Diss. ETH No. 26805

High-dimensional Stochastic Approximation: Algorithms and Convergence Rates

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZÜRICH
(Dr. sc. ETH Zürich)

presented by
TIMO WELTI

MSc ETH Mathematics, ETH Zürich
born on October 1, 1993
citizen of Küsnacht ZH

accepted on the recommendation of
Prof. Dr. Arnulf Jentzen, University of Münster, examiner
Prof. Dr. Philipp Grohs, University of Vienna, co-examiner
Prof. Dr. Peter Kloeden, Goethe University Frankfurt, co-examiner
Prof. Dr. Siddhartha Mishra, ETH Zürich, co-examiner

2020

Examination committee:
Prof. Dr. Arnulf Jentzen, University of Münster, examiner
Prof. Dr. Philipp Grohs, University of Vienna, co-examiner
Prof. Dr. Peter Kloeden, Goethe University Frankfurt, co-examiner
Prof. Dr. Siddhartha Mishra, ETH Zürich, co-examiner

ETH Diss. No. 26805

Funding: ETH Research Grant ETH-47 15-2
Swiss National Science Foundation Project No. 184220

Abstract

In virtually any scenario where there is a desire to make quantitative or qualitative predictions, mathematical models are of crucial importance for predicting quantities of interest. Unfortunately, many models are based on mathematical objects that cannot be calculated exactly. As a consequence, numerical approximation algorithms that approximate such objects are indispensable in practice. These algorithms are often stochastic in nature, which means that there is some form of randomness involved when running them. In this thesis we consider four possibly high-dimensional approximation problems and corresponding stochastic numerical approximation algorithms. More specifically, we prove essentially sharp rates of convergence in the probabilistically weak sense for spatial spectral Galerkin approximations of semi-linear stochastic wave equations driven by multiplicative noise. In addition, we develop an abstract framework that allows us to view full-history recursive multilevel Picard approximation methods from a new perspective and to work out more clearly how these stochastic methods beat the curse of dimensionality in the numerical approximation of semi-linear heat equations. Furthermore, we tackle high-dimensional optimal stopping problems by proposing a stochastic numerical approximation algorithm that is based on deep learning. And finally, we study convergence in the probabilistically strong sense of the overall error arising in deep learning based empirical risk minimisation, one of the main pillars of supervised learning.

Zusammenfassung

In geradezu jeder Situation, in der quantitative oder qualitative Prognosen gefragt sind, sind mathematische Modelle von enormer Wichtigkeit, um relevante Grössen vorauszuberechnen. Unglücklicherweise beruhen viele Modelle auf mathematischen Objekten, die nicht exakt berechenbar sind. Folglich sind numerische Näherungsverfahren, die solche Objekte approximieren, in der Praxis unabdingbar. Oft sind diese Verfahren stochastischer Natur, was bedeutet, dass eine Art von Zufälligkeit bei ihrer Ausführung involviert ist. In vorliegender Dissertation betrachten wir vier möglicherweise hochdimensionale Näherungsprobleme und zugehörige stochastische numerische Näherungsverfahren. Genauer gesagt beweisen wir Raten der Konvergenz im probabilistisch schwachen Sinne, die im Wesentlichen scharf sind, für räumliche spektrale Galerkinapproximationen von semilinearen stochastischen Wellengleichungen, die von multiplikativem Rauschen angetrieben werden. Ausserdem entwickeln wir einen abstrakten Formalismus, der uns erlaubt, vollhistorisch rekursive Multilevel-Picard-Approximationsverfahren aus einem neuen Blickwinkel zu betrachten und klarer herauszuarbeiten, wie diese stochastischen Verfahren den Fluch der Dimensionalität in der numerischen Approximation von semilinearen Wärmeleitungsgleichungen überwinden. Des Weiteren gehen wir hochdimensionale optimale Stoppprobleme an, indem wir ein stochastisches numerisches Näherungsverfahren einführen, das auf Deep Learning beruht. Und schliesslich untersuchen wir Konvergenz im probabilistisch starken Sinne des Gesamtfehlers, der bei Deep-Learning-basierter empirischer Risikominimierung entsteht, einem der zentralen Standbeine des überwachten Lernens.

Preface

The present thesis is a cumulative dissertation. More precisely, Section 1.1 combined with Chapter 2 is a slightly modified version of the preprint Jacobe de Naurois, Jentzen, & Welti [185], Section 1.2 combined with Chapter 3 is a slightly modified version of the preprint Giles, Jentzen, & Welti [134], Section 1.3 combined with Chapter 4 is a slightly modified version of the preprint Becker, Cheridito, Jentzen, & Welti [30], and parts of two paragraphs in the beginning of Chapter 1 combined with Section 1.4 and Chapter 5 are a slightly modified version of the preprint Jentzen & Welti [196].

In the case of each of the preprints [134, 185, 196] I have made major contributions in all aspects of the creation of the work, that is, in the development of the ideas in the work, in the development of the proofs in the work, and in the writing of the work. In the case of the preprint [30] I have made major contributions in the development of the proofs in the work, in the development of the computational examples in the work, in the development of the MATLAB source codes in the work, and in the writing of the work.

Acknowledgements

First and foremost I express my deep gratitude to my doctoral advisor Arnulf Jentzen for offering me the chance to do this PhD, for teaching me invaluable lessons about the nature of mathematics, for having countless exciting and inspiring research discussions with me, and for being a constant source of great encouragement and support. I consider myself extremely lucky to have had both the freedom to work independently whenever I wanted to and an advisor who would spare time for me as soon as I had any item to discuss.

I also owe very special thanks to Philipp Grohs, who welcomed me for a research stay at the University of Vienna, and to Mike Giles, who welcomed me for a research stay at the University of Oxford. I thank them for very enriching research discussions and for creating such a friendly atmosphere that allowed me to enjoy my research stays to the fullest. In addition, many thanks are due to my other co-authors Diyora Salimova, Sebastian Becker, Patrick Cheridito, Sonja Cox, Martin Hutzenthaler, Jan van Neerven, Adam Andersson, Ryan Kurniawan, and Ladislas Jacobe de Naurois for very pleasant and productive collaborations. In addition, I am indebted to my co-examiners Philipp Grohs, Peter Kloeden, and Siddhartha Mishra for their readiness to review this thesis and their valuable support.

Financial support through the ETH Research Grant ETH-47 15-2 *Mild stochastic calculus and numerical approximations for nonlinear stochastic evolution equations with Lévy noise*, through the project *Deep artificial neural network approximations for stochastic partial differential equations: Algorithms and convergence proofs* (project number 184220) funded by the Swiss National Science Foundation, and through the project *Construction of New Smoothness Spaces on Domains* (project number I 3403) funded by the Austrian Science Fund is gratefully acknowledged. Moreover, I gratefully acknowledge financial support from the Zurich Graduate School in Mathematics.

My time at the SAM has been a truly enjoyable one, for which I thank all current and former SAMmies very much. I especially thank Adrian Ruf, Alice Vanel, Bryn Davies, Carlo Marcati, Carlos Parés Pulido, Diyora Salimova, Fabian Müller, Gina Magnussen, Jakob Zech, Kjetil Olsen Lye, Larisa Yaroslavtseva, Luc Grosheintz Laval, Lukas Herrmann, Maxim Rakhuba, Michaela Szölgyenyi, Pancho Romero, Philippe von Wurstemberger, Pratyuksh Bansal, Primož Pušnik, Soumil Gurjar, and Tandri Gauksson for all the laughs, discussions, lunches, and good memories. In addition, I am grateful to David Rottensteiner, Dennis Elbrächter, Julius Berner, Martin Rathmair, Pavol Harár, Rafael Reisenhofer, Sarah Koppensteiner, and Thomas Peter for so kindly welcoming me as one of their peers at the University of Vienna. I am also indebted to Alberto Paganini, Carolina Urzúa Torres, Jaroslav Fowkes, Priya Subramanian, and Yuji Nakatsukasa for making my working environment at the University of Oxford such a pleasant one.

Friends from outside of maths are of crucial importance and have prevented me from going crazy. I thank Ramon Hess for life-long friendship. I thank Jakob Merane, Judith Suter, and Seraina Hügli for their constant support and for these great conversations. I thank my long-distance friends Philipp Dufter, Cynthia Ho, and Leo Wossnig for being close and for letting also the physical distance occasionally be short. I thank Sophie Beck, Jelena Hess, Mauro Hermann, Onur Sazpinar, Carmen Weber, Dominique Eichenberger, Elena Steinrisser, Jakob Herzog, Maria Engel, Rafael Hug, Jana Huisman, Nicole Speck, Beda Büchel, Lukas Ryffel, Philippe Degen, and Robert Knell for making me laugh, for lending an ear, for giving words of advice, and for keeping in touch. I thank Frederik Benzing and Alexis Podolak for being fantastic flatmates. I thank Christopher Gordon and Ludwig Hruza for playing with me in the string trio des amis. I thank Franzi Schiegl, Freede Metje, Anna Siegert, Cinzia Zanovello, Laura Kömürcü, Marlene Baumgartner, and Sebastian Maurer for making my time in Vienna unforgettable. I thank Anja Bitterwolf, Constanze Cavalier, Maike Ludley, Mario Hörbe, and Sara Schöchli for making me feel at home in Oxford.

I would not have been able to achieve anything without the life-long support from my parents Renate Salzgeber and Martin Welti and my sister Nora Welti. I thank them for everything, for their counsel, for cutting me some slack.

Finally, with all my heart I thank Andrea Gollner for her love, for always being there for me, for being awesome. You are the best thing that has ever happened to me.

Timo Welti
Zürich, April 2020

Contents

Chapter 1

Introduction

Mathematical modelling is ubiquitous. In virtually any scenario where there is a desire to make quantitative or qualitative predictions, theoretical insights and gathered data are used to derive mathematical models with the aim to predict values of quantities of interest. To name but a few examples, mathematical models for randomness allow us to make predictions based on statistics, mathematical models for the lift force enable us to construct wings that carry a modern airplane, mathematical models from quantum mechanics pave the way for the design and study of quantum computers, mathematical models that find a majority in parliament and the electorate tell us how much taxes we are obligated to pay, and mathematical models for infectious diseases provide us with possible scenarios for the evolution of a pandemic. In many cases reaching an actual prediction from the derived mathematical model requires solving equations or computing numbers that are not given in an explicit way. Unfortunately, this cannot be done exactly for the vast majority of mathematical models of practical relevance. For this reason, numerical approximation algorithms which are capable of approximatively solving certain equations or approximatively computing certain numbers are required.

Often such numerical approximation algorithms are stochastic in nature, which means that there is some form of randomness involved when running them. On the one hand, this randomness may stem directly from the mathematical model, which may incorporate randomness itself in order to achieve more realistic predictions. This is, for example, the case when numerically approximating solutions of stochastic partial differential equations (SPDEs), when numerically solving optimal stopping problems, and when using machine learning algorithms in order to approximatively solve supervised learning problems. On the other hand, randomness in numerical approximation algorithms may be beneficial even though the mathematical model under consideration is purely deterministic. Examples for this situation include Monte Carlo methods for approximatively computing high-dimensional deterministic integrals or stochastic numerical approximation methods for solving deterministic partial differential equations (deterministic PDEs). In this thesis we consider four possibly high-dimensional approximation problems and corresponding stochastic numerical approximation algorithms.

We first analyse spatial spectral Galerkin approximations of a certain class of semilinear stochastic wave equations (cf. Section 1.1 and Chapter 2). In order to develop our analysis for the approximation error and to prove essentially sharp convergence rates, we interpret solutions of the considered stochastic wave equations as solutions of suitable infinite-dimensional stochastic evolution equations. In this sense the problem of numerically approximating semi-linear stochastic wave equations is extremely high-dimensional.

Thereafter, we examine full-history recursive multilevel Picard (MLP) approximation methods (cf. E et al. [111] and Hutzenthaler et al. [181]), which are stochastic numerical approximation methods capable of solving high-dimensional PDEs (cf. Section 1.2 and Chapter 3). We establish an abstract framework for suitably generalised MLP approximation methods which take values in possibly infinite-dimensional Banach spaces and use this framework to prove a computational complexity result for suitable MLP approximation methods for semi-linear heat equations. This computational complexity result illustrates, in particular, how the overall computational cost grows with the space dimension of the approximated semi-linear heat equation.

In connection with the third and fourth numerical approximation problems, we consider stochastic numerical approximation algorithms that are based on deep learning.[1] Deep learning based algorithms have been applied extremely successfully to overcome fundamental challenges in many different areas, such as image recognition, natural language processing, game intelligence, autonomous driving, and computational advertising, just to name a few. Accordingly, researchers from a wide range of different fields, including, for example, computer science, mathematics, chemistry, medicine, and finance, are investing significant efforts into studying such algorithms and employing them to tackle challenges arising in their fields. In this spirit we propose a deep learning based stochastic numerical approximation algorithm for solving possibly high-dimensional optimal stopping problems (cf. Section 1.3 and Chapter 4). We provide evidence for the effectiveness and efficiency of the proposed algorithm in the case of high-dimensional optimal stopping problems by presenting results for many numerical example problems with different numbers of dimensions.

In spite of the, as mentioned, broad research interest and the accomplishments of deep learning based algorithms in numerous applications, at the moment there is still no rigorous understanding of the reasons why such algorithms produce useful results in certain situations.[2] Consequently, there is no rigorous way to predict, before actually implementing a deep learning based algorithm, in which situations it might perform reliably and in which situations it might fail. This necessitates in many cases a trial-and-error approach in order to move forward, which can cost a lot of time and resources. A thorough mathematical analysis of deep learning based algorithms (in scenarios where it is possible to formulate such an analysis) seems to be crucial in order to make progress on these issues. Moreover, such an analysis may lead to new insights that enable the design of more effective and efficient algorithms. With this situation in mind, we investigate deep learning based empirical risk minimisation, a stochastic numerical approximation algorithm that is one of the main pillars of supervised learning (cf. Section 1.4 and Chapter 5). More specifically, we prove mathematically rigorous convergence results for the overall error by deriving convergence rates for the three different sources of error, that is, for the approximation error, the generalisation error, and the optimisation error. These convergence results are applicable without any restriction on the number of dimensions of the domain on which the function to be learned is defined.

In the following we explain our work on these four numerical approximation problems and the corresponding stochastic numerical approximation algorithms in more detail.

[1] Parts of this paragraph are a slightly modified extract of the preprint Jentzen & Welti [196].
[2] Parts of this paragraph are a slightly modified extract of the preprint Jentzen & Welti [196].

1.1 Stochastic wave equations

In the field of numerical approximation of stochastic evolution equations one distinguishes between two conceptually fundamentally different error criteria, that is, strong convergence and weak convergence. In the case of the finite-dimensional stochastic ordinary differential equations, both strong and weak convergence are quite well understood nowadays; cf., e.g., the standard monographs Kloeden & Platen [203] and Milstein [246]. However, the situation is different in the case of the infinite-dimensional SPDEs (cf., e.g., Walsh [299], Da Prato & Zabczyk [94], Liu & Röckner [230]). In the case of SPDEs with regular non-linearities, strong convergence rates are essentially well understood, whereas a proper understanding of weak convergence rates has still not been reached (cf., e.g., [7, 8, 54–56, 82, 86, 98–100, 132, 162, 165, 166, 185, 191, 210–212, 214, 215, 228, 283, 302–304] for several weak convergence results in the literature). In Chapter 2 and the preprint Jacobe de Naurois, Jentzen, & Welti [185], of which the current section combined with Chapter 2 is a slightly modified version, we derive weak convergence rates for stochastic wave equations. Stochastic wave equations can be used for modelling several evolutionary processes subject to random forces. Examples include the motion of a DNA molecule floating in a fluid and the dilatation of shock waves throughout the sun (cf., e.g., Dalang [95, Section 1]) as well as heat conduction around a ring (cf., e.g., Thomas [292]). Unfortunately, such problems usually involve complicated non-linearities and are inaccessible for current numerical analysis approaches. Nonetheless, rigorous examination of simpler model problems such as the ones considered in Chapter 2 and the preprint [185], respectively, is a key first step. Even though a number of strong convergence rates for stochastic wave equations are available (cf., e.g., [9, 79, 80, 213, 265, 300, 302, 305]), apart from the findings of the preprint [185] and of the subsequent works Harms & Müller [162] and Cox, Jentzen, & Lindner [86] the existing weak convergence results for stochastic wave equations in the literature (cf., e.g., [166, 211, 212, 214, 302, 303]) assume that the diffusion coefficient is constant or, in other words, that the equation is driven by additive noise.

The main contribution of Chapter 2 and the preprint [185], respectively, is the derivation of essentially sharp weak convergence rates for a class of stochastic wave equations large enough to also include the case of multiplicative noise. Roughly speaking, the main result of Chapter 2 (cf. Theorem 2.12 in Subsection 2.2.2) and of the preprint [185] (cf. [185, Theorem 3.7]), respectively, establishes upper bounds for weak errors associated to spatial spectral Galerkin approximations of semi-linear stochastic wave equations under suitable Lipschitz and smoothness assumptions on the drift non-linearity and on the diffusion coefficient as well as under suitable integrability and regularity assumptions on the initial value. In order to employ a mild solution framework, the second-order stochastic wave equations are formulated as first-order two-component systems of stochastic evolution equations on an extended state space. The first component process of the solution process of such a first-order system corresponds to the solution process of the original second-order equation, while the second component process corresponds to the time derivative of the first component process. As is often the case in the context of spatial spectral Galerkin approximations, convergence is obtained in terms of the in absolute value increasing sequence of eigenvalues of a symmetric linear operator.

To illustrate the main result of Chapter 2 and the preprint [185], respectively, in more detail, we consider the following setting as a special case of our general framework (cf. Subsection 2.2.1). Consider the notation in Subsection 2.1.1, let $(H, \langle \cdot, \cdot \rangle_H, \|\cdot\|_H)$

and $(U, \langle\cdot,\cdot\rangle_U, \|\cdot\|_U)$ be separable $\mathbb{R}$-Hilbert spaces, let $T \in (0,\infty)$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a normal filtration $(\mathbb{F}_t)_{t\in[0,T]}$, let $(W_t)_{t\in[0,T]}$ be an id_U-cylindrical $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$-Wiener process, let $(e_n)_{n\in\mathbb{N}=\{1,2,3,\dots\}} \subseteq H$ be an orthonormal basis of H, let $(\lambda_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$ be an increasing sequence, let $A\colon D(A) \subseteq H \to H$ be the linear operator which satisfies $D(A) = \big\{v \in H\colon \sum_{n=1}^\infty |\lambda_n \langle e_n, v\rangle_H|^2 < \infty\big\}$ and $\forall v \in D(A)\colon Av = \sum_{n=1}^\infty -\lambda_n \langle e_n, v\rangle_H e_n$, let $(H_r, \langle\cdot,\cdot\rangle_{H_r}, \|\cdot\|_{H_r})$, $r \in \mathbb{R}$, be a family of interpolation spaces associated to $-A$ (cf., e.g., [279, Section 3.7]), let $(\mathbf{H}_r, \langle\cdot,\cdot\rangle_{\mathbf{H}_r}, \|\cdot\|_{\mathbf{H}_r})$, $r \in \mathbb{R}$, be the family of $\mathbb{R}$-Hilbert spaces which satisfies for all $r \in \mathbb{R}$ that $(\mathbf{H}_r, \langle\cdot,\cdot\rangle_{\mathbf{H}_r}, \|\cdot\|_{\mathbf{H}_r}) = \big(H_{r/2} \times H_{r/2-1/2}, \langle\cdot,\cdot\rangle_{H_{r/2}\times H_{r/2-1/2}}, \|\cdot\|_{H_{r/2}\times H_{r/2-1/2}}\big)$, let $\mathbf{P}_N \in L(\mathbf{H}_0)$, $N \in \mathbb{N}\cup\{\infty\}$, be the linear operators which satisfy for all $N \in \mathbb{N}\cup\{\infty\}$, $v, w \in H$ that $\mathbf{P}_N(v,w) = \big(\sum_{n=1}^N \langle e_n, v\rangle_H e_n, \sum_{n=1}^N \langle e_n, w\rangle_H e_n\big)$, let $\mathbf{A}\colon D(\mathbf{A}) \subseteq \mathbf{H}_0 \to \mathbf{H}_0$ be the linear operator which satisfies $D(\mathbf{A}) = \mathbf{H}_1$ and $\forall (v,w) \in \mathbf{H}_1\colon \mathbf{A}(v,w) = (w, Av)$, and let $\gamma \in (0,\infty)$, $\beta \in (\gamma/2, \gamma]$, $\rho \in [0, 2(\gamma-\beta)]$, $\varrho, C_{\mathbf{F}}, C_{\mathbf{B}} \in [0,\infty)$, $\xi \in L^2(\mathbb{P}|_{\mathbb{F}_0}; \mathbf{H}_{2(\gamma-\beta)})$, $\mathbf{F} \in \mathrm{Lip}^0(\mathbf{H}_0, \mathbf{H}_0)$, $\mathbf{B} \in \mathrm{Lip}^0(\mathbf{H}_0, L_2(U, \mathbf{H}_0))$ satisfy $(-A)^{-\beta} \in L_1(H)$, $\mathbf{F}(\mathbf{H}_\rho) \subseteq \mathbf{H}_{2(\gamma-\beta)}$, $(\mathbf{H}_\rho \ni v \mapsto \mathbf{F}(v) \in \mathbf{H}_{2(\gamma-\beta)}) \in \mathrm{Lip}^0(\mathbf{H}_\rho, \mathbf{H}_{2(\gamma-\beta)})$, $\forall v \in \mathbf{H}_\rho, u \in U\colon \mathbf{B}(v)u \in \mathbf{H}_\gamma$, $\forall v \in \mathbf{H}_\rho\colon (U \ni u \mapsto \mathbf{B}(v)u \in \mathbf{H}_\rho) \in L_2(U, \mathbf{H}_\rho)$, $(\mathbf{H}_\rho \ni v \mapsto (U \ni u \mapsto \mathbf{B}(v)u \in \mathbf{H}_\rho) \in L_2(U, \mathbf{H}_\rho)) \in \mathrm{Lip}^0(\mathbf{H}_\rho, L_2(U, \mathbf{H}_\rho))$, $\forall v \in \mathbf{H}_\rho\colon (U \ni u \mapsto \mathbf{B}(v)u \in \mathbf{H}_\gamma) \in L(U, \mathbf{H}_\gamma)$, $(\mathbf{H}_\rho \ni v \mapsto (U \ni u \mapsto \mathbf{B}(v)u \in \mathbf{H}_\gamma) \in L(U, \mathbf{H}_\gamma)) \in \mathrm{Lip}^0(\mathbf{H}_\rho, L(U, \mathbf{H}_\gamma))$, $\mathbf{F}|_{\mathbf{H}_\varrho} \in C^2(\mathbf{H}_\varrho, \mathbf{H}_0)$, $\mathbf{B}|_{\mathbf{H}_\varrho} \in C^2(\mathbf{H}_\varrho, L_2(U, \mathbf{H}_0))$, $C_{\mathbf{F}} = \sup_{x, v_1, v_2 \in \cap_{r\in\mathbb{R}} \mathbf{H}_r, \max\{\|v_1\|_{\mathbf{H}_0}, \|v_2\|_{\mathbf{H}_0}\}\leq 1} \|\mathbf{F}''(x)(v_1,v_2)\|_{\mathbf{H}_0} < \infty$, and $C_{\mathbf{B}} = \sup_{x, v_1, v_2 \in \cap_{r\in\mathbb{R}} \mathbf{H}_r, \max\{\|v_1\|_{\mathbf{H}_0}, \|v_2\|_{\mathbf{H}_0}\}\leq 1} \|\mathbf{B}''(x)(v_1,v_2)\|_{L_2(U,\mathbf{H}_0)} < \infty$.

