*, 2023.

[PZ20] Victor M. Panaretos and Yoav Zemel, *An invitation to statistics in Wasserstein space*, SpringerBriefs in Probability and Mathematical Statistics, Springer, Cham, 2020.

[Rog97] L. C. G. Rogers, *Arbitrage with fractional Brownian motion*, Math. Finance **7** (1997), no. 1, 95–105. MR 1434408

[Rud87] Walter Rudin, *Real and complex analysis, 3rd ed.*, McGraw-Hill, Inc., USA, 1987.

[RW00] L. C. G. Rogers and David Williams, *Diffusions, Markov processes, and martingales. Vol. 2*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Itô calculus, Reprint of the second (1994) edition.

[RY99] Daniel Revuz and Marc Yor, *Continuous martingales and Brownian motion*, third ed., Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 293, Springer-Verlag, Berlin, 1999.

[RZL17] Prajit Ramachandran, Barret Zoph, and Quoc V. Le, *Searching for activation functions*, 2017.

[Sko61] A. V. Skorohod, *On the existence and uniqueness of solutions of stochastic differential equations*, Sibirsk. Mat. Ž. **2** (1961), 129–137.

[SLG22] Cristopher Salvi, Maud Lemercier, and Andris Gerasimovics, *Neural stochastic pdes: Resolution-invariant learning of continuous spatiotemporal dynamics*, NeurIPS, 2022.

[SV79] Daniel W. Stroock and S. R. Srinivasa Varadhan, *Multidimensional diffusion processes*, vol. 233, Springer-Verlag, Berlin-New York, 1979.

[SW09] Galen R. Shorack and Jon A. Wellner, *Empirical processes with applications to statistics*, Society for Industrial and Applied Mathematics, 2009.

[SWY13] Yufeng Shi, Tianxiao Wang, and Jiongmin Yong, *Mean-field backward stochastic Volterra integral equations*, Discrete Contin. Dyn. Syst. Ser. B **18** (2013), no. 7, 1929–1967.

[Szn91] Alain-Sol Sznitman, *Topics in propagation of chaos*, École d'Été de Probabilités de Saint-Flour XIX—1989, Lecture Notes in Math., vol. 1464, Springer, Berlin, 1991, pp. 165–251.

[TE99] Erkan Tuzel and Ayse Erzan, *Dissipative dynamics and the statistics of energy states of a hookean model for protein folding*, 1999.

[UO30] G. E. Uhlenbeck and L. S. Ornstein, *On the theory of the brownian motion*, Phys. Rev. **36** (1930), 823–841.

[Vas12] Oldrich Vasicek, *An equilibrium characterization of the term structure [reprint of J. Financ. Econ.* **5** *(1977), no. 2, 177–188]*, Financial risk measurement and management, Internat. Lib. Crit. Writ. Econ., vol. 267, Edward Elgar, Cheltenham, 2012, pp. 724–735. MR 3235239

[Ver12] Mark Veraar, *The stochastic Fubini theorem revisited*, Stochastics **84** (2012), no. 4, 543–551.

[Vla67] Alexander A. Vlasov, *The vibrational properties of an electron gas*, Uspekhi Fizicheskih Nauk (1967).

[Vla68] Anatoliĭ Aleksandrovich Vlasov, *The vibrational properties of an electron gas*, Soviet Physics Uspekhi **10** (1968), no. 6, 721.

[Vol90] Vito Volterra, *Leçons sur la théorie mathématique de la lutte pour la vie*, Les Grands Classiques Gauthier-Villars. [Gauthier-Villars Great Classics], Éditions Jacques Gabay, Sceaux, 1990, Reprint of the 1931 original. MR 1189803

[Wan08] Zhidong Wang, *Existence and uniqueness of solutions to stochastic Volterra equations with singular kernels and non-Lipschitz coefficients*, Statist. Probab. Lett. **78** (2008), no. 9, 1062–1071.

[Wat20] Jeremy Watt, *Machine learning refined : foundations, algorithms, and applications*, second edition. ed., Cambridge, 2020 (eng).

[YW71] Toshio Yamada and Shinzo Watanabe, *On the uniqueness of solutions of stochastic differential equations*, J. Math. Kyoto Univ. **11** (1971), 155–167.

[YYK15] Neha Yadav, Anupam Yadav, and Manoj Kumar, *An introduction to neural network methods for differential equations*, SpringerBriefs in Applied Sciences and Technology, Springer, Dordrecht, 2015. MR 3328375

[Zö5] Henryk Zähle, *Space-time regularity of catalytic super-Brownian motion*, Math. Nachr. **278** (2005), no. 7-8, 942–970. MR 2141969

[Zha08] Xicheng Zhang, *Euler schemes and large deviations for stochastic Volterra equations with singular kernels*, J. Differential Equations **244** (2008), no. 9, 2226–2250. MR 2413840

[Zha10] ———, *Stochastic Volterra equations in Banach spaces and stochastic partial differential equation*, J. Funct. Anal. **258** (2010), no. 4, 1361–1425.