Theorem 1.1. *Assume the above setting. Then*

(i) it holds that there exist up to modifications unique $(\mathbb{F}_t)_{t\in[0,T]}$-predictable stochastic processes $\mathbf{X}^N = (X^N, \mathcal{X}^N)\colon [0,T]\times\Omega \to \mathbf{P}_N(\mathbf{H}_\rho)$, $N \in \mathbb{N}\cup\{\infty\}$, such that for all $N \in \mathbb{N}\cup\{\infty\}$, $t \in [0,T]$ it holds that $\sup_{s\in[0,T]} \mathbb{E}\big[\|\mathbf{X}_s^N\|_{\mathbf{H}_\rho}^2\big] < \infty$ and $\mathbb{P}$-a.s. that

$$\mathbf{X}_t^N = e^{t\mathbf{A}}\mathbf{P}_N\xi + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_N\mathbf{F}(\mathbf{X}_s^N)\,\mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_N\mathbf{B}(\mathbf{X}_s^N)\,\mathrm{d}W_s \tag{1.1}$$

and

(ii) it holds that

$$\begin{aligned}
&\sup_{N\in\mathbb{N}} \sup_{\varphi\in C_{\mathrm{b}}^2(\mathbf{H}_0,\mathbb{R})\setminus\{0\}} \left(\frac{(\lambda_N)^{\gamma-\beta}\,\big|\mathbb{E}\big[\varphi(\mathbf{X}_T^\infty)\big] - \mathbb{E}\big[\varphi(\mathbf{X}_T^N)\big]\big|}{\|\varphi\|_{C_{\mathrm{b}}^2(\mathbf{H}_0,\mathbb{R})}}\right)\\
&\leq \max\big\{1, \mathbb{E}\big[\|\xi\|_{\mathbf{H}_\rho}^2\big]\big\}\Big[\mathbb{E}\big[\|\xi\|_{\mathbf{H}_{2(\gamma-\beta)}}\big] + T\,\big\|\mathbf{F}|_{\mathbf{H}_\rho}\big\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,\mathbf{H}_{2(\gamma-\beta)})}\\
&\quad + 2T\,\|(-A)^{-\beta}\|_{L_1(H)}\big\|\mathbf{B}|_{\mathbf{H}_\rho}\big\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,L(U,\mathbf{H}_\gamma))}^2\Big]\max\Big\{1, \big[T\big((C_{\mathbf{F}})^2 + 2\,(C_{\mathbf{B}})^2\big)\big]^{1/2}\Big\}\\
&\quad \cdot \exp\big(T\big[\tfrac{1}{2} + 3\,|\mathbf{F}|_{\mathrm{Lip}^0(\mathbf{H}_0,\mathbf{H}_0)} + 4\,|\mathbf{B}|_{\mathrm{Lip}^0(\mathbf{H}_0,L_2(U,\mathbf{H}_0))}^2\big]\big)\\
&\quad \cdot \exp\Big(T\Big[2\,\big\|\mathbf{F}|_{\mathbf{H}_\rho}\big\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,\mathbf{H}_\rho)} + \big\|\mathbf{B}|_{\mathbf{H}_\rho}\big\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,L_2(U,\mathbf{H}_\rho))}^2\Big]\Big) < \infty.
\end{aligned} \tag{1.2}$$

Theorem 1.1 follows directly from Remark 2.7 and Corollary 2.14 in Subsection 2.2.2, the latter of which is a consequence of Theorem 2.12 in Subsection 2.2.2. Let us now add a few remarks regarding Theorem 1.1.

First, we briefly outline our proof of Theorem 1.1. As usual in the case of weak convergence analysis, the Kolmogorov equation (cf. (2.58) in the proof of Theorem 2.12) is used. Another key ingredient is the Hölder inequality for Schatten norms (cf. (2.61) in the proof of Theorem 2.12). In addition, the proof of Theorem 1.1 employs the mild Itô formula (cf. Da Prato, Jentzen, & Röckner [93, Corollary 1]) to obtain suitable a priori estimates for solutions of (1.1) above (cf. Lemma 2.8 in Subsection 2.2.2 for details). The detailed proof of Theorems 1.1 and 2.12, respectively, can be found in Subsection 2.2.2.

Second, we would like to emphasise that in the general setting of Theorem 1.1, the weak convergence rate established in Theorem 1.1 can *essentially not be improved*. More precisely, Jacobe de Naurois, Jentzen, & Welti [186, Theorem 1.1] proves that for every $\eta \in (0, \infty)$ and every infinite-dimensional separable $\mathbb{R}$-Hilbert space $(H, \langle \cdot, \cdot \rangle_H, \|\cdot\|_H)$ there exist $(U, \langle \cdot, \cdot \rangle_U, \|\cdot\|_U)$, $A\colon D(A) \subseteq H \to H$, $\gamma, c \in (0, \infty)$, $(C_\varepsilon)_{\varepsilon \in (0,\infty)} \subseteq [0, \infty)$, $\rho \in [0, \gamma/2]$, $\xi \in L^2(\mathbb{P}|_{\mathbb{F}_0}; \mathbf{H}_\gamma)$, $\varphi \in C^2_{\mathrm{b}}(\mathbf{H}_0, \mathbb{R})$, $\mathbf{F} \in C^2_{\mathrm{b}}(\mathbf{H}_0, \mathbf{H}_0)$, $\mathbf{B} \in C^2_{\mathrm{b}}(\mathbf{H}_0, L_2(U, \mathbf{H}_0))$ such that $\mathbf{F}(\mathbf{H}_\rho) \subseteq \mathbf{H}_\gamma$, $(\mathbf{H}_\rho \ni v \mapsto \mathbf{F}(v) \in \mathbf{H}_\gamma) \in \operatorname{Lip}^0(\mathbf{H}_\rho, \mathbf{H}_\gamma)$, $\forall v \in \mathbf{H}_\rho, u \in U\colon \mathbf{B}(v)u \in \mathbf{H}_\gamma$, $\forall v \in \mathbf{H}_\rho\colon (U \ni u \mapsto \mathbf{B}(v)u \in \mathbf{H}_\rho) \in L_2(U, \mathbf{H}_\rho)$, $(\mathbf{H}_\rho \ni v \mapsto (U \ni u \mapsto \mathbf{B}(v)u \in \mathbf{H}_\rho) \in L_2(U, \mathbf{H}_\rho)) \in \operatorname{Lip}^0(\mathbf{H}_\rho, L_2(U, \mathbf{H}_\rho))$, $\forall v \in \mathbf{H}_\rho\colon (U \ni u \mapsto \mathbf{B}(v)u \in \mathbf{H}_\gamma) \in L(U, \mathbf{H}_\gamma)$, and $(\mathbf{H}_\rho \ni v \mapsto (U \ni u \mapsto \mathbf{B}(v)u \in \mathbf{H}_\gamma) \in L(U, \mathbf{H}_\gamma)) \in \operatorname{Lip}^0(\mathbf{H}_\rho, L(U, \mathbf{H}_\gamma))$ and such that for all $\varepsilon \in (0, \infty)$, $N \in \mathbb{N}$ it holds that

$$c \cdot (\lambda_N)^{-\eta} \leq \big|\mathbb{E}\big[\varphi\big(\mathbf{X}^\infty_T\big)\big] - \mathbb{E}\big[\varphi\big(\mathbf{X}^N_T\big)\big]\big| \leq C_\varepsilon \cdot (\lambda_N)^{\varepsilon - \eta}. \tag{1.3}$$

Further results on lower bounds for strong and weak errors for stochastic parabolic equations can be found, e.g., in Davie & Gaines [96], Müller-Gronbach, Ritter, & Wagner [251], Müller-Gronbach & Ritter [250], Conus, Jentzen, & Kurniawan [82], and Jentzen & Kurniawan [191].

Third, we illustrate Theorem 1.1 by a simple example (cf. Corollary 2.18 in Subsection 2.2.3). For this let $P_N \in L(H)$, $N \in \mathbb{N} \cup \{\infty\}$, be the linear operators which satisfy for all $N \in \mathbb{N} \cup \{\infty\}$, $v \in H$ that $P_N(v) = \sum_{n=1}^N \langle e_n, v \rangle_H e_n$. In the case where $(H, \langle \cdot, \cdot \rangle_H, \|\cdot\|_H) = (U, \langle \cdot, \cdot \rangle_U, \|\cdot\|_U) = \big(L^2(\mu_{(0,1)}; \mathbb{R}), \langle \cdot, \cdot \rangle_{L^2(\mu_{(0,1)};\mathbb{R})}, \|\cdot\|_{L^2(\mu_{(0,1)};\mathbb{R})}\big)$, $\xi = (\xi_0, \xi_1) \in H^1_0((0,1); \mathbb{R}) \times H$, and $\mathbf{F} = 0$, where $A\colon D(A) \subseteq H \to H$ is the Laplacian with Dirichlet boundary conditions on H, and where $\mathbf{B}\colon H \times H_{-1/2} \to L_2(H, H \times H_{-1/2})$ is the function which satisfies for all $(v, w) \in H \times H_{-1/2}$, $u \in C([0,1], \mathbb{R})$ and $\mu_{(0,1)}$-a.e. $x \in (0,1)$ that $\big(\mathbf{B}(v,w)u\big)(x) = \big(0, v(x)u(x)\big)$, the first component processes $X^N\colon [0,T] \times \Omega \to P_N(H)$, $N \in \mathbb{N} \cup \{\infty\}$, are mild solutions of the SPDEs

$$\ddot{X}_t(x) = \tfrac{\partial^2}{\partial x^2} X_t(x) + P_N X_t(x) \dot{W}_t(x) \tag{1.4}$$

with $X_0(x) = (P_N \xi_0)(x)$, $\dot{X}_0(x) = (P_N \xi_1)(x)$, and $X_t(0) = X_t(1) = 0$ for $x \in (0,1)$, $t \in [0,T]$, $N \in \mathbb{N} \cup \{\infty\}$. In the case $N = \infty$, (1.4) is known as the continuous version of the hyperbolic Anderson model in the literature (cf., e.g., Conus et al. [83]). Theorem 1.1 applied to (1.4) ensures for all $\varphi \in C^2_{\mathrm{b}}(H, \mathbb{R})$, $\varepsilon \in (0, \infty)$ that there exists a real number $C \in [0, \infty)$ such that for all $N \in \mathbb{N}$ it holds that

$$\big|\mathbb{E}\big[\varphi\big(X^\infty_T\big)\big] - \mathbb{E}\big[\varphi\big(X^N_T\big)\big]\big| \leq C \cdot N^{\varepsilon - 1} \tag{1.5}$$

(cf. Corollary 2.18). We thus prove that the spatial spectral Galerkin approximations converge with weak rate 1- to the solution of the continuous version of the hyperbolic Anderson model. The weak rate 1- is exactly twice the well-known strong convergence rate for the continuous version of the hyperbolic Anderson model. To the best of our knowledge, Theorem 1.1 and the corresponding result [185, Theorem 1.1], respectively, are the first result in the literature that establishes an essentially sharp weak convergence rate for the continuous version of the hyperbolic Anderson model. Theorem 1.1 also establishes essentially sharp weak convergence rates for more general semi-linear stochastic wave equations (cf. Corollaries 2.16 and 2.18 in Subsection 2.2.3).

1.2 Generalised multilevel Picard approximations

It is one of the most challenging problems in applied mathematics to approximatively solve high-dimensional PDEs. In particular, most of the numerical approximation schemes studied in the scientific literature, such as finite differences, finite elements, and sparse grids, suffer under the curse of dimensionality (cf. Bellman [32]) in the sense that the number of computational operations needed to compute an approximation with an error of size at most $\varepsilon > 0$ grows at least exponentially in the PDE dimension $d \in \mathbb{N}$ or in the reciprocal of ε. Computing such an approximation with reasonably small error thus becomes unfeasible in dimension greater than, say, ten. Therefore, a fundamental goal of current research activities is to propose and analyse numerical methods with the power to beat the curse of dimensionality in such way that the number of computational operations needed to compute an approximation with an error of size at most $\varepsilon > 0$ grows at most polynomially in both the PDE dimension $d \in \mathbb{N}$ and the reciprocal of ε (cf., e.g., Novak & Woźniakowski [253, Chapter 1; 254, Chapter 9]). In the recent years a number of numerical schemes have been proposed to tackle the problem of approximately solving high-dimensional PDEs, which include deep learning based approximation methods (cf., e.g., [21–23, 28, 30, 45, 66, 75, 106, 110, 117, 123, 127, 146, 158, 159, 169, 178, 187, 231, 234, 238, 262, 266, 288] and the references mentioned therein), branching diffusion approximation methods (cf., e.g., [1, 31, 46, 51, 52, 67, 168, 170–172, 241, 267, 289, 306, 307]), approximation methods based on discretising a corresponding backward stochastic differential equation (BSDE) using iterative regression on function Hamel bases (cf., e.g., [13, 40, 43, 53, 68–71, 87–90, 101, 107, 137–142, 177, 219, 225, 229, 235–237, 247, 248, 261, 272–274, 296, 312] and the references mentioned therein) or using Wiener chaos expansions (cf. Briand & Labart [57] and Geiss & Labart [131]), and MLP approximation methods (cf. [25, 26, 111, 112, 134, 179, 181–183]). So far, deep learning based approximation methods for PDEs seem to work very well in the case of high-dimensional PDEs judging from a large number of numerical experiments. However, there exist only partial results (cf., e.g., [47, 120, 144, 148–150, 180, 194, 218, 268]) and no full error analysis in the scientific literature rigorously justifying their effectiveness in the numerical approximation of high-dimensional PDEs. Moreover, while for branching diffusion methods there is a full error analysis available proving that the curse of dimensionality can be beaten for instances of PDEs with small time horizon and small initial condition, respectively, numerical simulations suggest that such methods fail to work if the time horizon or the initial condition, respectively, are not small anymore. To sum it up, MLP approximation methods currently are, to the best of our knowledge, the only methods for parabolic semilinear PDEs with general time horizons and general initial conditions for which there is a rigorous proof that they are indeed able to beat the curse of dimensionality.

The main purpose of Chapter 3 and the preprint Giles, Jentzen, & Welti [134], of which the current section combined with Chapter 3 is a slightly modified version, is to investigate MLP methods in more depth, to reveal more clearly how these methods can overcome the curse of dimensionality, and to generalise the MLP scheme proposed in Hutzenthaler et al. [181]. In particular, in the main result of Chapter 3 (cf. Theorem 3.14 in Subsection 3.1.6) and of the preprint [134] (cf. [134, Theorem 2.14]), respectively, we develop an abstract framework in which suitably generalised MLP approximations can be formulated (cf. (1.6) in Theorem 1.2 below) and analysed (cf. (i)–(iii) in Theorem 1.2) and, thereafter, apply this abstract framework to derive a computational complexity result for suitable MLP approximations for semi-linear heat equations (cf. Corollary 1.3 below). These resulting MLP approximations for semi-linear heat equations essentially are generalisations of the MLP approximations introduced in [181]. To make the reader more familiar with the contributions of Chapter 3 and the preprint [134], respectively, we now illustrate in Theorem 1.2 below the findings of the main result of Chapter 3 (cf. Theorem 3.14 in Subsection 3.1.6 and [134, Theorem 2.14]) in a simplified situation.

Theorem 1.2. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be a separable $\mathbb{R}$-Banach space, let $\mathfrak{z}, \mathfrak{B}, \kappa, C, c \in [1,\infty)$, $\Theta = \bigcup_{n=1}^{\infty} \mathbb{Z}^n$, $(M_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$, $y \in \mathcal{Y}$ satisfy $\liminf_{j\to\infty} M_j = \infty$, $\sup_{j\in\mathbb{N}} {}^{M_{j+1}}/_{M_j} \le \mathfrak{B}$, and $\sup_{j\in\mathbb{N}} {}^{M_j}/_j \le \kappa$, let $(\mathcal{Z}, \mathscr{Z})$ be a measurable space, let $Z^\theta \colon \Omega \to \mathcal{Z}$, $\theta \in \Theta$, be i.i.d. $\mathscr{F}/\mathscr{Z}$-measurable functions, let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}}, \|\cdot\|_{\mathcal{H}})$ be a separable $\mathbb{R}$-Hilbert space, let $\mathscr{S}$ be the strong σ-algebra on $L(\mathcal{Y},\mathcal{H})$, let $\psi_k \colon \Omega \to L(\mathcal{Y},\mathcal{H})$, $k \in \mathbb{N}_0 = \{0,1,2,\dots\}$, be $\mathscr{F}/\mathscr{S}$-measurable functions, assume that $(Z^\theta)_{\theta\in\Theta}$ and $(\psi_k)_{k\in\mathbb{N}_0}$ are independent, let $\Phi_l \colon \mathcal{Y}\times\mathcal{Y}\times\mathcal{Z} \to \mathcal{Y}$, $l \in \mathbb{N}_0$, be $(\mathscr{B}(\mathcal{Y})\otimes\mathscr{B}(\mathcal{Y})\otimes\mathscr{Z})/\mathscr{B}(\mathcal{Y})$-measurable functions, let $Y^\theta_{n,j} \colon \Omega \to \mathcal{Y}$, $\theta \in \Theta$, $j \in \mathbb{N}$, $n \in (\mathbb{N}_0 \cup \{-1\})$, satisfy for all $n, j \in \mathbb{N}$, $\theta \in \Theta$ that $Y^\theta_{-1,j} = Y^\theta_{0,j} = 0$ and*

$$Y^\theta_{n,j} = \sum_{l=0}^{n-1} \tfrac{1}{(M_j)^{n-l}} \Bigg[\sum_{i=1}^{(M_j)^{n-l}} \Phi_l\big(Y^{(\theta,l,i)}_{l,j}, Y^{(\theta,-l,i)}_{l-1,j}, Z^{(\theta,l,i)}\big) \Bigg], \tag{1.6}$$

let $(\mathrm{Cost}_{n,j})_{(n,j)\in(\mathbb{N}_0\cup\{-1\})\times\mathbb{N}} \subseteq [0,\infty)$ satisfy for all $n, j \in \mathbb{N}$ that $\mathrm{Cost}_{-1,j} = \mathrm{Cost}_{0,j} = 0$ and

$$\mathrm{Cost}_{n,j} \le (M_j)^n \mathfrak{z} + \sum_{l=0}^{n-1} \big[(M_j)^{n-l}(\mathrm{Cost}_{l,j} + \mathrm{Cost}_{l-1,j} + 2\mathfrak{z})\big], \tag{1.7}$$

and assume for all $k \in \mathbb{N}_0$, $n, j \in \mathbb{N}$, $u, v \in \mathcal{Y}$ that $\mathbb{E}\big[\|\Phi_k(Y^0_{k,j}, Y^1_{k-1,j}, Z^0)\|_{\mathcal{Y}}\big] < \infty$ and

$$\max\big\{\mathbb{E}\big[\|\psi_k(\Phi_0(0,0,Z^0))\|^2_{\mathcal{H}}\big], \mathbb{E}\big[\|\psi_k(y)\|^2_{\mathcal{H}}\big]\big\} \le \tfrac{C^2}{k!}, \tag{1.8}$$

$$\mathbb{E}\big[\|\psi_k(\Phi_n(u,v,Z^0))\|^2_{\mathcal{H}}\big] \le c\,\mathbb{E}\big[\|\psi_{k+1}(u-v)\|^2_{\mathcal{H}}\big], \tag{1.9}$$

$$\mathbb{E}\Big[\big\|\psi_k\big(y - \textstyle\sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l\big(Y^0_{l,j}, Y^1_{l-1,j}, Z^0\big)\big]\big)\big\|^2_{\mathcal{H}}\Big] \le c\,\mathbb{E}\Big[\big\|\psi_{k+1}\big(Y^0_{n-1,j} - y\big)\big\|^2_{\mathcal{H}}\Big]. \tag{1.10}$$

Then

(i) it holds for all $n \in \mathbb{N}$ that $\big(\mathbb{E}\big[\|\psi_0(Y^0_{n,n} - y)\|^2_{\mathcal{H}}\big]\big)^{1/2} \le C\big[\tfrac{5ce^\kappa}{M_n}\big]^{n/2} < \infty$,

(ii) it holds for all $n \in \mathbb{N}$ that $\mathrm{Cost}_{n,n} \le (5M_n)^n \mathfrak{z}$, and

(iii) there exists $(N_\varepsilon)_{\varepsilon\in(0,1]} \subseteq \mathbb{N}$ such that it holds for all $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ that $\sup_{n\in\{N_\varepsilon, N_\varepsilon+1,\dots\}} \big(\mathbb{E}\big[\|\psi_0(Y^0_{n,n} - y)\|^2_{\mathcal{H}}\big]\big)^{1/2} \le \varepsilon$ and

$$\mathrm{Cost}_{N_\varepsilon,N_\varepsilon} \le 5\mathfrak{z}e^\kappa C^{2(1+\delta)}\big(1 + \sup_{n\in\mathbb{N}}\big[(M_n)^{-\delta}(25ce^{2\kappa}\mathfrak{B})^{(1+\delta)}\big]^n\big)\varepsilon^{-2(1+\delta)} < \infty. \tag{1.11}$$

Theorem 1.2 follows directly from the more general result in Corollary 3.15 in Subsection 3.1.6, which, in turn, is a consequence of the main result of Chapter 3, Theorem 3.14 in Subsection 3.1.6 (cf. [134, Theorem 2.14]).

In the following we provide some intuitions and further explanations for Theorem 1.2 and illustrate how it is applied in the context of numerically approximating semi-linear heat equations (cf. Corollary 1.3 below and Setting 3.1 in Section 3.2). Theorem 1.2 establishes an upper error bound (cf. (i) in Theorem 1.2) and an upper cost bound (cf. (ii) in Theorem 1.2) for the generalised MLP approximations in (1.6) as well as an abstract complexity result (cf. (iii) in Theorem 1.2), which states that for an approximation accuracy ε in a suitable root mean square sense the computational cost is essentially of order ε^{-2}. The separable $\mathbb{R}$-Banach space $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ is a set which the exact solution y is an element of and where the generalised MLP approximations $Y^{\theta}_{n,j}\colon \Omega \to \mathcal{Y}$, $\theta \in \Theta$, $j \in \mathbb{N}$, $n \in (\mathbb{N}_0 \cup \{-1\})$, which are random variables approximating $y \in \mathcal{Y}$ in an appropriate sense, take values in. When $y \in \mathcal{Y}$ is the solution of a suitable semi-linear heat equation (cf. (1.12) below), elements of $\mathcal{Y}$ are at most polynomially growing functions in $C([0,T] \times \mathbb{R}^d, \mathbb{R})$, where $T \in (0,\infty)$, $d \in \mathbb{N}$ (cf. (3.129) in Subsection 3.2.2.1). The randomness of the generalised MLP approximations $Y^{\theta}_{n,j}$, $\theta \in \Theta$, $j \in \mathbb{N}$, $n \in (\mathbb{N}_0 \cup \{-1\})$, stems from the i.i.d. random variables $Z^{\theta}\colon \Omega \to \mathcal{Z}$, $\theta \in \Theta$, taking values in a measurable space $(\mathcal{Z}, \mathscr{Z})$, which in our example about semi-linear heat equations correspond to standard Brownian motions and on $[0,1]$ uniformly distributed random variables (cf. (1.13) below). Observe that the generalised MLP approximations in (1.6) above are full-history recursive since each iterate depends on all previous iterates. Together with the random variables Z^{θ}, $\theta \in \Theta$, the previous iterates enter through the functions $\Phi_l\colon \mathcal{Y} \times \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y}$, $l \in \mathbb{N}_0$, which thus govern the dynamics of the generalised MLP approximations. This recursive dependence, the consequential nesting of the generalised MLP approximations, and the Monte Carlo sums in (1.6) necessitate a large number of i.i.d. samples indexed by $\theta \in \Theta = \bigcup_{n=1}^{\infty} \mathbb{Z}^n$ in order to formulate the generalised MLP approximations. In connection with this note that it holds for every $n \in (\mathbb{N}_0 \cup \{-1\})$, $j \in \mathbb{N}$ that $Y^{\theta}_{n,j}$, $\theta \in \Theta$, are identically distributed (cf. (v) in Proposition 3.8 in Subsection 3.1.3).

On the other hand, the parameter $j \in \mathbb{N}$ of the generalised MLP approximations $Y^{\theta}_{n,j}$, $\theta \in \Theta$, $j \in \mathbb{N}$, $n \in (\mathbb{N}_0 \cup \{-1\})$, specifies the respective element of the sequence of Monte Carlo numbers $(M_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ (which are assumed to grow to infinity not faster than linearly) and thereby determines the numbers of Monte Carlo samples to be used in (1.6). Thus for every $j \in \mathbb{N}$ we can consider the family $(Y^0_{n,j})_{n\in(\mathbb{N}_0\cup\{-1\})}$ of generalised MLP approximations with Monte Carlo sample numbers based on M_j, of which we pick the j-th element $Y^0_{j,j}$ to approximate $y \in \mathcal{Y}$ (cf. (iii) in Theorem 1.2). More precisely, for every $n \in \mathbb{N}$ the approximation error for $Y^0_{n,n}$ is measured in the root mean square sense in a separable $\mathbb{R}$-Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}}, \|\cdot\|_{\mathcal{H}})$, after linearly mapping it from $\mathcal{Y}$ to $\mathcal{H}$ using the possibly random function $\psi_0\colon \Omega \to L(\mathcal{Y}, \mathcal{H})$ (cf. (i) and (iii) in Theorem 1.2 above). In our example about semi-linear heat equations, $\mathcal{H}$ is nothing but the set of real numbers $\mathbb{R}$ and ψ_0 is the deterministic evaluation of a function in $\mathcal{Y} \subseteq C([0,T] \times \mathbb{R}^d, \mathbb{R})$ at a deterministic approximation point in $[0,T] \times \mathbb{R}^d$ (cf. (3.132) in Subsection 3.2.2.1).

Conversely, the functions $\psi_k\colon \Omega \to L(\mathcal{Y}, \mathcal{H})$, $k \in \mathbb{N}$, correspond in our example to evaluations at suitable random points in $[0,T] \times \mathbb{R}^d$ multiplied with random factors that diminish quickly as $k \in \mathbb{N}$ increases (cf. (3.132) in Subsection 3.2.2.1). Indeed assumption (1.8) above essentially demands that mean square norms of point evaluations of the functions ψ_k, $k \in \mathbb{N}_0$, diminish at least as fast as the reciprocal of the factorial of their index. Due to this, the functions ψ_k, $k \in \mathbb{N}_0$, can be thought of encoding magnitude in an appropriate randomised sense. Assumption (1.9) hence essentially requires for every $k \in \mathbb{N}_0$, $n \in \mathbb{N}$ that the k-magnitude of the dynamics function Φ_n can be bounded (up to a constant) by the $(k+1)$-magnitude of the difference of its first two arguments, while assumption (1.10), roughly speaking, calls for suitable telescopic cancellations (cf. (3.194) in Subsection 3.2.2.5) such that for every $k \in \mathbb{N}_0$ the k-magnitude of the probabilistically weak approximation error of a given MLP iterate (cf. (3.44)–(3.45) in Subsection 3.1.4) can be bounded (up to a constant) by the $(k+1)$-magnitude of the approximation error of the previous MLP iterate.

Furthermore, we think of the real number $\mathfrak{z} \in [1,\infty)$ as a parameter associated to the computational cost of one realisation of Z^0 and for every $n \in (\mathbb{N}_0 \cup \{-1\})$, $j \in \mathbb{N}$ we think of the real number $\mathrm{Cost}_{n,j} \in [0,\infty)$ as an upper bound for the computational cost associated to one realisation of $\psi_0(Y^0_{n,j})$ (cf. (1.7) above). In our application of the abstract framework outlined above, we have that $\mathfrak{z}$ corresponds to the spacial dimension of the considered semi-linear heat equation and we have for every $n \in (\mathbb{N}_0 \cup \{-1\})$, $j \in \mathbb{N}$ that the number $\mathrm{Cost}_{n,j}$ corresponds to an upper bound for the sum of the number of realisations of standard normal random variables and the number of realisations of on $[0,1]$ uniformly distributed random variables used to compute one realisation of $\psi_0(Y^0_{n,j})$ (cf. (3.204) in Subsection 3.2.3.1).

The abstract framework in Theorem 1.2 can be applied to prove convergence and computational complexity results for MLP approximations in more specific settings. We demonstrate this for the example of MLP approximations for semi-linear heat equations. In particular, Corollary 1.3 below (cf. [134, Corollary 1.2]) establishes that the MLP approximations in (1.13), which essentially are generalisations of the MLP approximations introduced in [181], approximate solutions of semi-linear heat equations (1.12) at the origin without the curse of dimensionality (cf. [181, Theorem 1.1] and [182, Theorems 1.1 and 4.1]).

Corollary 1.3. *Let $T \in (0,\infty)$, $p \in [0,\infty)$, $\Theta = \bigcup_{n=1}^\infty \mathbb{Z}^n$, $(M_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ satisfy $\sup_{j\in\mathbb{N}} (M_{j+1}/M_j + M_j/j) < \infty = \liminf_{j\to\infty} M_j$, let $f\colon \mathbb{R} \to \mathbb{R}$ be a Lipschitz continuous function, let $g_d \in C(\mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, satisfy $\sup_{d\in\mathbb{N}} \sup_{x\in\mathbb{R}^d} |g_d(x)|/\max\{1, \|x\|^p_{\mathbb{R}^d}\} < \infty$, for every $d \in \mathbb{N}$ let $y_d \in C([0,T]\times\mathbb{R}^d, \mathbb{R})$ be an at most polynomially growing viscosity solution of*

$$\left(\tfrac{\partial y_d}{\partial t}\right)(t,x) + \tfrac{1}{2}(\Delta_x y_d)(t,x) + f(y_d(t,x)) = 0 \tag{1.12}$$

with $y_d(T,x) = g_d(x)$ for $(t,x) \in (0,T)\times\mathbb{R}^d$, let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $U^\theta\colon \Omega \to [0,1]$, $\theta \in \Theta$, be independent on $[0,1]$ uniformly distributed random variables, let $W^{d,\theta}\colon [0,T]\times\Omega \to \mathbb{R}^d$, $\theta \in \Theta$, $d \in \mathbb{N}$, be independent standard Brownian motions, assume that $(U^\theta)_{\theta\in\Theta}$ and $(W^{d,\theta})_{(d,\theta)\in\mathbb{N}\times\Theta}$ are independent, let $Y^{d,\theta}_{n,j}\colon [0,T]\times\mathbb{R}^d\times\Omega \to \mathbb{R}$, $\theta \in \Theta$, $d,j \in \mathbb{N}$, $n \in (\mathbb{N}_0 \cup \{-1\})$, satisfy for all $n,j,d \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that $Y^{d,\theta}_{-1,j}(t,x) = Y^{d,\theta}_{0,j}(t,x) = 0$ and

$$\begin{aligned} Y^{d,\theta}_{n,j}(T-t,x) = \sum_{l=0}^{n-1} \tfrac{t}{(M_j)^{n-l}} \Bigg[\sum_{i=1}^{(M_j)^{n-l}} \bigg[f\Big(Y^{d,(\theta,l,i)}_{l,j}\big(T-t+U^{(\theta,l,i)}t, x + W^{d,(\theta,l,i)}_{U^{(\theta,l,i)}t}\big)\Big) \\ - \mathbb{1}_{\mathbb{N}}(l) f\Big(Y^{d,(\theta,-l,i)}_{l-1,j}\big(T-t+U^{(\theta,l,i)}t, x + W^{d,(\theta,l,i)}_{U^{(\theta,l,i)}t}\big)\Big)\bigg]\Bigg] + \tfrac{1}{(M_j)^n} \Bigg[\sum_{i=1}^{(M_j)^n} g_d\big(x + W^{d,(\theta,0,i)}_t\big)\Bigg], \end{aligned} \tag{1.13}$$

and for every $d,n \in \mathbb{N}$ let $\mathrm{Cost}_{d,n} \in \mathbb{N}_0$ be the number of realisations of standard normal random variables used to compute one realisation of $Y^{d,0}_{n,n}(0,0)$ (cf. (3.229) in Subsection 3.2.3.2 for a precise definition). Then there exist $(N_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbb{N}$ and $(C_\delta)_{\delta\in(0,\infty)} \subseteq (0,\infty)$ such that it holds for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ that $\mathrm{Cost}_{d,N_{d,\varepsilon}} \leq C_\delta d^{1+p(1+\delta)} \varepsilon^{-2(1+\delta)}$ and

$$\sup_{n\in\{N_{d,\varepsilon}, N_{d,\varepsilon}+1,\ldots\}} \big(\mathbb{E}\big[|Y^{d,0}_{n,n}(0,0) - y_d(0,0)|^2\big]\big)^{1/2} \leq \varepsilon. \tag{1.14}$$

Corollary 1.3 is a direct consequence of Corollary 3.34 in Subsection 3.2.3.2, while the latter is a direct consequence of Theorem 3.33 in Subsection 3.2.3.2. Theorem 3.33, in turn, follows from either Corollary 3.15 in Subsection 3.1.6 or Theorem 1.2 above. Furthermore, Theorem 3.33 and the corresponding result [134, Theorem 3.17], respectively, essentially are a slight generalisation of [181, Theorem 1.1]. More specifically, the MLP approximations in (3.220) in Theorem 3.33 and in (1.13) above allow for general sequences of Monte Carlo numbers $(M_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ satisfying $\sup_{j\in\mathbb{N}}({}^{M_{j+1}}/{}_{M_j} + {}^{M_j}/{}_{j}) < \infty = \liminf_{j\to\infty} M_j$. This includes, in particular, the special case where $\forall\, j \in \mathbb{N}\colon M_j = j$, which essentially corresponds to the MLP approximations in [181] (cf. [181, (1) in Theorem 1.1]).

1.3 Optimal stopping problems

Nowadays many financial derivatives which are traded on stock and futures exchanges, such as American or Bermudan options, are of early exercise type. Contrary to European options, the holder of such an option has the right to exercise before the time of maturity. In models from mathematical finance for the appropriate pricing of early exercise options this aspect gives rise to optimal stopping problems. The dimension of such optimal stopping problems can often be quite high since it corresponds to the number of underlyings, that is, the number of considered financial assets in the hedging portfolio associated to the optimal stopping problem. Due to the curse of dimensionality (cf. Bellman [32]), high-dimensional optimal stopping problems are, however, notoriously difficult to solve. Such optimal stopping problems can in nearly all cases not be solved explicitly and it is an active topic of research to design and analyse approximation methods which are capable of approximately solving possibly high-dimensional optimal stopping problems. Many different approaches for numerically solving optimal stopping problems and, in particular, American and Bermudan option pricing problems have been studied in the literature; cf., e.g., [3, 5, 6, 14, 15, 19, 20, 28–30, 33–39, 41–43, 48, 53, 58–61, 65, 73, 75, 77, 81, 97, 104, 118, 119, 126, 127, 129, 135, 139, 143, 146, 155, 164, 188, 189, 197, 204–209, 217, 219–221, 223, 224, 232, 233, 257, 270, 271, 276, 278, 287, 288, 293–295, 301]. For example, such approaches include approximating the Snell envelope or continuation values (cf., e.g., [15, 65, 232, 293]), computing optimal exercise boundaries (cf., e.g., [5]), and dual methods (cf., e.g., [164, 270]). Whereas in [164, 207] artificial neural networks with one hidden layer have been employed to approximate continuation values, more recently numerical approximation methods for American and Bermudan option pricing that are based on deep learning have been introduced; cf., e.g., [28–30, 75, 127, 221, 287, 288]. More precisely, in [287, 288] deep neural networks (DNNs) are used to approximately solve the corresponding obstacle PDE problem, in [28, 30] the corresponding optimal stopping problem is tackled directly with deep learning based algorithms, [127] applies an extension of the deep BSDE solver from [110, 158] to the corresponding reflected BSDE problem, [75] suggests a different deep learning based algorithm that relies on discretising BSDEs, and in [29, 221] DNN based variants of the classical algorithm introduced by Longstaff & Schwartz [232] are examined.

In Chapter 4 and the preprint Becker et al. [30], of which the current section combined with Chapter 4 is a slightly modified version, we propose an algorithm for solving general possibly high-dimensional optimal stopping problems; cf. Framework 4.2 in Subsection 4.2.2 and [30, Framework 3.2]. In spirit it is similar to the algorithm introduced in

Becker, Cheridito, & Jentzen [28]. The proposed algorithm is based on deep learning and computes both approximations for an optimal stopping strategy and the optimal expected pay-off associated to the considered optimal stopping problem. In the context of pricing early exercise options these correspond to approximations for an optimal exercise strategy and the price of the considered option, respectively. The derivation and implementation of the proposed algorithm consist of essentially the following three steps.

(I) A neural network architecture for in an appropriate sense ‘randomised’ stopping times (cf. (4.31) in Subsection 4.1.4) is established in such a way that varying the neural network parameters leads to different randomised stopping times being expressed. This neural network architecture is used to replace the supremum of the expected pay-off over suitable stopping times (which constitutes the generic optimal stopping problem) by the supremum of a suitable objective function over neural network parameters (cf. (4.38)–(4.39) in Subsection 4.1.5).

(II) A stochastic gradient ascent-type optimisation algorithm is employed to compute neural network parameters that approximately maximise the objective function (cf. Subsection 4.1.6).

(III) From these neural network parameters and the corresponding randomised stopping time, a true stopping time is constructed which serves as the approximation for an optimal stopping strategy (cf. (4.44) and (4.46) in Subsection 4.1.7). In addition, an approximation for the optimal expected pay-off is obtained by computing a Monte Carlo approximation of the expected pay-off under this approximately optimal stopping strategy (cf. (4.45) in Subsection 4.1.7).

It follows from (III) that the proposed algorithm computes a low-biased approximation of the optimal expected pay-off (cf. (4.48) in Subsection 4.1.7). Yet a large number of numerical experiments where a reference value is available (cf. Section 4.3) show that the bias appears to become small quickly during training and that a very satisfying accuracy can be achieved in short computation time, even in high dimensions (cf. the introductory paragraph of Chapter 4 for a brief overview of the numerical computations that have been performed). Moreover, in (I) we resort to randomised stopping times in order to circumvent the discrete nature of stopping times that attain only finitely many different values. As a result it is possible in (II) to tackle the arising optimisation problem with a stochastic gradient ascent-type algorithm. Furthermore, while the focus in Chapter 4 lies on American and Bermudan option pricing, the proposed algorithm can also be applied to optimal stopping problems that arise in other areas where the underlying stochastic process can be efficiently simulated. Apart from this, we only rely on the assumption that the stochastic process to be optimally stopped is a Markov process (cf. Subsection 4.1.4). But this assumption is no substantial restriction since, on the one hand, it is automatically fulfilled in many relevant problems and, on the other hand, a discrete stochastic process that is not a Markov process can be replaced by a Markov process of higher dimension that aggregates all necessary information (cf., e.g., [28, Subsection 4.3] and, e.g., Subsection 4.3.4.4).

Next we make a short comparison to the algorithm introduced in [28]. The latter is based on introducing for every point in time where stopping is permitted an auxiliary optimal stopping problem, for which stopping is only allowed at that point in time or later (cf. [28, (4) in Subsection 2.1]). Starting at maturity, these auxiliary problems are solved recursively backwards until the initial time is reached. Thereby in every new step neural network parameters are learned for an objective function that depends, in particular, on the parameters found in the previous steps (cf. [28, Subsection 2.3]). In contrast, in (I) a single objective function is designed. This objective function allows to search in (II) for neural network parameters that maximise the expected pay-off simultaneously over (randomised) stopping times which may decide to stop at any of the admissible points in time. Therefore, the algorithm proposed here does not rely on a recursion over the different time points. In addition, the construction of the final approximation for an optimal stopping strategy in (III) differs from a corresponding construction in [28].

1.4 Empirical risk minimisation

The aim of Chapter 5 and the preprint Jentzen & Welti [196], of which the current section combined with Chapter 5 is a slightly modified version, is to provide a mathematically rigorous full error analysis of deep learning based empirical risk minimisation with quadratic loss function in the probabilistically strong sense, where the underlying DNNs are trained using stochastic gradient descent (SGD) with random initialisation (cf. Theorem 1.4 below and [196, Theorem 1.1]). For a brief illustration of deep learning based empirical risk minimisation with quadratic loss function, consider natural numbers $d, \mathbf{d} \in \mathbb{N}$, a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, random variables $X\colon \Omega \to [0,1]^d$ and $Y\colon \Omega \to [0,1]$, and a measurable function $\mathcal{E}\colon [0,1]^d \to [0,1]$ satisfying $\mathbb{P}$-a.s. that $\mathcal{E}(X) = \mathbb{E}[Y|X]$. The goal is to find a DNN with appropriate architecture and appropriate parameter vector $\theta \in \mathbb{R}^{\mathbf{d}}$ (collecting its weights and biases) such that its realisation $\mathscr{N}_\theta\colon \mathbb{R}^d \to \mathbb{R}$ approximates the target function $\mathcal{E}$ well in the sense that the error $\mathbb{E}[|\mathscr{N}_\theta(X) - \mathcal{E}(X)|^p] = \int_{[0,1]^d} |\mathscr{N}_\theta(x) - \mathcal{E}(x)|^p \, \mathbb{P}_X(\mathrm{d}x) \in [0,\infty)$ for some $p \in [1,\infty)$ is as small as possible. In other words, given X we want $\mathscr{N}_\theta(X)$ to predict Y as reliably as possible. Due to the well-known bias–variance decomposition (cf., e.g., Beck, Jentzen, & Kuckuck [27, Lemma 4.1]), for the case $p = 2$ minimising the error function $\mathbb{R}^{\mathbf{d}} \ni \theta \mapsto \mathbb{E}[|\mathscr{N}_\theta(X) - \mathcal{E}(X)|^2] \in [0,\infty)$ is equivalent to minimising the risk function $\mathbb{R}^{\mathbf{d}} \ni \theta \mapsto \mathbb{E}[|\mathscr{N}_\theta(X) - Y|^2] \in [0,\infty)$ (corresponding to a quadratic loss function). Since in practice the joint distribution of X and Y is typically not known, the risk function is replaced by an empirical risk function based on i.i.d. training samples of (X, Y). This empirical risk is then approximatively minimised using an optimisation method such as SGD. As is often the case for deep learning based algorithms, the overall error arising from this procedure consists of the following three different parts (cf. [27, Lemma 4.3] and Proposition 5.37 in Subsection 5.5.1): (i) the *approximation error* (cf., e.g., [16, 17, 49, 78, 92, 122, 128, 163, 173–176, 226, 255] and the references in the introductory paragraph in Section 5.2), which arises from approximating the target function $\mathcal{E}$ by the considered class of DNNs, (ii) the *generalisation error* (cf., e.g., [18, 27, 47, 91, 113–115, 156, 240, 281, 297]), which arises from replacing the true risk by the empirical risk, and (iii) the *optimisation error* (cf., e.g., [10, 12, 22, 27, 44, 72, 102, 103, 108, 109, 124, 161, 190, 195, 200, 222, 282, 311, 313]), which arises from computing only an approximate minimiser using the selected optimisation method.

In Chapter 5 and the preprint [196], respectively, we derive strong convergence rates for the approximation error, the generalisation error, and the optimisation error separately and combine these findings to prove strong convergence results for the overall error (cf. Subsections 5.5.2 and 5.5.3), as illustrated in Theorem 1.4 below. The convergence speed we obtain (cf. (1.18) in Theorem 1.4) suffers under the curse of dimensionality (cf., e.g., Bellman [32] and Novak & Woźniakowski [253, Chapter 1; 254, Chapter 9]) and is, as a consequence, very slow. To the best of our knowledge, we establish in Chapter 5 and the preprint [196], however, the first full error analysis in the scientific literature for a deep learning based algorithm in the probabilistically strong sense and, moreover, the first full error analysis in the scientific literature for a deep learning based algorithm where SGD with random initialisation is the employed optimisation method. We now present Theorem 1.4, the statement of which is entirely self-contained, before we add further explanations and intuitions for the mathematical objects that are introduced.

Theorem 1.4. *Let* $d, \mathbf{d}, \mathbf{L}, \mathbf{J}, M, K, N \in \mathbb{N}$, $\gamma, L \in \mathbb{R}$, $c \in [\max\{2, L\}, \infty)$, $\mathbf{l} = (\mathbf{l}_0, \dots, \mathbf{l}_\mathbf{L}) \in \mathbb{N}^{\mathbf{L}+1}$, $\mathbf{N} \subseteq \{0, \dots, N\}$, *assume* $0 \in \mathbf{N}$, $\mathbf{l}_0 = d$, $\mathbf{l}_\mathbf{L} = 1$, *and* $\mathbf{d} \geq \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1} + 1)$, *for every* $m, n \in \mathbb{N}$, $s \in \mathbb{N}_0$, $\theta = (\theta_1, \dots, \theta_\mathbf{d}) \in \mathbb{R}^\mathbf{d}$ *with* $\mathbf{d} \geq s + mn + m$ *let* $\mathcal{A}_{m,n}^{\theta,s} \colon \mathbb{R}^n \to \mathbb{R}^m$ *satisfy for all* $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ *that*

$$\mathcal{A}_{m,n}^{\theta,s}(x) = \begin{pmatrix} \theta_{s+1} & \theta_{s+2} & \cdots & \theta_{s+n} \\ \theta_{s+n+1} & \theta_{s+n+2} & \cdots & \theta_{s+2n} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{s+(m-1)n+1} & \theta_{s+(m-1)n+2} & \cdots & \theta_{s+mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \theta_{s+mn+1} \\ \theta_{s+mn+2} \\ \vdots \\ \theta_{s+mn+m} \end{pmatrix}, \tag{1.15}$$

let $\mathbf{a}_i \colon \mathbb{R}^{\mathbf{l}_i} \to \mathbb{R}^{\mathbf{l}_i}$, $i \in \{1, \dots, \mathbf{L}\}$, *satisfy for all* $i \in \mathbb{N} \cap [0, \mathbf{L})$, $x = (x_1, \dots, x_{\mathbf{l}_i}) \in \mathbb{R}^{\mathbf{l}_i}$ *that* $\mathbf{a}_i(x) = (\max\{x_1, 0\}, \dots, \max\{x_{\mathbf{l}_i}, 0\})$, *assume for all* $x \in \mathbb{R}$ *that* $\mathbf{a}_\mathbf{L}(x) = \max\{\min\{x, 1\}, 0\}$, *for every* $\theta \in \mathbb{R}^\mathbf{d}$ *let* $\mathscr{N}_\theta \colon \mathbb{R}^d \to \mathbb{R}$ *satisfy* $\mathscr{N}_\theta = \mathbf{a}_\mathbf{L} \circ \mathcal{A}_{\mathbf{l}_\mathbf{L}, \mathbf{l}_{\mathbf{L}-1}}^{\theta, \sum_{i=1}^{\mathbf{L}-1} \mathbf{l}_i(\mathbf{l}_{i-1}+1)} \circ \mathbf{a}_{\mathbf{L}-1} \circ \mathcal{A}_{\mathbf{l}_{\mathbf{L}-1}, \mathbf{l}_{\mathbf{L}-2}}^{\theta, \sum_{i=1}^{\mathbf{L}-2} \mathbf{l}_i(\mathbf{l}_{i-1}+1)} \circ \ldots \circ \mathbf{a}_1 \circ \mathcal{A}_{\mathbf{l}_1, \mathbf{l}_0}^{\theta, 0}$, *let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, let* $X_j^{k,n} \colon \Omega \to [0,1]^d$, $k, n, j \in \mathbb{N}_0$, *and* $Y_j^{k,n} \colon \Omega \to [0,1]$, $k, n, j \in \mathbb{N}_0$, *be functions, assume that* $(X_j^{0,0}, Y_j^{0,0})$, $j \in \mathbb{N}$, *are i.i.d. random variables, let* $\mathcal{E} \colon [0,1]^d \to [0,1]$ *satisfy* $\mathbb{P}$*-a.s. that* $\mathcal{E}(X_1^{0,0}) = \mathbb{E}[Y_1^{0,0} | X_1^{0,0}]$, *assume for all* $x, y \in [0,1]^d$ *that* $|\mathcal{E}(x) - \mathcal{E}(y)| \leq L\|x - y\|_1$, *let* $\Theta_{k,n} \colon \Omega \to \mathbb{R}^\mathbf{d}$, $k, n \in \mathbb{N}_0$, *and* $\mathbf{k} \colon \Omega \to (\mathbb{N}_0)^2$ *be random variables, assume* $\left(\bigcup_{k=1}^\infty \Theta_{k,0}(\Omega)\right) \subseteq [-c, c]^\mathbf{d}$, *assume that* $\Theta_{k,0}$, $k \in \mathbb{N}$, *are i.i.d., assume that* $\Theta_{1,0}$ *is continuous uniformly distributed on* $[-c, c]^\mathbf{d}$, *let* $\mathcal{R}_J^{k,n} \colon \mathbb{R}^\mathbf{d} \times \Omega \to [0, \infty)$, $k, n, J \in \mathbb{N}_0$, *and* $\mathcal{G}^{k,n} \colon \mathbb{R}^\mathbf{d} \times \Omega \to \mathbb{R}^\mathbf{d}$, $k, n \in \mathbb{N}$, *satisfy for all* $k, n \in \mathbb{N}$, $\omega \in \Omega$, $\theta \in \{\vartheta \in \mathbb{R}^\mathbf{d} \colon (\mathcal{R}_\mathbf{J}^{k,n}(\cdot, \omega) \colon \mathbb{R}^\mathbf{d} \to [0, \infty) \text{ is differentiable at } \vartheta)\}$ *that* $\mathcal{G}^{k,n}(\theta, \omega) = (\nabla_\theta \mathcal{R}_\mathbf{J}^{k,n})(\theta, \omega)$, *assume for all* $k, n \in \mathbb{N}$ *that* $\Theta_{k,n} = \Theta_{k,n-1} - \gamma \mathcal{G}^{k,n}(\Theta_{k,n-1})$, *and assume for all* $k, n \in \mathbb{N}_0$, $J \in \mathbb{N}$, $\theta \in \mathbb{R}^\mathbf{d}$, $\omega \in \Omega$ *that*

$$\mathcal{R}_J^{k,n}(\theta, \omega) = \frac{1}{J} \left[\sum_{j=1}^J |\mathscr{N}_\theta(X_j^{k,n}(\omega)) - Y_j^{k,n}(\omega)|^2 \right] \qquad \text{and} \tag{1.16}$$

$$\mathbf{k}(\omega) \in \arg\min\nolimits_{(l,m) \in \{1, \dots, K\} \times \mathbf{N},\, \|\Theta_{l,m}(\omega)\|_\infty \leq c} \mathcal{R}_M^{0,0}(\Theta_{l,m}(\omega), \omega). \tag{1.17}$$

Then

$$\begin{aligned} &\mathbb{E}\Big[\int_{[0,1]^d} |\mathscr{N}_{\Theta_\mathbf{k}}(x) - \mathcal{E}(x)| \, \mathbb{P}_{X_1^{0,0}}(\mathrm{d}x)\Big] \\ &\leq \frac{dc^3}{[\min\{\mathbf{L}, \mathbf{l}_1, \dots, \mathbf{l}_{\mathbf{L}-1}\}]^{1/d}} + \frac{c^3 \mathbf{L}(\|\mathbf{l}\|_\infty + 1) \ln(eM)}{M^{1/4}} + \frac{\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^\mathbf{L} c^{\mathbf{L}+1}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty + 1)^{-2}]}}. \end{aligned} \tag{1.18}$$

Recall that we denote for every $p \in [1, \infty]$ by $\|\cdot\|_p \colon \left(\bigcup_{n=1}^{\infty} \mathbb{R}^n\right) \to [0, \infty)$ the p-norm of vectors in $\bigcup_{n=1}^{\infty} \mathbb{R}^n$ (cf. Definition 5.9 in Subsection 5.2.1). In addition, note that the function $\Omega \times [0,1]^d \ni (\omega, x) \mapsto |\mathscr{N}_{\Theta_{\mathbf{k}(\omega)}(\omega)}(x) - \mathcal{E}(x)| \in [0, \infty)$ is measurable (cf. Lemma 5.38 in Subsection 5.5.2) and that the expression on the left hand side of (1.18) above is thus well-defined. Theorem 1.4 follows directly from Corollary 5.45 in Subsection 5.5.3, which, in turn, is a consequence of the main result of Chapter 5, Theorem 5.41 in Subsection 5.5.2 (cf. [196, Theorem 6.5]).

In the following we provide additional explanations and intuitions for Theorem 1.4. For every $\theta \in \mathbb{R}^{\mathbf{d}}$ the function $\mathscr{N}_\theta \colon \mathbb{R}^d \to \mathbb{R}$ is the realisation of a fully connected feedforward artificial neural network with $\mathbf{L} + 1$ layers consisting of an input layer of dimension $\mathbf{l}_0 = d$, of $\mathbf{L} - 1$ hidden layers of dimensions $\mathbf{l}_1, \ldots, \mathbf{l}_{\mathbf{L}-1}$, respectively, and of an output layer of dimension $\mathbf{l}_{\mathbf{L}} = 1$ (cf. Definition 5.8 in Subsection 5.1.3). The weights and biases stored in the DNN parameter vector $\theta \in \mathbb{R}^{\mathbf{d}}$ determine the corresponding $\mathbf{L}$ affine linear transformations (cf. (1.15) above). As activation functions we employ the multidimensional versions $\mathbf{a}_1, \ldots, \mathbf{a}_{\mathbf{L}-1}$ (cf. Definition 5.3 in Subsection 5.1.2) of the *rectifier function* $\mathbb{R} \ni x \mapsto \max\{x, 0\} \in \mathbb{R}$ (cf. Definition 5.4 in Subsection 5.1.2) just in front of each of the hidden layers and the *clipping function* $\mathbf{a}_{\mathbf{L}}$ (cf. Definition 5.6 in Subsection 5.1.2) just in front of the output layer.

Furthermore, observe that we assume the target function $\mathcal{E} \colon [0,1]^d \to [0,1]$, the values of which we intend to approximately predict with the trained DNN, to be Lipschitz continuous with Lipschitz constant L. Moreover, for every $k, n \in \mathbb{N}_0$, $J \in \mathbb{N}$ the function $\mathcal{R}_J^{k,n} \colon \mathbb{R}^{\mathbf{d}} \times \Omega \to [0, \infty)$ is the empirical risk based on the J training samples $(X_j^{k,n}, Y_j^{k,n})$, $j \in \{1, \ldots, J\}$ (cf. (1.16) above). Derived from the empirical risk, for every $k, n \in \mathbb{N}$ the function $\mathcal{G}^{k,n} \colon \mathbb{R}^{\mathbf{d}} \times \Omega \to \mathbb{R}^{\mathbf{d}}$ is a (generalised) gradient of the empirical risk $\mathcal{R}_{\mathbf{J}}^{k,n}$ with respect to its first argument, that is, with respect to the DNN parameter vector $\theta \in \mathbb{R}^{\mathbf{d}}$. These gradients are required in order to formulate the training dynamics of the (random) DNN parameter vectors $\Theta_{k,n} \in \mathbb{R}^{\mathbf{d}}$, $k \in \mathbb{N}$, $n \in \mathbb{N}_0$, given by the SGD optimisation method with learning rate γ. Note that the subscript $n \in \mathbb{N}_0$ of these SGD iterates (i.e., DNN parameter vectors) is the current training step number, whereas the subscript $k \in \mathbb{N}$ counts the number of times the SGD iteration has been started from scratch so far. Such a new start entails the corresponding initial DNN parameter vector $\Theta_{k,0} \in \mathbb{R}^{\mathbf{d}}$ to be drawn continuous uniformly from the hypercube $[-c, c]^{\mathbf{d}}$, in accordance with Xavier initialisation (cf. Glorot & Bengio [136]). The (random) double index $\mathbf{k} \in \mathbb{N} \times \mathbb{N}_0$ represents the final choice made for the DNN parameter vector $\Theta_{\mathbf{k}} \in \mathbb{R}^{\mathbf{d}}$ (cf. (1.18) above), concluding the training procedure, and is selected as follows. During training the empirical risk $\mathcal{R}_M^{0,0}$ has been calculated for the subset of the SGD iterates indexed by $\mathbf{N} \subseteq \{0, \ldots, N\}$ provided that they have not left the hypercube $[-c, c]^{\mathbf{d}}$ (cf. (1.17) above). After the SGD iteration has been started and finished K times (with maximally N training steps in each case) the final choice for the DNN parameter vector $\Theta_{\mathbf{k}} \in \mathbb{R}^{\mathbf{d}}$ is made among those SGD iterates for which the calculated empirical risk is minimal (cf. (1.17) above). Observe that we use mini-batches of size $\mathbf{J}$ consisting, during SGD iteration number $k \in \{1, \ldots, K\}$ for training step number $n \in \{1, \ldots, N\}$, of the training samples $(X_j^{k,n}, Y_j^{k,n})$, $j \in \{1, \ldots, \mathbf{J}\}$, and that we reserve the M training samples $(X_j^{0,0}, Y_j^{0,0})$, $j \in \{1, \ldots, M\}$, for checking the value of the empirical risk $\mathcal{R}_M^{0,0}$.

Regarding the conclusion of Theorem 1.4, note that the left hand side of (1.18) is the expectation of the overall L^1-error, that is, the expected L^1-distance between the trained DNN $\mathscr{N}_{\Theta_{\mathbf{k}}}$ and the target function $\mathcal{E}$. It is bounded from above by the right hand side of (1.18), which consists of following three summands: (i) the first summand corresponds to the *approximation error* and converges to zero as the number of hidden layers $\mathbf{L} - 1$ as well as the hidden layer dimensions $\mathbf{l}_1, \ldots, \mathbf{l}_{\mathbf{L}-1}$ increase to infinity, (ii) the second summand corresponds to the *generalisation error* and converges to zero as the number of training samples M used for calculating the empirical risk increases to infinity, and (iii) the third summand corresponds to the *optimisation error* and converges to zero as the total number of times K the SGD iteration has been started from scratch increases to infinity. We would like to point out that the second summand (corresponding to the generalisation error) does not suffer under the curse of dimensionality with respect to any of the variables involved.

The main result of Chapter 5 (cf. Theorem 5.41 in Subsection 5.5.2) and of the preprint [196] (cf. [196, Theorem 6.5]), respectively, covers, in comparison with Theorem 1.4, the more general cases where L^p-norms of the overall L^2-error instead of the expectation of the overall L^1-error are considered (cf. (5.163) in Theorem 5.41), where the training samples are not restricted to unit hypercubes, and where a general stochastic approximation algorithm (cf., e.g., Robbins & Monro [269]) with random initialisation is used for optimisation. Our convergence proof for the optimisation error relies, in fact, on the convergence of the Minimum Monte Carlo method (cf. Proposition 5.34 in Section 5.4) and thus only exploits random initialisation but not the specific dynamics of the employed optimisation method (cf. (5.151) in the proof of Proposition 5.39 in Subsection 5.5.2). In this regard, note that Theorem 1.4 above also includes the application of deterministic gradient descent instead of SGD for optimisation since we do not assume the samples used for gradient iterations to be i.i.d.

Chapter 2

Weak convergence rates for spatial spectral Galerkin approximations of semi-linear stochastic wave equations with multiplicative noise

The content of this chapter is a slightly modified extract of the preprint Jacobe de Naurois, Jentzen, & Welti [185].

In this chapter we investigate weak convergence rates for stochastic wave equations that may be driven by multiplicative noise (cf. Section 1.1 in Chapter 1). More precisely, Theorem 2.12 in Subsection 2.2.2, which is the main result of this chapter, yields upper bounds for weak errors associated to spatial spectral Galerkin approximations of abstract wave-type stochastic evolution equations (SEEs). With the help of Theorem 2.12 we establish, in particular, Theorem 1.1 in Section 1.1, which provides more explicit estimates for weak errors in the case of suitably numbered spatial spectral Galerkin approximations.

This chapter is organised as follows. In Subsection 2.1.1 we present some notation often used in this chapter. Subsection 2.1.2 states mostly well-known existence, uniqueness, and regularity results, while Subsection 2.1.4 collects basic properties about interpolation spaces and semigroups associated to deterministic wave equations. The main result of this chapter, Theorem 2.12, is stated and proven in Subsection 2.2.2. Finally, Subsection 2.2.3 shows how Theorem 2.12 can be applied to stochastic wave equations and, in particular, to the continuous version of the hyperbolic Anderson model (cf. Corollaries 2.16 and 2.18).

2.1 Preliminary results

For the proof of our key results in Section 2.2 below we require a number of basic properties of solutions of Kolmogorov equations and of semigroups associated to wave-type evolution equations, which we collect in this section. More concretely, after presenting some notation in Subsection 2.1.1 we state in Subsection 2.1.2 a well-known existence and uniqueness result for solutions of SEEs with Lipschitz continuous drift and diffusion coefficients (cf. Proposition 2.1) as well as an elementary result providing bounds for solutions of Kolmogorov equations associated to finite-dimensional SEEs (cf. Lemma 2.2).

Furthermore, in Subsection 2.1.4 we recall several elementary and well-known facts about linear operators, semigroups, and interpolation spaces associated to deterministic linear wave-type evolution equations (cf. the setting in Subsection 2.1.3).

2.1.1 Notation

In this subsection we introduce some notation which we employ throughout this chapter. For a set A we denote by $\mathcal{P}(A)$ the power set of A and by $\mathcal{P}_0(A)$ the set of all finite subsets of A. For a metric space (E, d_E), a dense subset $A \subseteq E$, a complete metric space (F, d_F), a uniformly continuous function $f\colon A \to F$, and the unique function $\tilde{f} \in C(E, F)$ which satisfies $\tilde{f}|_A = f$ we often write, for simplicity of presentation, f instead of $\tilde{f}$. For two $\mathbb{R}$-Banach spaces $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ with $V \neq \{0\}$, a natural number $k \in \mathbb{N} = \{1, 2, 3, \ldots\}$, and a function $f \in C^k(V, W)$ we denote by $|f|_{C_b^k(V,W)}, \|f\|_{C_b^k(V,W)} \in [0, \infty]$ the extended real numbers given by

$$|f|_{C_b^k(V,W)} = \sup_{x\in V}\|f^{(k)}(x)\|_{L^{(k)}(V,W)} = \sup_{x\in V}\sup_{v_1,\ldots,v_k\in V\setminus\{0\}} \frac{\|f^{(k)}(x)(v_1, \ldots, v_k)\|_W}{\|v_1\|_V \cdot \ldots \cdot \|v_k\|_V}, \tag{2.1}$$

$$\|f\|_{C_b^k(V,W)} = \|f(0)\|_W + \sum_{\ell=1}^{k}|f|_{C_b^\ell(V,W)} \tag{2.2}$$

and we denote by $C_b^k(V, W)$ the set given by $C_b^k(V, W) = \big\{g \in C^k(V, W)\colon \|g\|_{C_b^k(V,W)} < \infty\big\}$. For two $\mathbb{R}$-Banach spaces $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ with $V \neq \{0\}$, a number $k \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}$, and a function $f \in C^k(V, W)$ we denote by $|f|_{\mathrm{Lip}^k(V,W)}, \|f\|_{\mathrm{Lip}^k(V,W)} \in [0, \infty]$ the extended real numbers given by

$$|f|_{\mathrm{Lip}^k(V,W)} = \begin{cases} \sup_{\substack{x,y\in V,\\ x\neq y}}\left(\frac{\|f(x)-f(y)\|_W}{\|x-y\|_V}\right) & : k = 0 \\ \sup_{\substack{x,y\in V,\\ x\neq y}}\left(\frac{\|f^{(k)}(x)-f^{(k)}(y)\|_{L^{(k)}(V,W)}}{\|x-y\|_V}\right) & : k \in \mathbb{N} \end{cases}, \tag{2.3}$$

$$\|f\|_{\mathrm{Lip}^k(V,W)} = \|f(0)\|_W + \sum_{\ell=0}^{k}|f|_{\mathrm{Lip}^\ell(V,W)} \tag{2.4}$$

and we denote by $\mathrm{Lip}^k(V, W)$ the set given by $\mathrm{Lip}^k(V, W) = \big\{g \in C^k(V, W)\colon \|g\|_{\mathrm{Lip}^k(V,W)} < \infty\big\}$. For two $\mathbb{R}$-inner product spaces $(V, \langle\cdot,\cdot\rangle_V, \|\cdot\|_V)$ and $(W, \langle\cdot,\cdot\rangle_W, \|\cdot\|_W)$ we denote by $(V \times W, \langle\cdot,\cdot\rangle_{V\times W}, \|\cdot\|_{V\times W})$ the $\mathbb{R}$-inner product space which satisfies for all $x_1 = (v_1, w_1), x_2 = (v_2, w_2) \in V \times W$ that $\langle x_1, x_2\rangle_{V\times W} = \langle v_1, v_2\rangle_V + \langle w_1, w_2\rangle_W$. For $\mathbb{R}$-Hilbert spaces $(H_i, \langle\cdot,\cdot\rangle_{H_i}, \|\cdot\|_{H_i})$, $i \in \{1, 2\}$, a real number $p \in [1, \infty)$, and linear operators $A \in L(H_1, H_2)$ and $B \in L(H_1)$ we denote by $\|A\|_{L_p(H_1,H_2)} \in [0, \infty]$ the extended real number given by $\|A\|_{L_p(H_1,H_2)} = [\mathrm{trace}_{H_1}((A^\star A)^{p/2})]^{1/p}$, we denote by $\|B\|_{L_p(H_1)} \in [0, \infty]$ the extended real number given by $\|B\|_{L_p(H_1)} = \|B\|_{L_p(H_1,H_1)}$, we denote by $L_p(H_1, H_2)$ the set given by $L_p(H_1, H_2) = \big\{C \in L(H_1, H_2)\colon \|C\|_{L_p(H_1,H_2)} < \infty\big\}$, and we denote by $L_p(H_1)$ the set given by $L_p(H_1) = L_p(H_1, H_1)$. For an $\mathbb{R}$-Hilbert space $(H, \langle\cdot,\cdot\rangle_H, \|\cdot\|_H)$, an orthonormal basis $\mathbb{B} \subseteq H$ of H, a function $\lambda\colon \mathbb{B} \to \mathbb{R}$, a linear operator $A\colon D(A) \subseteq H \to H$ which satisfies $D(A) = \big\{v \in H\colon \sum_{b\in\mathbb{B}}|\lambda_b\langle b, v\rangle_H|^2 < \infty\big\}$ and $\forall v \in D(A)\colon Av = \sum_{b\in\mathbb{B}} \lambda_b\langle b, v\rangle_H b$, and a function $\varphi\colon \mathbb{R} \to \mathbb{R}$ we denote by $\varphi(A)\colon D(\varphi(A)) \subseteq H \to H$ the linear operator which satisfies $D(\varphi(A)) = \big\{v \in H\colon \sum_{b\in\mathbb{B}}|\varphi(\lambda_b)\langle b, v\rangle_H|^2 < \infty\big\}$ and $\forall v \in D(\varphi(A))\colon \varphi(A)v = \sum_{b\in\mathbb{B}} \varphi(\lambda_b)\langle b, v\rangle_H b$. For a Borel measurable set $A \in \mathcal{B}(\mathbb{R})$ we denote by $\mu_A\colon \mathcal{B}(A) \to [0, \infty]$ the Lebesgue–Borel measure on A.

2.1.2 Existence, uniqueness, and regularity results for SEEs

Proposition 2.1 below is a direct consequence of Da Prato & Zabczyk [94, Theorem 7.4].

Proposition 2.1. *Consider the notation in Subsection 2.1.1, let* $T \in (0,\infty)$, $p \in [2,\infty)$, *let* $(H, \langle\cdot,\cdot\rangle_H, \|\cdot\|_H)$ *and* $(U, \langle\cdot,\cdot\rangle_U, \|\cdot\|_U)$ *be separable* $\mathbb{R}$*-Hilbert spaces with* $H \neq \{0\}$, *let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space with a normal filtration* $(\mathbb{F}_t)_{t\in[0,T]}$, *let* $(W_t)_{t\in[0,T]}$ *be an* id_U*-cylindrical* $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$*-Wiener process, let* $S\colon [0,\infty) \to L(H)$ *be a strongly continuous semigroup, and let* $F \in \mathrm{Lip}^0(H,H)$, $B \in \mathrm{Lip}^0(H, L_2(U,H))$, $\xi \in L^p(\mathbb{P}|_{\mathbb{F}_0}; H)$. *Then there exists an up to modifications unique* $(\mathbb{F}_t)_{t\in[0,T]}$*-predictable stochastic process* $X\colon [0,T]\times\Omega \to H$ *such that for all* $t \in [0,T]$ *it holds that* $\sup_{s\in[0,T]} \mathbb{E}\big[\|X_s\|_H^p\big] < \infty$ *and* $\mathbb{P}$*-a.s. that*

$$X_t = S_t\xi + \int_0^t S_{t-s}F(X_s)\,\mathrm{d}s + \int_0^t S_{t-s}B(X_s)\,\mathrm{d}W_s. \tag{2.5}$$

In the next elementary and well-known result, Lemma 2.2, we present bounds for spatial derivatives of solutions of Kolmogorov equations associated to finite-dimensional SEEs.

Lemma 2.2. *Consider the notation in Subsection 2.1.1, let* $(H, \langle\cdot,\cdot\rangle_H, \|\cdot\|_H)$ *be a finite-dimensional* $\mathbb{R}$*-vector space with* $H \neq \{0\}$, *let* $(U, \langle\cdot,\cdot\rangle_U, \|\cdot\|_U)$ *be a separable* $\mathbb{R}$*-Hilbert space, let* $\mathbb{U} \subseteq U$ *be an orthonormal basis of* U, *let* $T \in (0,\infty)$, $A \in L(H)$, $F \in C^2_{\mathrm{b}}(H,H)$, $B \in C^2_{\mathrm{b}}(H, L_2(U,H))$, $\varphi \in C^2_{\mathrm{b}}(H,\mathbb{R})$, *let* $(\Omega,\mathcal{F},\mathbb{P})$ *be a probability space with a normal filtration* $(\mathbb{F}_t)_{t\in[0,T]}$, *let* $(W_t)_{t\in[0,T]}$ *be an* id_U*-cylindrical* $(\Omega,\mathcal{F},\mathbb{P},(\mathbb{F}_t)_{t\in[0,T]})$*-Wiener process, let* $X^x\colon [0,T]\times\Omega \to H$, $x \in H$, *be* $(\mathbb{F}_t)_{t\in[0,T]}$*-predictable stochastic processes such that for all* $x \in H$, $t \in [0,T]$ *it holds that* $\sup_{s\in[0,T]} \mathbb{E}\big[\|X_s^x\|_H^2\big] < \infty$ *and* $\mathbb{P}$*-a.s. that*

$$X_t^x = e^{tA}x + \int_0^t e^{(t-s)A}F(X_s^x)\,\mathrm{d}s + \int_0^t e^{(t-s)A}B(X_s^x)\,\mathrm{d}W_s, \tag{2.6}$$

and let $u\colon [0,T]\times H \to \mathbb{R}$ *be the function which satisfies for all* $t \in [0,T]$, $x \in H$ *that* $u(t,x) = \mathbb{E}[\varphi(X_t^x)]$. *Then*

(i) it holds that $u \in C^{1,2}([0,T]\times H, \mathbb{R})$,

(ii) it holds for all $(t,x) \in [0,T]\times H$ *that*

$$\big(\tfrac{\partial}{\partial t}u\big)(t,x) = \big(\tfrac{\partial}{\partial x}u\big)(t,x)[Ax + F(x)] + \tfrac{1}{2}\sum_{b\in\mathbb{U}}\big(\tfrac{\partial^2}{\partial x^2}u\big)(t,x)(B(x)b, B(x)b), \tag{2.7}$$

(iii) it holds that

$$\begin{aligned}\sup_{t\in[0,T]} |u(t,\cdot)|_{C^1_{\mathrm{b}}(H,\mathbb{R})} &\leq |\varphi|_{C^1_{\mathrm{b}}(H,\mathbb{R})}\left[\sup_{s\in[0,T]}\|e^{sA}\|_{L(H)}\right] \\ &\quad\cdot\exp\!\left(T\big[|F|_{C^1_{\mathrm{b}}(H,H)} + \tfrac{1}{2}|B|^2_{C^1_{\mathrm{b}}(H,L_2(U,H))}\big]\sup_{s\in[0,T]}\|e^{sA}\|^2_{L(H)}\right) < \infty,\end{aligned} \tag{2.8}$$

and

(iv) it holds that

$$\begin{aligned}&\sup_{t\in[0,T]}|u(t,\cdot)|_{C^2_{\mathrm{b}}(H,\mathbb{R})}\\&\leq\|\varphi\|_{C^2_{\mathrm{b}}(H,\mathbb{R})}\left[\sup_{s\in[0,T]}\|e^{sA}\|^3_{L(H)}\right]\max\Big\{1,\Big[T\big(|F|^2_{C^2_{\mathrm{b}}(H,H)}+2|B|^2_{C^2_{\mathrm{b}}(H,L_2(U,H))}\big)\Big]^{1/2}\Big\}\\&\quad\cdot\exp\!\left(T\big[\tfrac{1}{2}+3|F|_{C^1_{\mathrm{b}}(H,H)}+4|B|^2_{C^1_{\mathrm{b}}(H,L_2(U,H))}\big]\sup_{s\in[0,T]}\|e^{sA}\|^4_{L(H)}\right)<\infty.\end{aligned}\tag{2.9}$$

Proof of Lemma 2.2. Note that it is well-known (cf., e.g., Krylov [216, Sections 4 and 5 and Lemma 5.10]) that the assumption that H is finite-dimensional and the assumptions that $\varphi\in C^2_{\mathrm{b}}(H,\mathbb{R})$, $F\in C^2_{\mathrm{b}}(H,H)$, and $B\in C^2_{\mathrm{b}}(H,L_2(U,H))$ imply (i), (ii), that there exist up to modifications unique $(\mathbb{F}_t)_{t\in[0,T]}$-predictable stochastic processes $X^{x,v_1},X^{x,v_1,v_2}\colon[0,T]\times\Omega\to H$, $x,v_1,v_2\in H$, such that for all $x,v_1,v_2\in H$, $t\in[0,T]$, $p\in[2,\infty)$ it holds that $\sup_{s\in[0,T]}\big(\mathbb{E}\big[\|X^{x,v_1}_s\|^p_H\big]+\mathbb{E}\big[\|X^{x,v_1,v_2}_s\|^p_H\big]\big)<\infty$ and $\mathbb{P}$-a.s. that

$$X^{x,v_1}_t=e^{tA}v_1+\int_0^t e^{(t-s)A}F'(X^x_s)X^{x,v_1}_s\,\mathrm{d}s+\int_0^t e^{(t-s)A}B'(X^x_s)X^{x,v_1}_s\,\mathrm{d}W_s,\tag{2.10}$$

$$\begin{aligned}X^{x,v_1,v_2}_t&=\int_0^t e^{(t-s)A}\big(F''(X^x_s)(X^{x,v_1}_s,X^{x,v_2}_s)+F'(X^x_s)X^{x,v_1,v_2}_s\big)\,\mathrm{d}s\\&\quad+\int_0^t e^{(t-s)A}\big(B''(X^x_s)(X^{x,v_1}_s,X^{x,v_2}_s)+B'(X^x_s)X^{x,v_1,v_2}_s\big)\,\mathrm{d}W_s,\end{aligned}\tag{2.11}$$

and that it holds for all $(t,x)\in[0,T]\times H$, $v_1,v_2\in H$ that

$$\big(\tfrac{\partial}{\partial x}u\big)(t,x)v_1=\mathbb{E}\big[\varphi'(X^x_t)X^{x,v_1}_t\big],\tag{2.12}$$

$$\big(\tfrac{\partial^2}{\partial x^2}u\big)(t,x)(v_1,v_2)=\mathbb{E}\big[\varphi''(X^x_t)(X^{x,v_1}_t,X^{x,v_2}_t)+\varphi'(X^x_t)X^{x,v_1,v_2}_t\big].\tag{2.13}$$

It thus remains to prove (iii)–(iv). For this let $\psi_p\colon H\to\mathbb{R}$, $p\in[2,\infty)$, be the functions which satisfy for all $p\in[2,\infty)$, $x\in H$ that $\psi_p(x)=\|x\|^p_H$. Observe that it holds for all $p\in[2,\infty)$, $x,v_1,v_2\in H$ that $\psi_p\in C^2(H,\mathbb{R})$, $\psi_p'(x)v_1=p\|x\|^{p-2}\langle x,v_1\rangle_H$, and

$$\begin{aligned}&\psi_p''(x)(v_1,v_2)\\&=\begin{cases}2\langle v_1,v_2\rangle_H & :p=2\\ 0 & :(p\neq2)\wedge(x=0)\\ p\|x\|^{p-2}_H\langle v_1,v_2\rangle_H+p(p-2)\|x\|^{p-4}_H\langle x,v_1\rangle_H\langle x,v_2\rangle_H & :x\neq0\end{cases}.\end{aligned}\tag{2.14}$$

An application of the mild Itô formula in Da Prato, Jentzen, & Röckner [93, Corollary 1] on the test functions ψ_p, $p\in[2,\infty)$, and the Cauchy–Schwarz inequality hence yield for all $p\in[2,\infty)$, $x,v\in H$, $t\in[0,T]$ that

$$\begin{aligned}&\mathbb{E}\big[\|X^{x,v}_t\|^p_H\big]=\mathbb{E}\big[\psi_p(X^{x,v}_t)\big]\\&=\psi_p(e^{tA}v)+\int_0^t\mathbb{E}\big[\psi_p'(e^{(t-s)A}X^{x,v}_s)e^{(t-s)A}F'(X^x_s)X^{x,v}_s\big]\,\mathrm{d}s\\&\quad+\tfrac{1}{2}\sum_{b\in\mathbb{U}}\int_0^t\mathbb{E}\big[\psi_p''(e^{(t-s)A}X^{x,v}_s)\big(e^{(t-s)A}(B'(X^x_s)X^{x,v}_s)b,e^{(t-s)A}(B'(X^x_s)X^{x,v}_s)b\big)\big]\,\mathrm{d}s\end{aligned}$$

$$
\begin{aligned}
&\leq \|v\|_H^p \Big[\sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}^p\Big] + p\Big[\sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}^p\Big] |F|_{C_{\mathrm{b}}^1(H,H)} \int_0^t \mathbb{E}\big[\|X_s^{x,v}\|_H^p\big]\,\mathrm{d}s \\
&\quad + \tfrac{p}{2}\Big[\sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}^p\Big] |B|^2_{C_{\mathrm{b}}^1(H,L_2(U,H))} \int_0^t \mathbb{E}\big[\|X_s^{x,v}\|_H^p\big]\,\mathrm{d}s \qquad (2.15)\\
&\quad + \tfrac{p(p-2)}{2}\Big[\sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}^p\Big] |B|^2_{C_{\mathrm{b}}^1(H,L_2(U,H))} \int_0^t \mathbb{E}\big[\|X_s^{x,v}\|_H^p\big]\,\mathrm{d}s \\
&= \|v\|_H^p \Big[\sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}^p\Big] \\
&\quad + p\Big[\sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}^p\Big] \big(|F|_{C_{\mathrm{b}}^1(H,H)} + \tfrac{p-1}{2}|B|^2_{C_{\mathrm{b}}^1(H,L_2(U,H))}\big) \int_0^t \mathbb{E}\big[\|X_s^{x,v}\|_H^p\big]\,\mathrm{d}s.
\end{aligned}
$$

This and Gronwall's lemma show for all $p \in [2,\infty)$, $x, v \in H$ that

$$
\begin{aligned}
&\sup_{t\in[0,T]} \|X_t^{x,v}\|_{L^p(\mathbb{P};H)} \qquad (2.16)\\
&\leq \|v\|_H \Big[\sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}\Big] \exp\Big(T\big[|F|_{C_{\mathrm{b}}^1(H,H)} + \tfrac{p-1}{2}|B|^2_{C_{\mathrm{b}}^1(H,L_2(U,H))}\big] \sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}^p\Big).
\end{aligned}
$$

Furthermore, applying again [93, Corollary 1] on the test function ψ_2, the Cauchy–Schwarz inequality, and the fact that $\forall a, b \in \mathbb{R}\colon ab \leq \frac{a^2+b^2}{2}$ imply for all $x, v_1, v_2 \in H$, $t \in [0,T]$ that

$$
\begin{aligned}
&\mathbb{E}\big[\|X_t^{x,v_1,v_2}\|_H^2\big] \\
&= 2\int_0^t \mathbb{E}\big[\big\langle e^{(t-s)A} X_s^{x,v_1,v_2}, e^{(t-s)A}\big(F''(X_s^x)(X_s^{x,v_1}, X_s^{x,v_2}) + F'(X_s^x)X_s^{x,v_1,v_2}\big)\big\rangle_H\big]\,\mathrm{d}s \\
&\quad + \int_0^t \mathbb{E}\big[\big\|e^{(t-s)A}\big(B''(X_s^x)(X_s^{x,v_1}, X_s^{x,v_2}) + B'(X_s^x)X_s^{x,v_1,v_2}\big)\big\|^2_{L_2(U,H)}\big]\,\mathrm{d}s \\
&\leq \int_0^t \mathbb{E}\big[\|e^{(t-s)A}F''(X_s^x)(X_s^{x,v_1}, X_s^{x,v_2})\|_H^2\big] + \mathbb{E}\big[\|e^{(t-s)A}X_s^{x,v_1,v_2}\|_H^2\big]\,\mathrm{d}s \\
&\quad + 2\Big[\sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}^2\Big] |F|_{C_{\mathrm{b}}^1(H,H)} \int_0^t \mathbb{E}\big[\|X_s^{x,v_1,v_2}\|_H^2\big]\,\mathrm{d}s \qquad (2.17)\\
&\quad + 2\int_0^t \mathbb{E}\big[\|e^{(t-s)A}B''(X_s^x)(X_s^{x,v_1}, X_s^{x,v_2})\|^2_{L_2(U,H)} + \|e^{(t-s)A}B'(X_s^x)X_s^{x,v_1,v_2}\|^2_{L_2(U,H)}\big]\,\mathrm{d}s \\
&\leq \Big[\sup_{s\in[0,T]} \|X_s^{x,v_1}\|^2_{L^4(\mathbb{P};H)} \|X_s^{x,v_2}\|^2_{L^4(\mathbb{P};H)}\Big] T\big(|F|^2_{C_{\mathrm{b}}^2(H,H)} + 2|B|^2_{C_{\mathrm{b}}^2(H,L_2(U,H))}\big) \Big[\sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}^2\Big] \\
&\quad + 2\Big[\sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}^2\Big] \big(\tfrac{1}{2} + |F|_{C_{\mathrm{b}}^1(H,H)} + |B|^2_{C_{\mathrm{b}}^1(H,L_2(U,H))}\big) \int_0^t \mathbb{E}\big[\|X_s^{x,v_1,v_2}\|_H^2\big]\,\mathrm{d}s.
\end{aligned}
$$

Gronwall's lemma and (2.16) hence imply for all $x, v_1, v_2 \in H$ that

$$
\begin{aligned}
&\sup_{t\in[0,T]} \|X_t^{x,v_1,v_2}\|_{L^2(\mathbb{P};H)} \\
&\leq \Big[\sup_{s\in[0,T]} \|X_s^{x,v_1}\|_{L^4(\mathbb{P};H)} \|X_s^{x,v_2}\|_{L^4(\mathbb{P};H)}\Big] \Big[T\big(|F|^2_{C_{\mathrm{b}}^2(H,H)} + 2|B|^2_{C_{\mathrm{b}}^2(H,L_2(U,H))}\big)\Big]^{1/2} \\
&\quad \cdot \Big[\sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}\Big] \exp\Big(T\big[\tfrac{1}{2} + |F|_{C_{\mathrm{b}}^1(H,H)} + |B|^2_{C_{\mathrm{b}}^1(H,L_2(U,H))}\big] \sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}^2\Big) \qquad (2.18)
\end{aligned}
$$

$$\leq \|v_1\|_H \|v_2\|_H \left[\sup_{s\in[0,T]} \|e^{sA}\|^3_{L(H)}\right] \left[T\big(|F|^2_{C^2_{\mathrm{b}}(H,H)} + 2|B|^2_{C^2_{\mathrm{b}}(H,L_2(U,H))}\big)\right]^{1/2}$$
$$\cdot \exp\!\left(T\big[\tfrac{1}{2} + 3|F|_{C^1_{\mathrm{b}}(H,H)} + 4|B|^2_{C^1_{\mathrm{b}}(H,L_2(U,H))}\big] \sup_{s\in[0,T]} \|e^{sA}\|^4_{L(H)}\right).$$

Next note that (2.12), (2.13), (2.16), and (2.18) ensure for all $(t,x) \in [0,T]\times H$, $v_1, v_2 \in H$ that

$$\begin{aligned} \big|\big(\tfrac{\partial}{\partial x}u\big)(t,x)v_1\big| &= |\mathbb{E}[\varphi'(X^x_t)X^{x,v_1}_t]| \leq |\varphi|_{C^1_{\mathrm{b}}(H,\mathbb{R})}\mathbb{E}[\|X^{x,v_1}_t\|_H] \\ &\leq \|v_1\|_H |\varphi|_{C^1_{\mathrm{b}}(H,\mathbb{R})} \left[\sup_{s\in[0,T]} \|e^{sA}\|_{L(H)}\right] \\ &\quad \cdot \exp\!\left(T\big[|F|_{C^1_{\mathrm{b}}(H,H)} + \tfrac{1}{2}|B|^2_{C^1_{\mathrm{b}}(H,L_2(U,H))}\big] \sup_{s\in[0,T]} \|e^{sA}\|^2_{L(H)}\right) \end{aligned} \tag{2.19}$$

and

$$\begin{aligned} &\big|\big(\tfrac{\partial^2}{\partial x^2}u\big)(t,x)(v_1,v_2)\big| = \big|\mathbb{E}\big[\varphi''(X^x_t)(X^{x,v_1}_t, X^{x,v_2}_t) + \varphi'(X^x_t)X^{x,v_1,v_2}_t\big]\big| \\ &\leq |\varphi|_{C^2_{\mathrm{b}}(H,\mathbb{R})} \|X^{x,v_1}_t\|_{L^2(\mathbb{P};H)} \|X^{x,v_2}_t\|_{L^2(\mathbb{P};H)} + |\varphi|_{C^1_{\mathrm{b}}(H,\mathbb{R})}\mathbb{E}[\|X^{x,v_1,v_2}_t\|_H] \\ &\leq \|v_1\|_H\|v_2\|_H |\varphi|_{C^2_{\mathrm{b}}(H,\mathbb{R})} \left[\sup_{s\in[0,T]} \|e^{sA}\|^2_{L(H)}\right] \\ &\qquad \cdot \exp\!\left(T\big[2|F|_{C^1_{\mathrm{b}}(H,H)} + |B|^2_{C^1_{\mathrm{b}}(H,L_2(U,H))}\big] \sup_{s\in[0,T]} \|e^{sA}\|^2_{L(H)}\right) \\ &\quad + \|v_1\|_H\|v_2\|_H |\varphi|_{C^1_{\mathrm{b}}(H,\mathbb{R})} \left[\sup_{s\in[0,T]} \|e^{sA}\|^3_{L(H)}\right] \left[T\big(|F|^2_{C^2_{\mathrm{b}}(H,H)} + 2|B|^2_{C^2_{\mathrm{b}}(H,L_2(U,H))}\big)\right]^{1/2} \\ &\qquad \cdot \exp\!\left(T\big[\tfrac{1}{2} + 3|F|_{C^1_{\mathrm{b}}(H,H)} + 4|B|^2_{C^1_{\mathrm{b}}(H,L_2(U,H))}\big] \sup_{s\in[0,T]} \|e^{sA}\|^4_{L(H)}\right) \\ &\leq \|v_1\|_H\|v_2\|_H \|\varphi\|_{C^2_{\mathrm{b}}(H,\mathbb{R})} \left[\sup_{s\in[0,T]} \|e^{sA}\|^3_{L(H)}\right] \max\!\left\{1, \left[T\big(|F|^2_{C^2_{\mathrm{b}}(H,H)} + 2|B|^2_{C^2_{\mathrm{b}}(H,L_2(U,H))}\big)\right]^{1/2}\right\} \\ &\quad \cdot \exp\!\left(T\big[\tfrac{1}{2} + 3|F|_{C^1_{\mathrm{b}}(H,H)} + 4|B|^2_{C^1_{\mathrm{b}}(H,L_2(U,H))}\big] \sup_{s\in[0,T]} \|e^{sA}\|^4_{L(H)}\right). \end{aligned} \tag{2.20}$$

This completes the proof of Lemma 2.2. □

2.1.3 Setting

Setting 2.1. *Consider the notation in Subsection 2.1.1, let $(H, \langle\cdot,\cdot\rangle_H, \|\cdot\|_H)$ be a separable $\mathbb{R}$-Hilbert space, let $\mathbb{H} \subseteq H$ be a non-empty orthonormal basis of H, let $\lambda\colon \mathbb{H} \to \mathbb{R}$ be a function which satisfies $\sup_{h\in\mathbb{H}} \lambda_h < 0$, let $A\colon D(A) \subseteq H \to H$ be the linear operator which satisfies $D(A) = \big\{v \in H\colon \sum_{h\in\mathbb{H}} |\lambda_h \langle h, v\rangle_H|^2 < \infty\big\}$ and $\forall\, v \in D(A)\colon Av = \sum_{h\in\mathbb{H}} \lambda_h \langle h, v\rangle_H h$, let $(H_r, \langle\cdot,\cdot\rangle_{H_r}, \|\cdot\|_{H_r})$, $r \in \mathbb{R}$, be a family of interpolation spaces associated to $-A$, let $(\mathbf{H}_r, \langle\cdot,\cdot\rangle_{\mathbf{H}_r}, \|\cdot\|_{\mathbf{H}_r})$, $r \in \mathbb{R}$, be the family of $\mathbb{R}$-Hilbert spaces which satisfies for all $r \in \mathbb{R}$ that $(\mathbf{H}_r, \langle\cdot,\cdot\rangle_{\mathbf{H}_r}, \|\cdot\|_{\mathbf{H}_r}) = \big(H_{r/2} \times H_{r/2-1/2}, \langle\cdot,\cdot\rangle_{H_{r/2}\times H_{r/2-1/2}}, \|\cdot\|_{H_{r/2}\times H_{r/2-1/2}}\big)$, and let $\mathbf{A}\colon D(\mathbf{A}) \subseteq \mathbf{H}_0 \to \mathbf{H}_0$ be the linear operator which satisfies $D(\mathbf{A}) = \mathbf{H}_1$ and $\forall\, (v,w) \in \mathbf{H}_1\colon \mathbf{A}(v,w) = (w, Av)$.*

2.1.4 Basic properties of deterministic linear wave equations

The following elementary result, Lemma 2.3, provides a characterisation for the family of $\mathbb{R}$-Hilbert spaces $(\mathbf{H}_r, \langle\cdot,\cdot\rangle_{\mathbf{H}_r}, \|\cdot\|_{\mathbf{H}_r})$, $r \in \mathbb{R}$, from the setting in Subsection 2.1.3.

Lemma 2.3. *Assume Setting 2.1 and let $\mathbf{\Lambda}\colon D(\mathbf{\Lambda}) \subseteq \mathbf{H}_0 \to \mathbf{H}_0$ be the linear operator which satisfies for all $(v,w) \in \mathbf{H}_1$ that $D(\mathbf{\Lambda}) = \mathbf{H}_1$ and*

$$\mathbf{\Lambda}(v,w) = \begin{pmatrix} \sum_{h\in\mathbb{H}} |\lambda_h|^{1/2} \langle h, v\rangle_H h \\ \sum_{h\in\mathbb{H}} |\lambda_h|^{1/2} \langle h, w\rangle_H h \end{pmatrix}. \tag{2.21}$$

Then the $\mathbb{R}$-Hilbert spaces $(\mathbf{H}_r, \langle\cdot,\cdot\rangle_{\mathbf{H}_r}, \|\cdot\|_{\mathbf{H}_r})$, $r \in \mathbb{R}$, are a family of interpolation spaces associated to $\mathbf{\Lambda}$.

Proof of Lemma 2.3. Observe that $\mathbf{\Lambda}\colon D(\mathbf{\Lambda}) \subseteq \mathbf{H}_0 \to \mathbf{H}_0$ is a symmetric diagonal linear operator (cf., e.g., Sell & You [279, Section 3.2]) with $\inf(\sigma_{\mathrm{P}}(\mathbf{\Lambda})) > 0$ and that for all $r \in [0,\infty)$ it holds that

$$\begin{aligned} D(\mathbf{\Lambda}^r) &= \left\{x \in \mathbf{H}_0 \colon \sum_{h\in\mathbb{H}} \Big(|\lambda_h|^r |\langle (h,0), x\rangle_{\mathbf{H}_0}|^2 + |\lambda_h|^r \big|\big\langle \big(0, |\lambda_h|^{1/2} h\big), x\big\rangle_{\mathbf{H}_0}\big|^2\Big) < \infty\right\} \\ &= \left\{(v,w) \in \mathbf{H}_0 \colon \sum_{h\in\mathbb{H}} \Big(|\lambda_h|^r |\langle h, v\rangle_H|^2 + |\lambda_h|^r \big|\big\langle |\lambda_h|^{1/2} h, w\big\rangle_{H_{-1/2}}\big|^2\Big) < \infty\right\} \\ &= \left\{v \in H \colon \sum_{h\in\mathbb{H}} |\lambda_h|^r |\langle h, v\rangle_H|^2 < \infty\right\} \times \left\{w \in H_{-1/2} \colon \sum_{h\in\mathbb{H}} |\lambda_h|^{r-1} \big|\langle h, w\rangle_H\big|^2 < \infty\right\} \\ &= H_{r/2} \times H_{r/2-1/2} = \mathbf{H}_r. \end{aligned} \tag{2.22}$$

Moreover, note that for all $r \in [0,\infty)$, $x_1 = (v_1,w_1), x_2 = (v_2,w_2) \in \mathbf{H}_r$ it holds that

$$\begin{aligned} \langle \mathbf{\Lambda}^r x_1, \mathbf{\Lambda}^r x_2\rangle_{\mathbf{H}_0} &= \left\langle \sum_{h\in\mathbb{H}} |\lambda_h|^{r/2} \langle h, v_1\rangle_H h, \sum_{h\in\mathbb{H}} |\lambda_h|^{r/2} \langle h, v_2\rangle_H h \right\rangle_H \\ &\quad + \left\langle \sum_{h\in\mathbb{H}} |\lambda_h|^{r/2} \big\langle |\lambda_h|^{1/2} h, w_1\big\rangle_{H_{-1/2}} |\lambda_h|^{1/2} h, \sum_{h\in\mathbb{H}} |\lambda_h|^{r/2} \big\langle |\lambda_h|^{1/2} h, w_2\big\rangle_{H_{-1/2}} |\lambda_h|^{1/2} h \right\rangle_{H_{-1/2}} \\ &= \langle (-A)^{r/2} v_1, (-A)^{r/2} v_2\rangle_H + \langle (-A)^{r/2} w_1, (-A)^{r/2} w_2\rangle_{H_{-1/2}} \\ &= \langle v_1, v_2\rangle_{H_{r/2}} + \langle w_1, w_2\rangle_{H_{r/2-1/2}} = \langle x_1, x_2\rangle_{\mathbf{H}_r}. \end{aligned} \tag{2.23}$$

In addition, observe that for all $r \in (-\infty, 0]$, $x = (v,w) \in \mathbf{H}_0$ it holds that

$$\begin{aligned} \|\mathbf{\Lambda}^r x\|_{\mathbf{H}_0}^2 &= \left\| \sum_{h\in\mathbb{H}} |\lambda_h|^{r/2} \langle h, v\rangle_H h \right\|_H^2 + \left\| \sum_{h\in\mathbb{H}} |\lambda_h|^{r/2} \big\langle |\lambda_h|^{1/2} h, w\big\rangle_{H_{-1/2}} |\lambda_h|^{1/2} h \right\|_{H_{-1/2}}^2 \\ &= \|(-A)^{r/2} v\|_H^2 + \|(-A)^{r/2} w\|_{H_{-1/2}}^2 = \|v\|_{H_{r/2}}^2 + \|w\|_{H_{r/2-1/2}}^2 = \|x\|_{\mathbf{H}_r}^2. \end{aligned} \tag{2.24}$$

This completes the proof of Lemma 2.3. ☐

The next elementary and well-known result, Lemma 2.4, can be found, e.g., in a slightly different form in Lindgren [227, Section 5.3].

Lemma 2.4. *Assume Setting 2.1 and let* $\mathbf{S}\colon [0,\infty) \to L(\mathbf{H}_0)$ *be the function which satisfies for all* $t \in [0,\infty)$, $(v,w) \in \mathbf{H}_0$ *that*

$$\mathbf{S}_t \begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} \cos\big(t(-A)^{1/2}\big)v + (-A)^{-1/2}\sin\big(t(-A)^{1/2}\big)w \\ -(-A)^{1/2}\sin\big(t(-A)^{1/2}\big)v + \cos\big(t(-A)^{1/2}\big)w \end{pmatrix}. \tag{2.25}$$

Then $\mathbf{S}\colon [0,\infty) \to L(\mathbf{H}_0)$ *is a strongly continuous semigroup of bounded linear operators on* $\mathbf{H}_0$ *and* $\mathbf{A}\colon D(\mathbf{A}) \subseteq \mathbf{H}_0 \to \mathbf{H}_0$ *is the generator of* $\mathbf{S}$.

The following two elementary and well-known assertions state that the semigroup in Lemma 2.4 above is a semigroup of isometries and that both this semigroup and its generator commute with Galerkin projections.

Lemma 2.5. *Assume Setting 2.1. Then*

(i) it holds for all $t \in [0,\infty)$, $x \in \mathbf{H}_0$ *that* $\|e^{t\mathbf{A}}x\|_{\mathbf{H}_0} = \|x\|_{\mathbf{H}_0}$ *and*

(ii) it holds that $\sup_{t\in[0,\infty)}\|e^{t\mathbf{A}}\|_{L(\mathbf{H}_0)} = 1$.

Proof of Lemma 2.5. Lemma 2.4 implies for all $t \in [0,\infty)$, $x = (v,w) \in \mathbf{H}_1$ that

$$\begin{aligned}
\|e^{t\mathbf{A}}x\|_{\mathbf{H}_0}^2 &= \big\|\cos\big(t(-A)^{1/2}\big)v + (-A)^{-1/2}\sin\big(t(-A)^{1/2}\big)w\big\|_H^2 \\
&\quad + \big\|-(-A)^{1/2}\sin\big(t(-A)^{1/2}\big)v + \cos\big(t(-A)^{1/2}\big)w\big\|_{H_{-1/2}}^2 \\
&= \big\|\cos\big(t(-A)^{1/2}\big)v\big\|_H^2 + \big\|(-A)^{-1/2}\sin\big(t(-A)^{1/2}\big)w\big\|_H^2 \\
&\quad + \big\|(-A)^{1/2}\sin\big(t(-A)^{1/2}\big)v\big\|_{H_{-1/2}}^2 + \big\|\cos\big(t(-A)^{1/2}\big)w\big\|_{H_{-1/2}}^2 \\
&\quad + 2\big\langle\cos\big(t(-A)^{1/2}\big)v, (-A)^{-1/2}\sin\big(t(-A)^{1/2}\big)w\big\rangle_H \\
&\quad - 2\big\langle\sin\big(t(-A)^{1/2}\big)v, (-A)^{1/2}\cos\big(t(-A)^{1/2}\big)w\big\rangle_{H_{-1/2}} \\
&= \big\|\cos\big(t(-A)^{1/2}\big)v\big\|_H^2 + \big\|\sin\big(t(-A)^{1/2}\big)v\big\|_H^2 \\
&\quad + \big\|\sin\big(t(-A)^{1/2}\big)w\big\|_{H_{-1/2}}^2 + \big\|\cos\big(t(-A)^{1/2}\big)w\big\|_{H_{-1/2}}^2 \\
&= \|v\|_H^2 + \|w\|_{H_{-1/2}}^2 = \|x\|_{\mathbf{H}_0}^2.
\end{aligned} \tag{2.26}$$

This shows (i). In addition, note that (i) implies (ii). The proof of Lemma 2.5 is thus complete. □

Lemma 2.6. *Assume Setting 2.1 and let* $\mathbf{P}_I \in L(\mathbf{H}_0)$, $I \in \mathcal{P}(\mathbb{H})$, *be the linear operators which satisfy for all* $I \in \mathcal{P}(\mathbb{H})$, $v,w \in H$ *that* $\mathbf{P}_I(v,w) = \big(\sum_{h\in I}\langle h,v\rangle_H h, \sum_{h\in I}\langle h,w\rangle_H h\big)$. *Then*

(i) it holds for all $I \in \mathcal{P}(\mathbb{H})$, $x \in \mathbf{H}_1$ *that* $\mathbf{A}\mathbf{P}_I(x) = \mathbf{P}_I\mathbf{A}x$ *and*

(ii) it holds for all $I \in \mathcal{P}(\mathbb{H})$, $t \in [0,\infty)$, $x \in \mathbf{H}_0$ *that* $e^{t\mathbf{A}}\mathbf{P}_I(x) = \mathbf{P}_I e^{t\mathbf{A}}x$.

Proof of Lemma 2.6. Throughout this proof let $P_I \in L(H_{-1/2})$, $I \in \mathcal{P}(\mathbb{H})$, be the linear operators which satisfy for all $I \in \mathcal{P}(\mathbb{H})$, $w \in H_{-1/2}$ that

$$P_I(w) = \sum_{h\in I}\big\langle|\lambda_h|^{1/2}h, w\big\rangle_{H_{-1/2}}|\lambda_h|^{1/2}h. \tag{2.27}$$

Observe that for all $I \in \mathcal{P}(\mathbb{H})$, $x = (v, w) \in \mathbf{H}_1$ it holds that

$$\mathbf{P}_I \mathbf{A} x = \mathbf{P}_I(w, Av) = (P_I(w), P_I A v) = (P_I(w), A P_I(v)) = \mathbf{A} \mathbf{P}_I(x). \tag{2.28}$$

This proves (i). In addition, Lemma 2.4 ensures for all $I \in \mathcal{P}(\mathbb{H})$, $t \in [0, \infty)$, $x = (v, w) \in \mathbf{H}_0$ that

$$\begin{aligned} e^{t\mathbf{A}} \mathbf{P}_I(x) &= \begin{pmatrix} \cos\big(t(-A)^{1/2}\big) P_I(v) + (-A)^{-1/2} \sin\big(t(-A)^{1/2}\big) P_I(w) \\ -(-A)^{1/2} \sin\big(t(-A)^{1/2}\big) P_I(v) + \cos\big(t(-A)^{1/2}\big) P_I(w) \end{pmatrix} \\ &= \begin{pmatrix} P_I\big[\cos\big(t(-A)^{1/2}\big) v + (-A)^{-1/2} \sin\big(t(-A)^{1/2}\big) w\big] \\ P_I\big[-(-A)^{1/2} \sin\big(t(-A)^{1/2}\big) v + \cos\big(t(-A)^{1/2}\big) w\big] \end{pmatrix} \\ &= \mathbf{P}_I e^{t\mathbf{A}} x. \end{aligned} \tag{2.29}$$

This establishes (ii) and thus completes the proof of Lemma 2.6. □

2.2 Upper bounds for weak errors

In this section we establish upper bounds for weak errors associated to spatial spectral Galerkin approximations of semi-linear stochastic wave equations; cf. Theorem 2.12 and Corollaries 2.13, 2.14, 2.16, and 2.18 below.

For many results in this section we consider an abstract setting of wave-type SEEs with appropriate Lipschitz and smoothness assumptions on the corresponding drift non-linearity and diffusion coefficients; cf. the setting in Subsection 2.2.1. In Subsection 2.2.2 we first present a suitable a priori estimate and a suitable perturbation estimate for solutions of certain wave-type SEEs; cf. Lemmas 2.8 and 2.9, respectively. Thereafter, we show an estimate for first and second order spatial derivatives of solutions to Kolmogorov equations associated to certain finite-dimensional wave-type SEEs; cf. Lemma 2.10. Following an elementary auxiliary lemma (cf. Lemma 2.11), we demonstrate the main theorem of this chapter, Theorem 2.12, which provides upper bounds for weak errors involving, among other terms, quantities depending on solutions of certain finite-dimensional wave-type SEEs as well as quantities depending on solutions of Kolmogorov equations associated to these SEEs (cf. also Corollary 2.13). Using the a priori estimate in Lemma 2.8 and the estimate for solutions of Kolmogorov equations in Lemma 2.10, the upper bounds in Theorem 2.12 are subsequently specialised in order to obtain upper bounds depending in an explicit way on the drift non-linearity, the diffusion coefficient, and the initial value; cf. Corollary 2.14.

Finally, in Subsection 2.2.3 we apply Corollary 2.13 to prove essentially sharp weak convergence rates for spatial spectral Galerkin approximations of semi-linear stochastic wave equations. In Corollary 2.16 we consider a setting with specialised drift non-linearity but still quite general diffusion coefficient, while in Corollary 2.18 we consider a class of semi-linear stochastic wave equations driven by multiplicative noise, that includes, in particular, the continuous version of the hyperbolic Anderson model. For the proofs of these results we recall two well-known facts about families of interpolation spaces associated to symmetric diagonal linear operators (cf., e.g., Sell & You [279, Section 3.2]); cf. Lemmas 2.15 and 2.17.

2.2.1 Setting

Setting 2.2. *Consider the notation in Subsection 2.1.1, let* $(H, \langle\cdot,\cdot\rangle_H, \|\cdot\|_H)$ *and* $(U, \langle\cdot,\cdot\rangle_U, \|\cdot\|_U)$ *be separable* $\mathbb{R}$*-Hilbert spaces, let* $\mathbb{H} \subseteq H$ *be a non-empty orthonormal basis of* H*, let* $\lambda\colon \mathbb{H} \to \mathbb{R}$ *be a function which satisfies* $\sup_{h\in\mathbb{H}} \lambda_h < 0$*, let* $A\colon D(A) \subseteq H \to H$ *be the linear operator which satisfies* $D(A) = \big\{v \in H\colon \sum_{h\in\mathbb{H}} |\lambda_h \langle h, v\rangle_H|^2 < \infty\big\}$ *and* $\forall\, v \in D(A)\colon Av = \sum_{h\in\mathbb{H}} \lambda_h \langle h, v\rangle_H h$*, let* $(H_r, \langle\cdot,\cdot\rangle_{H_r}, \|\cdot\|_{H_r})$*,* $r \in \mathbb{R}$*, be a family of interpolation spaces associated to* $-A$*, let* $(\mathbf{H}_r, \langle\cdot,\cdot\rangle_{\mathbf{H}_r}, \|\cdot\|_{\mathbf{H}_r})$*,* $r \in \mathbb{R}$*, be the family of* $\mathbb{R}$*-Hilbert spaces which satisfies for all* $r \in \mathbb{R}$ *that* $(\mathbf{H}_r, \langle\cdot,\cdot\rangle_{\mathbf{H}_r}, \|\cdot\|_{\mathbf{H}_r}) = \big(H_{r/2} \times H_{r/2-1/2}, \langle\cdot,\cdot\rangle_{H_{r/2}\times H_{r/2-1/2}}, \|\cdot\|_{H_{r/2}\times H_{r/2-1/2}}\big)$*, let* $\mathbf{P}_I \in L(\mathbf{H}_0)$*,* $I \in \mathcal{P}(\mathbb{H})$*, be the linear operators which satisfy for all* $I \in \mathcal{P}(\mathbb{H})$*,* $v, w \in H$ *that* $\mathbf{P}_I(v, w) = \big(\sum_{h\in I} \langle h, v\rangle_H h, \sum_{h\in I} \langle h, w\rangle_H h\big)$*, let* $\mathbf{A}\colon D(\mathbf{A}) \subseteq \mathbf{H}_0 \to \mathbf{H}_0$ *be the linear operator which satisfies* $D(\mathbf{A}) = \mathbf{H}_1$ *and* $\forall\, (v, w) \in \mathbf{H}_1\colon \mathbf{A}(v, w) = (w, Av)$*, let* $\mathbf{\Lambda}\colon D(\mathbf{\Lambda}) \subseteq \mathbf{H}_0 \to \mathbf{H}_0$ *be the linear operator which satisfies* $D(\mathbf{\Lambda}) = \mathbf{H}_1$ *and* $\forall\, (v, w) \in \mathbf{H}_1\colon \mathbf{\Lambda}(v, w) = \big(\sum_{h\in\mathbb{H}} |\lambda_h|^{1/2} \langle h, v\rangle_H h, \sum_{h\in\mathbb{H}} |\lambda_h|^{1/2} \langle h, w\rangle_H h\big)$*, let* $T \in (0, \infty)$*, let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space with a normal filtration* $(\mathbb{F}_t)_{t\in[0,T]}$*, let* $(W_t)_{t\in[0,T]}$ *be an* id_U*-cylindrical* $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$*-Wiener process, and let* $\gamma \in (0, \infty)$*,* $\beta \in (\gamma/2, \gamma]$*,* $\rho \in [0, 2(\gamma-\beta)]$*,* $\varrho, C_{\mathbf{F}}, C_{\mathbf{B}} \in [0, \infty)$*,* $\xi \in L^2(\mathbb{P}|_{\mathbb{F}_0}; \mathbf{H}_{2(\gamma-\beta)})$*,* $\mathbf{F} \in \mathrm{Lip}^0(\mathbf{H}_0, \mathbf{H}_0)$*,* $\mathbf{B} \in \mathrm{Lip}^0(\mathbf{H}_0, L_2(U, \mathbf{H}_0))$ *satisfy* $\mathbf{\Lambda}^{-\beta} \in L_2(\mathbf{H}_0)$*,* $\mathbf{F}(\mathbf{H}_\rho) \subseteq \mathbf{H}_{2(\gamma-\beta)}$*,* $(\mathbf{H}_\rho \ni v \mapsto \mathbf{F}(v) \in \mathbf{H}_{2(\gamma-\beta)}) \in \mathrm{Lip}^0(\mathbf{H}_\rho, \mathbf{H}_{2(\gamma-\beta)})$*,* $\forall\, v \in \mathbf{H}_\rho, u \in U\colon \mathbf{B}(v)u \in \mathbf{H}_\gamma$*,* $\forall\, v \in \mathbf{H}_\rho\colon (U \ni u \mapsto \mathbf{B}(v)u \in \mathbf{H}_\rho) \in L_2(U, \mathbf{H}_\rho)$*,* $(\mathbf{H}_\rho \ni v \mapsto (U \ni u \mapsto \mathbf{B}(v)u \in \mathbf{H}_\rho) \in L_2(U, \mathbf{H}_\rho)) \in \mathrm{Lip}^0(\mathbf{H}_\rho, L_2(U, \mathbf{H}_\rho))$*,* $\forall\, v \in \mathbf{H}_\rho\colon (U \ni u \mapsto \mathbf{B}(v)u \in \mathbf{H}_\gamma) \in L(U, \mathbf{H}_\gamma)$*,* $(\mathbf{H}_\rho \ni v \mapsto (U \ni u \mapsto \mathbf{B}(v)u \in \mathbf{H}_\gamma) \in L(U, \mathbf{H}_\gamma)) \in \mathrm{Lip}^0(\mathbf{H}_\rho, L(U, \mathbf{H}_\gamma))$*,* $\mathbf{F}|_{\mathbf{H}_\varrho} \in C^2(\mathbf{H}_\varrho, \mathbf{H}_0)$*,* $\mathbf{B}|_{\mathbf{H}_\varrho} \in C^2(\mathbf{H}_\varrho, L_2(U, \mathbf{H}_0))$*,* $C_{\mathbf{F}} = \sup_{x, v_1, v_2 \in \cap_{r\in\mathbb{R}} \mathbf{H}_r, \max\{\|v_1\|_{\mathbf{H}_0}, \|v_2\|_{\mathbf{H}_0}\} \leq 1} \|\mathbf{F}''(x)(v_1, v_2)\|_{\mathbf{H}_0} < \infty$*, and* $C_{\mathbf{B}} = \sup_{x, v_1, v_2 \in \cap_{r\in\mathbb{R}} \mathbf{H}_r, \max\{\|v_1\|_{\mathbf{H}_0}, \|v_2\|_{\mathbf{H}_0}\} \leq 1} \|\mathbf{B}''(x)(v_1, v_2)\|_{L_2(U, \mathbf{H}_0)} < \infty$*.*

2.2.2 Weak convergence rates for Galerkin approximations

Remark 2.7. Assume Setting 2.2. Then note that the assumption that $(\mathbf{H}_\rho \ni v \mapsto \mathbf{F}(v) \in \mathbf{H}_{2(\gamma-\beta)}) \in \mathrm{Lip}^0(\mathbf{H}_\rho, \mathbf{H}_{2(\gamma-\beta)})$ ensures that $(\mathbf{H}_\rho \ni v \mapsto \mathbf{F}(v) \in \mathbf{H}_\rho) \in \mathrm{Lip}^0(\mathbf{H}_\rho, \mathbf{H}_\rho)$. The assumption that $(\mathbf{H}_\rho \ni v \mapsto (U \ni u \mapsto \mathbf{B}(v)u \in \mathbf{H}_\rho) \in L_2(U, \mathbf{H}_\rho)) \in \mathrm{Lip}^0(\mathbf{H}_\rho, L_2(U, \mathbf{H}_\rho))$ and Proposition 2.1 hence show that there exist up to modifications unique $(\mathbb{F}_t)_{t\in[0,T]}$-predictable stochastic processes $X^I\colon [0, T] \times \Omega \to \mathbf{P}_I(\mathbf{H}_\rho)$, $I \in \mathcal{P}(\mathbb{H})$, such that for all $I \in \mathcal{P}(\mathbb{H})$, $t \in [0, T]$ it holds that $\sup_{s\in[0,T]} \mathbb{E}\big[\|X_s^I\|^2_{\mathbf{H}_\rho}\big] < \infty$ and $\mathbb{P}$-a.s. that

$$X_t^I = e^{t\mathbf{A}} \mathbf{P}_I \xi + \int_0^t e^{(t-s)\mathbf{A}} \mathbf{P}_I \mathbf{F}(X_s^I)\, \mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}} \mathbf{P}_I \mathbf{B}(X_s^I)\, \mathrm{d}W_s. \tag{2.30}$$

The next elementary result, Lemma 2.8, provides global a priori L^2-bounds for the stochastic processes $X^I\colon [0, T] \times \Omega \to \mathbf{P}_I(\mathbf{H}_\rho)$, $I \in \mathcal{P}(\mathbb{H})$, from Remark 2.7.

Lemma 2.8. *Assume Setting 2.2 and let* $X^I\colon [0, T] \times \Omega \to \mathbf{P}_I(\mathbf{H}_\rho)$*,* $I \in \mathcal{P}(\mathbb{H})$*, be* $(\mathbb{F}_t)_{t\in[0,T]}$*-predictable stochastic processes such that for all* $I \in \mathcal{P}(\mathbb{H})$*,* $t \in [0, T]$ *it holds that* $\sup_{s\in[0,T]} \mathbb{E}\big[\|X_s^I\|^2_{\mathbf{H}_\rho}\big] < \infty$ *and* $\mathbb{P}$*-a.s. that*

$$X_t^I = e^{t\mathbf{A}} \mathbf{P}_I \xi + \int_0^t e^{(t-s)\mathbf{A}} \mathbf{P}_I \mathbf{F}(X_s^I)\, \mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}} \mathbf{P}_I \mathbf{B}(X_s^I)\, \mathrm{d}W_s. \tag{2.31}$$

Then

$$\begin{aligned}&\sup_{I\in\mathcal{P}(\mathbb{H})}\sup_{t\in[0,T]}\max\bigl\{1,\mathbb{E}\bigl[\|X_t^I\|_{\mathbf{H}_\rho}^2\bigr]\bigr\}\\&\le\max\bigl\{1,\mathbb{E}\bigl[\|\xi\|_{\mathbf{H}_\rho}^2\bigr]\bigr\}\exp\Bigl(T\Bigl[2\bigl\|\mathbf{F}|_{\mathbf{H}_\rho}\bigr\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,\mathbf{H}_\rho)}+\bigl\|\mathbf{B}|_{\mathbf{H}_\rho}\bigr\|^2_{\mathrm{Lip}^0(\mathbf{H}_\rho,L_2(U,\mathbf{H}_\rho))}\Bigr]\Bigr)<\infty.\end{aligned}\tag{2.32}$$

Proof of Lemma 2.8. Observe that Da Prato, Jentzen, & Röckner [93, Corollary 1], Lemma 2.5, and the Cauchy–Schwarz inequality ensure for all $I\in\mathcal{P}(\mathbb{H})$, $t\in[0,T]$ that

$$\begin{aligned}\mathbb{E}\bigl[\|X_t^I\|_{\mathbf{H}_\rho}^2\bigr]&=\mathbb{E}\bigl[\|e^{t\mathbf{A}}\mathbf{P}_I\xi\|_{\mathbf{H}_\rho}^2\bigr]+2\int_0^t\mathbb{E}\bigl[\bigl\langle e^{(t-s)\mathbf{A}}X_s^I,e^{(t-s)\mathbf{A}}\mathbf{P}_I\mathbf{F}(X_s^I)\bigr\rangle_{\mathbf{H}_\rho}\bigr]\,\mathrm{d}s\\&\quad+\int_0^t\mathbb{E}\bigl[\|e^{(t-s)\mathbf{A}}\mathbf{P}_I\mathbf{B}(X_s^I)\|^2_{L_2(U,\mathbf{H}_\rho)}\bigr]\,\mathrm{d}s\\&\le\mathbb{E}\bigl[\|\mathbf{P}_I\xi\|_{\mathbf{H}_\rho}^2\bigr]+2\int_0^t\|\mathbf{P}_I\mathbf{F}(0)\|_{\mathbf{H}_\rho}\mathbb{E}\bigl[\|X_s^I\|_{\mathbf{H}_\rho}\bigr]+\bigl|\mathbf{P}_I\mathbf{F}|_{\mathbf{H}_\rho}\bigr|_{\mathrm{Lip}^0(\mathbf{H}_\rho,\mathbf{H}_\rho)}\mathbb{E}\bigl[\|X_s^I\|_{\mathbf{H}_\rho}^2\bigr]\,\mathrm{d}s\\&\quad+\int_0^t\|\mathbf{P}_I\mathbf{B}(0)\|^2_{L_2(U,\mathbf{H}_\rho)}+2\|\mathbf{P}_I\mathbf{B}(0)\|_{L_2(U,\mathbf{H}_\rho)}\bigl|\mathbf{P}_I\mathbf{B}|_{\mathbf{H}_\rho}\bigr|_{\mathrm{Lip}^0(\mathbf{H}_\rho,L_2(U,\mathbf{H}_\rho))}\mathbb{E}\bigl[\|X_s^I\|_{\mathbf{H}_\rho}\bigr]\\&\qquad+\bigl|\mathbf{P}_I\mathbf{B}|_{\mathbf{H}_\rho}\bigr|^2_{\mathrm{Lip}^0(\mathbf{H}_\rho,L_2(U,\mathbf{H}_\rho))}\mathbb{E}\bigl[\|X_s^I\|_{\mathbf{H}_\rho}^2\bigr]\,\mathrm{d}s\\&\le\mathbb{E}\bigl[\|\mathbf{P}_I\xi\|_{\mathbf{H}_\rho}^2\bigr]\\&\quad+\Bigl(2\bigl\|\mathbf{P}_I\mathbf{F}|_{\mathbf{H}_\rho}\bigr\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,\mathbf{H}_\rho)}+\bigl\|\mathbf{P}_I\mathbf{B}|_{\mathbf{H}_\rho}\bigr\|^2_{\mathrm{Lip}^0(\mathbf{H}_\rho,L_2(U,\mathbf{H}_\rho))}\Bigr)\int_0^t\max\bigl\{1,\mathbb{E}\bigl[\|X_s^I\|_{\mathbf{H}_\rho}^2\bigr]\bigr\}\,\mathrm{d}s.\end{aligned}\tag{2.33}$$

Gronwall's lemma hence implies for all $I\in\mathcal{P}(\mathbb{H})$ that

$$\begin{aligned}&\sup_{t\in[0,T]}\max\{1,\mathbb{E}\bigl[\|X_t^I\|_{\mathbf{H}_\rho}^2\bigr]\}\\&\le\max\bigl\{1,\mathbb{E}\bigl[\|\mathbf{P}_I\xi\|_{\mathbf{H}_\rho}^2\bigr]\bigr\}\exp\Bigl(T\Bigl[2\bigl\|\mathbf{P}_I\mathbf{F}|_{\mathbf{H}_\rho}\bigr\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,\mathbf{H}_\rho)}+\bigl\|\mathbf{P}_I\mathbf{B}|_{\mathbf{H}_\rho}\bigr\|^2_{\mathrm{Lip}^0(\mathbf{H}_\rho,L_2(U,\mathbf{H}_\rho))}\Bigr]\Bigr)\\&\le\max\bigl\{1,\mathbb{E}\bigl[\|\xi\|_{\mathbf{H}_\rho}^2\bigr]\bigr\}\exp\Bigl(T\Bigl[2\bigl\|\mathbf{F}|_{\mathbf{H}_\rho}\bigr\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,\mathbf{H}_\rho)}+\bigl\|\mathbf{B}|_{\mathbf{H}_\rho}\bigr\|^2_{\mathrm{Lip}^0(\mathbf{H}_\rho,L_2(U,\mathbf{H}_\rho))}\Bigr]\Bigr).\end{aligned}\tag{2.34}$$

The proof of Lemma 2.8 is thus complete. □

In the next result, Lemma 2.9, we present an elementary perturbation estimate.

Lemma 2.9. *Assume Setting 2.2 and let* $X^I\colon[0,T]\times\Omega\to\mathbf{P}_I(\mathbf{H}_0)$, $I\in\mathcal{P}(\mathbb{H})$, *be* $(\mathbb{F}_t)_{t\in[0,T]}$*-predictable stochastic processes such that for all* $I\in\mathcal{P}(\mathbb{H})$, $t\in[0,T]$ *it holds that* $\sup_{s\in[0,T]}\mathbb{E}\bigl[\|X_s^I\|_{\mathbf{H}_0}^2\bigr]<\infty$ *and* $\mathbb{P}$*-a.s. that*

$$X_t^I=e^{t\mathbf{A}}\mathbf{P}_I\xi+\int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_I\mathbf{F}(X_s^I)\,\mathrm{d}s+\int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_I\mathbf{B}(X_s^I)\,\mathrm{d}W_s.\tag{2.35}$$

Then it holds for all $I,J\in\mathcal{P}(\mathbb{H})$ *that*

$$\begin{aligned}&\sup_{t\in[0,T]}\mathbb{E}\bigl[\|X_t^I-X_t^J\|_{\mathbf{H}_0}^2\bigr]\le2\Bigl[\sup_{t\in[0,T]}\mathbb{E}\bigl[\|\mathbf{P}_{I\setminus J}X_t^I+\mathbf{P}_{J\setminus I}X_t^J\|_{\mathbf{H}_0}^2\bigr]\Bigr]\\&\cdot\exp\bigl(\bigl[\sqrt{2}T\,|\mathbf{P}_{I\cap J}\mathbf{F}|_{\mathrm{Lip}^0(\mathbf{H}_0,\mathbf{H}_0)}+\sqrt{2T}\,|\mathbf{P}_{I\cap J}\mathbf{B}|_{\mathrm{Lip}^0(\mathbf{H}_0,L_2(U,\mathbf{H}_0))}\bigr]^2\bigr)<\infty.\end{aligned}\tag{2.36}$$

Proof of Lemma 2.9. Note that Jentzen & Kurniawan [191, Corollary 3.1] and Lemma 2.5 imply for all $I, J \in \mathcal{P}(\mathbb{H})$ that

$$\begin{aligned}
&\sup_{t\in[0,T]} \big\|X_t^I - X_t^J\big\|_{L^2(\mathbb{P};\mathbf{H}_0)} \\
&\leq \sqrt{2}\exp\!\big(\tfrac{1}{2}\big[\sqrt{2}T|\mathbf{P}_{I\cap J}\mathbf{F}|_{\operatorname{Lip}^0(\mathbf{H}_0,\mathbf{H}_0)} + \sqrt{2T}|\mathbf{P}_{I\cap J}\mathbf{B}|_{\operatorname{Lip}^0(\mathbf{H}_0,L_2(U,\mathbf{H}_0))}\big]^2\big) \\
&\quad\cdot \sup_{t\in[0,T]} \bigg\| X_t^I - \bigg[\int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_{I\cap J}\mathbf{F}(X_s^I)\,\mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_{I\cap J}\mathbf{B}(X_s^I)\,\mathrm{d}W_s\bigg] \\
&\qquad\qquad + \bigg[\int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_{I\cap J}\mathbf{F}(X_s^J)\,\mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_{I\cap J}\mathbf{B}(X_s^J)\,\mathrm{d}W_s\bigg] - X_t^J\bigg\|_{L^2(\mathbb{P};\mathbf{H}_0)} \\
&= \sqrt{2}\exp\!\big(\tfrac{1}{2}\big[\sqrt{2}T|\mathbf{P}_{I\cap J}\mathbf{F}|_{\operatorname{Lip}^0(\mathbf{H}_0,\mathbf{H}_0)} + \sqrt{2T}|\mathbf{P}_{I\cap J}\mathbf{B}|_{\operatorname{Lip}^0(\mathbf{H}_0,L_2(U,\mathbf{H}_0))}\big]^2\big) \qquad (2.37)\\
&\quad\cdot \sup_{t\in[0,T]} \bigg\| X_t^I - \mathbf{P}_J\bigg(e^{(t-s)\mathbf{A}}\mathbf{P}_I\xi + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_I\mathbf{F}(X_s^I)\,\mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_I\mathbf{B}(X_s^I)\,\mathrm{d}W_s\bigg) \\
&\quad + \mathbf{P}_I\bigg(e^{(t-s)\mathbf{A}}\mathbf{P}_J\xi + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_J\mathbf{F}(X_s^J)\,\mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_J\mathbf{B}(X_s^J)\,\mathrm{d}W_s\bigg) - X_t^J\bigg\|_{L^2(\mathbb{P};\mathbf{H}_0)} \\
&= \sqrt{2}\exp\!\big(\tfrac{1}{2}\big[\sqrt{2}T|\mathbf{P}_{I\cap J}\mathbf{F}|_{\operatorname{Lip}^0(\mathbf{H}_0,\mathbf{H}_0)} + \sqrt{2T}|\mathbf{P}_{I\cap J}\mathbf{B}|_{\operatorname{Lip}^0(\mathbf{H}_0,L_2(U,\mathbf{H}_0))}\big]^2\big) \\
&\quad\cdot \sup_{t\in[0,T]} \big\|\mathbf{P}_{I\setminus J}X_t^I - \mathbf{P}_{J\setminus I}X_t^J\big\|_{L^2(\mathbb{P};\mathbf{H}_0)}.
\end{aligned}$$

This implies (2.36) and thus completes the proof of Lemma 2.9. □

Lemma 2.10. *Assume Setting 2.2, let $X^{J,x}\colon [0,T]\times\Omega \to \mathbf{P}_J(\mathbf{H}_0)$, $x \in \mathbf{P}_J(\mathbf{H}_0)$, $J \in \mathcal{P}_0(\mathbb{H})$, be $(\mathbb{F}_t)_{t\in[0,T]}$-predictable stochastic processes such that for all $J \in \mathcal{P}_0(\mathbb{H})$, $x \in \mathbf{P}_J(\mathbf{H}_0)$, $t \in [0,T]$ it holds that $\sup_{s\in[0,T]} \mathbb{E}\big[\|X_s^{J,x}\|_{\mathbf{H}_0}^2\big] < \infty$ and $\mathbb{P}$-a.s. that*

$$X_t^{J,x} = e^{t\mathbf{A}}x + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_J\mathbf{F}(X_s^{J,x})\,\mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_J\mathbf{B}(X_s^{J,x})\,\mathrm{d}W_s, \tag{2.38}$$

let $\varphi \in C_{\mathrm{b}}^2(\mathbf{H}_0,\mathbb{R})$, and let $u^J\colon [0,T]\times\mathbf{P}_J(\mathbf{H}_0)\to\mathbb{R}$, $J \in \mathcal{P}_0(\mathbb{H})$, be the functions which satisfy for all $J \in \mathcal{P}_0(\mathbb{H})$, $(t,x) \in [0,T]\times\mathbf{P}_J(\mathbf{H}_0)$ that $u^J(t,x) = \mathbb{E}\big[\varphi\big(X_t^{J,x}\big)\big]$. Then

(i) it holds for all $J \in \mathcal{P}_0(\mathbb{H})$ that $u^J \in C^{1,2}([0,T]\times\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})$,

(ii) it holds that

$$\begin{aligned}
&\sup_{J\in\mathcal{P}_0(\mathbb{H})}\sup_{t\in[0,T]} |u^J(t,\cdot)|_{C_{\mathrm{b}}^1(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})} \\
&\leq |\varphi|_{C_{\mathrm{b}}^1(\mathbf{H}_0,\mathbb{R})}\exp\!\big(T\big[|\mathbf{F}|_{\operatorname{Lip}^0(\mathbf{H}_0,\mathbf{H}_0)} + \tfrac{1}{2}|\mathbf{B}|^2_{\operatorname{Lip}^0(\mathbf{H}_0,L_2(U,\mathbf{H}_0))}\big]\big) < \infty,
\end{aligned} \tag{2.39}$$

and

(iii) it holds that

$$\begin{aligned}
&\sup_{J\in\mathcal{P}_0(\mathbb{H})}\sup_{t\in[0,T]} |u^J(t,\cdot)|_{C_{\mathrm{b}}^2(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})} \leq \|\varphi\|_{C_{\mathrm{b}}^2(\mathbf{H}_0,\mathbb{R})}\max\Big\{1, \big[T\big((C_{\mathbf{F}})^2 + 2\,(C_{\mathbf{B}})^2\big)\big]^{1/2}\Big\} \\
&\quad\cdot\exp\!\big(T\big[\tfrac{1}{2} + 3\,|\mathbf{F}|_{\operatorname{Lip}^0(\mathbf{H}_0,\mathbf{H}_0)} + 4\,|\mathbf{B}|^2_{\operatorname{Lip}^0(\mathbf{H}_0,L_2(U,\mathbf{H}_0))}\big]\big) < \infty.
\end{aligned} \tag{2.40}$$

Proof of Lemma 2.10. Observe that it holds for all $J \in \mathcal{P}_0(\mathbb{H})$ that $\mathbf{P}_J(\mathbf{H}_0) \subseteq \big(\bigcap_{r\in\mathbb{R}} \mathbf{H}_r\big)$ is a finite-dimensional $\mathbb{R}$-vector space. The assumptions that $\mathbf{F}|_{\mathbf{H}_\varrho} \in C^2(\mathbf{H}_\varrho, \mathbf{H}_0)$, $\mathbf{B}|_{\mathbf{H}_\varrho} \in C^2(\mathbf{H}_\varrho, L_2(U, \mathbf{H}_0))$, $\mathbf{F} \in \operatorname{Lip}^0(\mathbf{H}_0, \mathbf{H}_0)$, $\mathbf{B} \in \operatorname{Lip}^0(\mathbf{H}_0, L_2(U, \mathbf{H}_0))$, and $C_\mathbf{F} + C_\mathbf{B} < \infty$ hence ensure for all $J \in \mathcal{P}_0(\mathbb{H})$ that $(\mathbf{P}_J(\mathbf{H}_0) \ni v \mapsto \mathbf{P}_J\mathbf{F}(v) \in \mathbf{P}_J(\mathbf{H}_0)) \in C^2_{\mathrm{b}}(\mathbf{P}_J(\mathbf{H}_0), \mathbf{P}_J(\mathbf{H}_0))$ and $(\mathbf{P}_J(\mathbf{H}_0) \ni v \mapsto (U \ni u \mapsto \mathbf{P}_J\mathbf{B}(v)u \in \mathbf{P}_J(\mathbf{H}_0)) \in L_2(U, \mathbf{P}_J(\mathbf{H}_0))) \in C^2_{\mathrm{b}}(\mathbf{P}_J(\mathbf{H}_0), L_2(U, \mathbf{P}_J(\mathbf{H}_0)))$. Therefore, Lemma 2.2 and Lemma 2.5 prove for all $J \in \mathcal{P}_0(\mathbb{H})$ that

$$u^J \in C^{1,2}([0,T] \times \mathbf{P}_J(\mathbf{H}_0), \mathbb{R}), \tag{2.41}$$

$$\begin{aligned} &\sup_{t\in[0,T]} |u^J(t,\cdot)|_{C^1_{\mathrm{b}}(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})} \leq \big|\varphi|_{\mathbf{P}_J(\mathbf{H}_0)}\big|_{C^1_{\mathrm{b}}(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})} \\ &\quad\cdot \exp\!\Big(T\big[\big|\mathbf{P}_J\mathbf{F}|_{\mathbf{P}_J(\mathbf{H}_0)}\big|_{C^1_{\mathrm{b}}(\mathbf{P}_J(\mathbf{H}_0),\mathbf{P}_J(\mathbf{H}_0))} + \tfrac{1}{2}\big|\mathbf{P}_J\mathbf{B}|_{\mathbf{P}_J(\mathbf{H}_0)}\big|^2_{C^1_{\mathrm{b}}(\mathbf{P}_J(\mathbf{H}_0),L_2(U,\mathbf{P}_J(\mathbf{H}_0)))}\big]\Big), \end{aligned} \tag{2.42}$$

and

$$\begin{aligned} &\sup_{t\in[0,T]} |u^J(t,\cdot)|_{C^2_{\mathrm{b}}(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})} \leq \big\|\varphi|_{\mathbf{P}_J(\mathbf{H}_0)}\big\|_{C^2_{\mathrm{b}}(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})} \\ &\cdot \max\!\Big\{1, \Big[T\Big(\big|\mathbf{P}_J\mathbf{F}|_{\mathbf{P}_J(\mathbf{H}_0)}\big|^2_{C^2_{\mathrm{b}}(\mathbf{P}_J(\mathbf{H}_0),\mathbf{P}_J(\mathbf{H}_0))} + 2\big|\mathbf{P}_J\mathbf{B}|_{\mathbf{P}_J(\mathbf{H}_0)}\big|^2_{C^2_{\mathrm{b}}(\mathbf{P}_J(\mathbf{H}_0),L_2(U,\mathbf{P}_J(\mathbf{H}_0)))}\Big)\Big]^{1/2}\Big\} \\ &\cdot \exp\!\Big(T\big[\tfrac{1}{2} + 3\big|\mathbf{P}_J\mathbf{F}|_{\mathbf{P}_J(\mathbf{H}_0)}\big|_{C^1_{\mathrm{b}}(\mathbf{P}_J(\mathbf{H}_0),\mathbf{P}_J(\mathbf{H}_0))} + 4\big|\mathbf{P}_J\mathbf{B}|_{\mathbf{P}_J(\mathbf{H}_0)}\big|^2_{C^1_{\mathrm{b}}(\mathbf{P}_J(\mathbf{H}_0),L_2(U,\mathbf{P}_J(\mathbf{H}_0)))}\big]\Big) < \infty. \end{aligned} \tag{2.43}$$

This implies (i)–(iii) and thus completes the proof of Lemma 2.10. □

Before we present the main result of this chapter, Theorem 2.12 below, we recall the following elementary and well-known lemma, which is employed in the proof of Theorem 2.12.

Lemma 2.11. *Let $p \in [0,\infty)$, let $\mathcal{J}_n$, $n \in \mathbb{N}_0$, be sets which satisfy for all $n \in \mathbb{N}$ that $\mathcal{J}_n \subseteq \mathcal{J}_{n+1}$ and $\bigcup_{m=1}^\infty \mathcal{J}_m = \mathcal{J}_0$, and let $g\colon \mathcal{J}_0 \to (0,\infty)$ be a function which satisfies $\sum_{h\in\mathcal{J}_0} (g_h)^p < \infty$. Then*

$$\limsup_{n\to\infty} \sup\!\big(\{g_h \colon h \in \mathcal{J}_0 \setminus \mathcal{J}_n\} \cup \{0\}\big) = 0. \tag{2.44}$$

Proof of Lemma 2.11. Without loss of generality we assume that $p \in (0,\infty)$ (otherwise (2.44) is clear). Observe that for all $n \in \mathbb{N}$ it holds that

$$\big[\sup\!\big(\{g_h \colon h \in \mathcal{J}_0 \setminus \mathcal{J}_n\} \cup \{0\}\big)\big]^p \leq \sum_{h\in\mathcal{J}_0\setminus\mathcal{J}_n} (g_h)^p = \sum_{h\in\mathcal{J}_0} (g_h)^p - \sum_{h\in\mathcal{J}_n} (g_h)^p. \tag{2.45}$$

Moreover, note that Lebesgue's theorem of dominated convergence proves that

$$\limsup_{n\to\infty} \left[\sum_{h\in\mathcal{J}_0} (g_h)^p - \sum_{h\in\mathcal{J}_n} (g_h)^p\right] = 0. \tag{2.46}$$

Combining this with (2.45) completes the proof of Lemma 2.11. □

Theorem 2.12. *Assume Setting 2.2, let $X^I\colon [0,T] \times \Omega \to \mathbf{P}_I(\mathbf{H}_\rho)$, $I \in \mathcal{P}(\mathbb{H})$, and $X^{J,x}\colon [0,T] \times \Omega \to \mathbf{P}_J(\mathbf{H}_0)$, $x \in \mathbf{P}_J(\mathbf{H}_0)$, $J \in \mathcal{P}_0(\mathbb{H})$, be $(\mathbb{F}_t)_{t\in[0,T]}$-predictable stochastic* *processes such that for all* $I \in \mathcal{P}(\mathbb{H})$, $J \in \mathcal{P}_0(\mathbb{H})$, $x \in \mathbf{P}_J(\mathbf{H}_0)$, $t \in [0,T]$ *it holds that* $\sup_{s\in[0,T]} \mathbb{E}\big[\|X_s^I\|_{\mathbf{H}_\rho}^2 + \|X_s^{J,x}\|_{\mathbf{H}_0}^2\big] < \infty$ *and* $\mathbb{P}$*-a.s. that*

$$X_t^I = e^{t\mathbf{A}}\mathbf{P}_I\xi + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_I\mathbf{F}(X_s^I)\,\mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_I\mathbf{B}(X_s^I)\,\mathrm{d}W_s, \tag{2.47}$$

$$X_t^{J,x} = e^{t\mathbf{A}}x + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_J\mathbf{F}(X_s^{J,x})\,\mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_J\mathbf{B}(X_s^{J,x})\,\mathrm{d}W_s, \tag{2.48}$$

let $\varphi \in C_{\mathrm{b}}^2(\mathbf{H}_0,\mathbb{R})$, *and let* $u^J\colon [0,T]\times\mathbf{P}_J(\mathbf{H}_0)\to\mathbb{R}$, $J\in\mathcal{P}_0(\mathbb{H})$, *be the functions which satisfy for all* $J\in\mathcal{P}_0(\mathbb{H})$, $(t,x)\in[0,T]\times\mathbf{P}_J(\mathbf{H}_0)$ *that* $u^J(t,x) = \mathbb{E}\big[\varphi\big(X_t^{J,x}\big)\big]$. *Then it holds for all* $I \in \mathcal{P}(\mathbb{H})\setminus\{\mathbb{H}\}$ *that*

$$\begin{aligned}
&\big|\mathbb{E}\big[\varphi\big(X_T^{\mathbb{H}}\big)\big] - \mathbb{E}\big[\varphi\big(X_T^I\big)\big]\big| \\
&\leq \Bigg(\Bigg[\sup_{J\in\mathcal{P}_0(\mathbb{H})}\sup_{t\in[0,T]}\big|u^J(t,\cdot)\big|_{C_{\mathrm{b}}^1(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})}\Bigg]\Bigg[\mathbb{E}\big[\|\xi\|_{\mathbf{H}_{2(\gamma-\beta)}}\big] + \sup_{J\in\mathcal{P}_0(\mathbb{H})}\int_0^T \mathbb{E}\big[\big\|\mathbf{F}(X_s^J)\big\|_{\mathbf{H}_{2(\gamma-\beta)}}\big]\,\mathrm{d}s\Bigg] \\
&\quad + \|\mathbf{\Lambda}^{-\beta}\|_{L_2(\mathbf{H}_0)}^2\Bigg[\sup_{J\in\mathcal{P}_0(\mathbb{H})}\sup_{t\in[0,T]}\big|u^J(t,\cdot)\big|_{C_{\mathrm{b}}^2(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})}\Bigg]\sup_{J\in\mathcal{P}_0(\mathbb{H})}\int_0^T\mathbb{E}\big[\big\|\mathbf{B}(X_s^J)\big\|_{L(U,\mathbf{H}_\gamma)}^2\big]\,\mathrm{d}s\Bigg) \\
&\quad \cdot \Bigg[\inf_{h\in\mathbb{H}\setminus I}|\lambda_h|\Bigg]^{\beta-\gamma} < \infty.
\end{aligned} \tag{2.49}$$

Proof of Theorem 2.12. Throughout this proof let $\mathbb{U}\subseteq U$ be an orthonormal basis of U, let $v^J, v_{1,0}^J\colon [0,T]\times\mathbf{P}_J(\mathbf{H}_0)\to\mathbb{R}$, $J\in\mathcal{P}_0(\mathbb{H})$, and $v_{0,\ell}^J\colon[0,T]\times\mathbf{P}_J(\mathbf{H}_0)\to L^{(\ell)}(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})$, $\ell\in\{1,2\}$, $J\in\mathcal{P}_0(\mathbb{H})$, be the functions which satisfy for all $J\in\mathcal{P}_0(\mathbb{H})$, $(k,\ell)\in\{(1,0),(0,1),(0,2)\}$, $(t,x)\in[0,T]\times\mathbf{P}_J(\mathbf{H}_0)$ that $v^J(t,x)=\mathbb{E}\big[\varphi\big(X_{T-t}^{J,x}\big)\big]$ and $v_{k,\ell}^J(t,x) = \big(\frac{\partial^{k+\ell}}{\partial t^k\partial x^\ell}v^J\big)(t,x)$, and let $R_{I,J,s}\colon\Omega\to L(\mathbf{P}_J(\mathbf{H}_0))$, $I\in\mathcal{P}(J)$, $J\in\mathcal{P}_0(\mathbb{H})$, $s\in[0,T]$, be the functions which satisfy for all $s\in[0,T]$, $J\in\mathcal{P}_0(\mathbb{H})$, $I\in\mathcal{P}(J)$, $\omega\in\Omega$, $y_1,y_2\in\mathbf{P}_J(\mathbf{H}_0)$ that

$$v_{0,2}^J\big(s,X_s^I(\omega)\big)(y_1,y_2) = \langle y_1, R_{I,J,s}(\omega)\,y_2\rangle_{\mathbf{H}_0}. \tag{2.50}$$

Note that for all $J\in\mathcal{P}_0(\mathbb{H})$, $(t,x)\in[0,T]\times\mathbf{P}_J(\mathbf{H}_0)$ it holds that $v^J(t,x) = u^J(T-t,x)$. Next observe that for all $J\in\mathcal{P}_0(\mathbb{H})$, $x\in\mathbf{P}_J(\mathbf{H}_0)$ it holds that

$$\varphi(x) = \mathbb{E}[\varphi(x)] = u^J(0,x) = v^J(T,x). \tag{2.51}$$

Moreover, note that for all $J\in\mathcal{P}_0(\mathbb{H})$ it holds that

$$\mathbb{E}\big[\varphi\big(X_T^J\big)\big] = \mathbb{E}\big[u^J\big(T,X_0^J\big)\big] = \mathbb{E}\big[v^J\big(0,X_0^J\big)\big]. \tag{2.52}$$

Combining (2.51) and (2.52) shows for all $J\in\mathcal{P}_0(\mathbb{H})$, $I\in\mathcal{P}(J)$ that

$$\begin{aligned}
&\big|\mathbb{E}\big[\varphi\big(X_T^J\big)\big] - \mathbb{E}\big[\varphi\big(X_T^I\big)\big]\big| = \big|\mathbb{E}\big[\varphi\big(X_T^I\big)\big] - \mathbb{E}\big[\varphi\big(X_T^J\big)\big]\big| \\
&= \big|\mathbb{E}\big[v^J\big(T,X_T^I\big)\big] - \mathbb{E}\big[v^J\big(0,X_0^J\big)\big]\big| \\
&\leq \big|\mathbb{E}\big[v^J\big(T,X_T^I\big)\big] - \mathbb{E}\big[v^J\big(0,X_0^I\big)\big]\big| + \big|\mathbb{E}\big[v^J\big(0,X_0^I\big)\big] - \mathbb{E}\big[v^J\big(0,X_0^J\big)\big]\big|.
\end{aligned} \tag{2.53}$$

In a first step we establish an estimate for the second summand on the right hand side of (2.53). For this observe that for all $J \in \mathcal{P}_0(\mathbb{H})$, $I \in \mathcal{P}(J)$ it holds that

$$\begin{aligned}
&\big|\mathbb{E}\big[v^J\big(0, X_0^I\big)\big] - \mathbb{E}\big[v^J\big(0, X_0^J\big)\big]\big| \\
&= \left|\mathbb{E}\left[\int_0^1 v_{0,1}^J\big(0, X_0^I + \tau\big(X_0^J - X_0^I\big)\big)\big(X_0^J - X_0^I\big)\,\mathrm{d}\tau\right]\right| \\
&\leq \big|u^J(T,\cdot)\big|_{C_{\mathrm{b}}^1(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})}\mathbb{E}\big[\big\|X_0^J - X_0^I\big\|_{\mathbf{H}_0}\big] \\
&= \big|u^J(T,\cdot)\big|_{C_{\mathrm{b}}^1(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})}\mathbb{E}\big[\big\|\mathbf{P}_I\big(X_0^J\big) - \mathbf{P}_J\big(X_0^J\big)\big\|_{\mathbf{H}_0}\big].
\end{aligned} \tag{2.54}$$

In addition, it holds for all $x \in \mathbf{H}_{2(\gamma-\beta)}$, $I, J \in \mathcal{P}(\mathbb{H})$ with $I \neq J$ that

$$\begin{aligned}
\|\mathbf{P}_I(x) - \mathbf{P}_J(x)\|_{\mathbf{H}_0} &\leq \|\mathbf{\Lambda}^{2(\beta-\gamma)}\mathbf{P}_{(I\setminus J)\cup(J\setminus I)}\|_{L(\mathbf{H}_0)}\|\mathbf{P}_{(I\setminus J)\cup(J\setminus I)}(x)\|_{\mathbf{H}_{2(\gamma-\beta)}} \\
&= \left[\inf_{h\in(I\setminus J)\cup(J\setminus I)}|\lambda_h|\right]^{\beta-\gamma}\|\mathbf{P}_{(I\setminus J)\cup(J\setminus I)}(x)\|_{\mathbf{H}_{2(\gamma-\beta)}} \\
&\leq \left[\inf_{h\in(I\setminus J)\cup(J\setminus I)}|\lambda_h|\right]^{\beta-\gamma}\|x\|_{\mathbf{H}_{2(\gamma-\beta)}}.
\end{aligned} \tag{2.55}$$

Putting (2.54) and (2.55) together proves for all $J \in \mathcal{P}_0(\mathbb{H})$, $I \in \mathcal{P}(J) \setminus \{\mathbb{H}\}$ that

$$\begin{aligned}
&\big|\mathbb{E}\big[v^J\big(0, X_0^I\big)\big] - \mathbb{E}\big[v^J\big(0, X_0^J\big)\big]\big| \\
&\leq \left[\sup_{K\in\mathcal{P}_0(\mathbb{H})}\sup_{t\in[0,T]}\big|u^K(t,\cdot)\big|_{C_{\mathrm{b}}^1(\mathbf{P}_K(\mathbf{H}_0),\mathbb{R})}\right]\mathbb{E}\big[\|\xi\|_{\mathbf{H}_{2(\gamma-\beta)}}\big]\left[\inf_{h\in\mathbb{H}\setminus I}|\lambda_h|\right]^{\beta-\gamma} < \infty.
\end{aligned} \tag{2.56}$$

Inequality (2.56) provides an estimate for the second summand on the right hand side of (2.53). In a second step we establish an estimate for the fist summand on the right hand side of (2.53). The chain rule and Lemma 2.2 show that for all $J \in \mathcal{P}_0(\mathbb{H})$, $(t,x) \in [0,T] \times \mathbf{P}_J(\mathbf{H}_0)$ it holds that

$$v_{1,0}^J(t,x) = -v_{0,1}^J(t,x)\big[\mathbf{A}x + \mathbf{P}_J\mathbf{F}(x)\big] - \tfrac{1}{2}\sum_{u\in\mathbb{U}} v_{0,2}^J(t,x)(\mathbf{P}_J\mathbf{B}(x)u, \mathbf{P}_J\mathbf{B}(x)u). \tag{2.57}$$

The standard Itô formula and (2.57) prove for all $J \in \mathcal{P}_0(\mathbb{H})$, $I \in \mathcal{P}(J)$ that

$$\begin{aligned}
&\mathbb{E}\big[v^J\big(T, X_T^I\big)\big] - \mathbb{E}\big[v^J\big(0, X_0^I\big)\big] = \int_0^T \mathbb{E}\big[v_{1,0}^J\big(s, X_s^I\big)\big]\,\mathrm{d}s + \int_0^T \mathbb{E}\big[v_{0,1}^J\big(s, X_s^I\big)\mathbf{A}X_s^I\big]\,\mathrm{d}s \\
&\quad + \int_0^T \mathbb{E}\big[v_{0,1}^J\big(s, X_s^I\big)\mathbf{P}_I\mathbf{F}(X_s^I)\big]\,\mathrm{d}s + \tfrac{1}{2}\sum_{b\in\mathbb{U}}\int_0^T \mathbb{E}\big[v_{0,2}^J\big(s, X_s^I\big)\big(\mathbf{P}_I\mathbf{B}(X_s^I)b, \mathbf{P}_I\mathbf{B}(X_s^I)b\big)\big]\,\mathrm{d}s \\
&= \int_0^T \mathbb{E}\big[v_{0,1}^J\big(s, X_s^I\big)\mathbf{P}_I\mathbf{F}(X_s^I)\big]\,\mathrm{d}s - \int_0^T \mathbb{E}\big[v_{0,1}^J\big(s, X_s^I\big)\mathbf{P}_J\mathbf{F}(X_s^I)\big]\,\mathrm{d}s \\
&\quad + \tfrac{1}{2}\sum_{b\in\mathbb{U}}\int_0^T \Big(\mathbb{E}\big[v_{0,2}^J\big(s, X_s^I\big)\big(\mathbf{P}_I\mathbf{B}(X_s^I)b, \mathbf{P}_I\mathbf{B}(X_s^I)b\big)\big] \\
&\qquad\qquad - \mathbb{E}\big[v_{0,2}^J\big(s, X_s^I\big)\big(\mathbf{P}_J\mathbf{B}(X_s^I)b, \mathbf{P}_J\mathbf{B}(X_s^I)b\big)\big]\Big)\,\mathrm{d}s.
\end{aligned} \tag{2.58}$$

This shows for all $J \in \mathcal{P}_0(\mathbb{H})$, $I \in \mathcal{P}(J)$ that

$$\begin{aligned}
&\big|\mathbb{E}\big[v^J\big(T, X_T^I\big)\big] - \mathbb{E}\big[v^J\big(0, X_0^I\big)\big]\big| \leq \int_0^T \big|\mathbb{E}\big[v_{0,1}^J\big(s, X_s^I\big)\big(\mathbf{P}_I\mathbf{F}(X_s^I) - \mathbf{P}_J\mathbf{F}(X_s^I)\big)\big]\big|\,\mathrm{d}s \\
&\quad + \left|\tfrac{1}{2}\sum_{b\in\mathbb{U}}\int_0^T \mathbb{E}\big[v_{0,2}^J\big(s, X_s^I\big)\big(\mathbf{P}_I\mathbf{B}(X_s^I)b + \mathbf{P}_J\mathbf{B}(X_s^I)b, \mathbf{P}_I\mathbf{B}(X_s^I)b - \mathbf{P}_J\mathbf{B}(X_s^I)b\big)\big]\,\mathrm{d}s\right|.
\end{aligned} \tag{2.59}$$

Inequality (2.55), Lemma 2.10, and Lemma 2.8 thus prove for all $J \in \mathcal{P}_0(\mathbb{H})$, $I \in \mathcal{P}(J) \setminus \{J\}$ that

$$\begin{aligned}
&\int_0^T \left|\mathbb{E}\left[v_{0,1}^J\left(s, X_s^I\right)\left(\mathbf{P}_I\mathbf{F}(X_s^I) - \mathbf{P}_J\mathbf{F}(X_s^I)\right)\right]\right| \mathrm{d}s \\
&\leq \int_0^T \mathbb{E}\left[\left|v_{0,1}^J\left(s, X_s^I\right)\left(\mathbf{P}_I\mathbf{F}(X_s^I) - \mathbf{P}_J\mathbf{F}(X_s^I)\right)\right|\right] \mathrm{d}s \\
&\leq \sup_{t\in[0,T]} \left|u^J(t,\cdot)\right|_{C_{\mathrm{b}}^1(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})} \int_0^T \mathbb{E}\left[\left\|\mathbf{P}_I\mathbf{F}(X_s^I) - \mathbf{P}_J\mathbf{F}(X_s^I)\right\|_{\mathbf{H}_0}\right] \mathrm{d}s \qquad (2.60)\\
&\leq \left[\sup_{K\in\mathcal{P}_0(\mathbb{H})} \sup_{t\in[0,T]} \left|u^K(t,\cdot)\right|_{C_{\mathrm{b}}^1(\mathbf{P}_K(\mathbf{H}_0),\mathbb{R})}\right] \sup_{K\in\mathcal{P}_0(\mathbb{H})} \int_0^T \mathbb{E}\left[\left\|\mathbf{F}(X_s^K)\right\|_{\mathbf{H}_{2(\gamma-\beta)}}\right] \mathrm{d}s \\
&\quad \cdot \left[\inf_{h\in J\setminus I} |\lambda_h|\right]^{\beta-\gamma} < \infty.
\end{aligned}$$

This estimates the first summand on the right hand side of (2.59). Next we consider the second summand on the right hand side of (2.59). Note that the Hölder inequality for Schatten norms implies for all $s \in [0,T]$, $J \in \mathcal{P}_0(\mathbb{H})$, $I \in \mathcal{P}(J)$ that

$$\begin{aligned}
&\left\|\mathbf{B}(X_s^I)^\star(\mathbf{P}_I + \mathbf{P}_J)R_{I,J,s}(\mathbf{P}_I - \mathbf{P}_J)\mathbf{B}(X_s^I)\right\|_{L_1(U)} \qquad (2.61)\\
&\leq \left\|\mathbf{B}(X_s^I)^\star(\mathbf{P}_I + \mathbf{P}_J)\right\|_{L_{(2\beta)/\gamma}(\mathbf{H}_0,U)} \|R_{I,J,s}\|_{L(\mathbf{P}_J(\mathbf{H}_0))} \left\|(\mathbf{P}_I - \mathbf{P}_J)\mathbf{B}(X_s^I)\right\|_{L_{(2\beta)/(2\beta-\gamma)}(U,\mathbf{H}_0)}.
\end{aligned}$$

Moreover, observe that for all $s \in [0,T]$, $J \in \mathcal{P}_0(\mathbb{H})$, $I \in \mathcal{P}(J) \setminus \{J\}$ it holds that

$$\begin{aligned}
&\left\|\mathbf{B}(X_s^I)^\star(\mathbf{P}_I + \mathbf{P}_J)\right\|_{L_{(2\beta)/\gamma}(\mathbf{H}_0,U)} = \left\|\mathbf{B}(X_s^I)^\star\mathbf{\Lambda}^\gamma\mathbf{\Lambda}^{-\gamma}(\mathbf{P}_I + \mathbf{P}_J)\right\|_{L_{(2\beta)/\gamma}(\mathbf{H}_0,U)} \\
&\leq \left\|\mathbf{B}(X_s^I)^\star\mathbf{\Lambda}^\gamma\right\|_{L(\mathbf{H}_0,U)} \|\mathbf{\Lambda}^{-\gamma}\|_{L_{(2\beta)/\gamma}(\mathbf{H}_0)} \|\mathbf{P}_I + \mathbf{P}_J\|_{L(\mathbf{H}_0)} \qquad (2.62)\\
&= \left\|\mathbf{B}(X_s^I)\right\|_{L(U,\mathbf{H}_\gamma)} \|\mathbf{\Lambda}^{-\beta}\|_{L_2(\mathbf{H}_0)}^{\gamma/\beta} \|\mathbf{P}_I + \mathbf{P}_J\|_{L(\mathbf{H}_0)} \leq 2\left\|\mathbf{B}(X_s^I)\right\|_{L(U,\mathbf{H}_\gamma)} \|\mathbf{\Lambda}^{-\beta}\|_{L_2(\mathbf{H}_0)}^{\gamma/\beta} < \infty
\end{aligned}$$

and

$$\begin{aligned}
&\left\|(\mathbf{P}_I - \mathbf{P}_J)\mathbf{B}(X_s^I)\right\|_{L_{(2\beta)/(2\beta-\gamma)}(U,\mathbf{H}_0)} \leq \|(\mathbf{P}_I - \mathbf{P}_J)|_{\mathbf{H}_\gamma}\|_{L_{(2\beta)/(2\beta-\gamma)}(\mathbf{H}_\gamma,\mathbf{H}_0)} \left\|\mathbf{B}(X_s^I)\right\|_{L(U,\mathbf{H}_\gamma)} \\
&\leq \|(\mathbf{P}_I - \mathbf{P}_J)\mathbf{\Lambda}^{2(\beta-\gamma)}\|_{L(\mathbf{H}_0)} \|\mathbf{\Lambda}^{2(\gamma-\beta)}|_{\mathbf{H}_\gamma}\|_{L_{(2\beta)/(2\beta-\gamma)}(\mathbf{H}_\gamma,\mathbf{H}_0)} \left\|\mathbf{B}(X_s^I)\right\|_{L(U,\mathbf{H}_\gamma)} \qquad (2.63)\\
&= \left[\inf_{h\in J\setminus I} |\lambda_h|\right]^{\beta-\gamma} \|\mathbf{\Lambda}^{-\beta}\|_{L_2(\mathbf{H}_0)}^{(2\beta-\gamma)/\beta} \left\|\mathbf{B}(X_s^I)\right\|_{L(U,\mathbf{H}_\gamma)} < \infty.
\end{aligned}$$

In addition, Lemma 2.10 establishes for all $s \in [0,T]$, $J \in \mathcal{P}_0(\mathbb{H})$, $I \in \mathcal{P}(J)$ that

$$\|R_{I,J,s}\|_{L(\mathbf{P}_J(\mathbf{H}_0))} \leq \sup_{t\in[0,T]} \left|u^J(t,\cdot)\right|_{C_{\mathrm{b}}^2(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})} < \infty. \qquad (2.64)$$

Combining (2.61)–(2.64) shows for all $s \in [0,T]$, $J \in \mathcal{P}_0(\mathbb{H})$, $I \in \mathcal{P}(J) \setminus \{J\}$ that

$$\begin{aligned}
&\left\|\mathbf{B}(X_s^I)^\star(\mathbf{P}_I + \mathbf{P}_J)R_{I,J,s}(\mathbf{P}_I - \mathbf{P}_J)\mathbf{B}(X_s^I)\right\|_{L_1(U)} \\
&\leq 2\|\mathbf{\Lambda}^{-\beta}\|_{L_2(\mathbf{H}_0)}^2 \left[\sup_{t\in[0,T]} \left|u^J(t,\cdot)\right|_{C_{\mathrm{b}}^2(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})}\right] \left\|\mathbf{B}(X_s^I)\right\|_{L(U,\mathbf{H}_\gamma)}^2 \left[\inf_{h\in J\setminus I} |\lambda_h|\right]^{\beta-\gamma} < \infty. \qquad (2.65)
\end{aligned}$$

This and (2.50) imply for all $s \in [0,T]$, $J \in \mathcal{P}_0(\mathbb{H})$, $I \in \mathcal{P}(J) \setminus \{J\}$ that

$$\begin{aligned}
&\Big| \sum_{b \in \mathbb{U}} \mathbb{E}\big[v^J_{0,2}\big(s, X^I_s\big)\big((\mathbf{P}_I + \mathbf{P}_J)\mathbf{B}(X^I_s)b, (\mathbf{P}_I - \mathbf{P}_J)\mathbf{B}(X^I_s)b\big)\big] \Big| \\
&= \Big| \mathbb{E}\Big[\sum_{b \in \mathbb{U}} \langle (\mathbf{P}_I + \mathbf{P}_J)\mathbf{B}(X^I_s)b, R_{I,J,s}(\mathbf{P}_I - \mathbf{P}_J)\mathbf{B}(X^I_s)b \rangle_{\mathbf{H}_0} \Big] \Big| \\
&= \Big| \mathbb{E}\Big[\sum_{b \in \mathbb{U}} \langle b, \mathbf{B}(X^I_s)^\star (\mathbf{P}_I + \mathbf{P}_J) R_{I,J,s}(\mathbf{P}_I - \mathbf{P}_J)\mathbf{B}(X^I_s)b \rangle_U \Big] \Big| \\
&= \big| \mathbb{E}\big[\operatorname{trace}_U\big(\mathbf{B}(X^I_s)^\star (\mathbf{P}_I + \mathbf{P}_J) R_{I,J,s}(\mathbf{P}_I - \mathbf{P}_J)\mathbf{B}(X^I_s)\big)\big] \big| \\
&\leq \mathbb{E}\big[\big\| \mathbf{B}(X^I_s)^\star (\mathbf{P}_I + \mathbf{P}_J) R_{I,J,s}(\mathbf{P}_I - \mathbf{P}_J)\mathbf{B}(X^I_s) \big\|_{L_1(U)} \big] \\
&\leq 2 \, \|\mathbf{\Lambda}^{-\beta}\|^2_{L_2(\mathbf{H}_0)} \Big[\sup_{t \in [0,T]} |u^J(t,\cdot)|_{C^2_{\mathrm{b}}(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})} \Big] \mathbb{E}\big[\|\mathbf{B}(X^I_s)\|^2_{L(U,\mathbf{H}_\gamma)} \big] \Big[\inf_{h \in J \setminus I} |\lambda_h| \Big]^{\beta-\gamma}.
\end{aligned} \tag{2.66}$$

Lemma 2.8 and Lemma 2.10 hence prove for all $J \in \mathcal{P}_0(\mathbb{H})$, $I \in \mathcal{P}(J) \setminus \{J\}$ that

$$\begin{aligned}
&\Big| \tfrac{1}{2} \sum_{b \in \mathbb{U}} \int_0^T \mathbb{E}\big[v^J_{0,2}\big(s, X^I_s\big)\big(\mathbf{P}_I\mathbf{B}(X^I_s)b + \mathbf{P}_J\mathbf{B}(X^I_s)b, \mathbf{P}_I\mathbf{B}(X^I_s)b - \mathbf{P}_J\mathbf{B}(X^I_s)b\big)\big] \, \mathrm{d}s \Big| \\
&\leq \|\mathbf{\Lambda}^{-\beta}\|^2_{L_2(\mathbf{H}_0)} \Big[\sup_{K \in \mathcal{P}_0(\mathbb{H})} \sup_{t \in [0,T]} |u^K(t,\cdot)|_{C^2_{\mathrm{b}}(\mathbf{P}_K(\mathbf{H}_0),\mathbb{R})} \Big] \sup_{K \in \mathcal{P}_0(\mathbb{H})} \int_0^T \mathbb{E}\big[\|\mathbf{B}(X^K_s)\|^2_{L(U,\mathbf{H}_\gamma)} \big] \, \mathrm{d}s \\
&\quad \cdot \Big[\inf_{h \in J \setminus I} |\lambda_h| \Big]^{\beta-\gamma} < \infty.
\end{aligned} \tag{2.67}$$

Combining this with (2.59) and (2.60) ensures for all $J \in \mathcal{P}_0(\mathbb{H})$, $I \in \mathcal{P}(J) \setminus \{\mathbb{H}\}$ that

$$\begin{aligned}
&\big| \mathbb{E}\big[v^J\big(T, X^I_T\big)\big] - \mathbb{E}\big[v^J\big(0, X^I_0\big)\big] \big| \\
&\leq \bigg(\Big[\sup_{K \in \mathcal{P}_0(\mathbb{H})} \sup_{t \in [0,T]} |u^K(t,\cdot)|_{C^1_{\mathrm{b}}(\mathbf{P}_K(\mathbf{H}_0),\mathbb{R})} \Big] \sup_{K \in \mathcal{P}_0(\mathbb{H})} \int_0^T \mathbb{E}\big[\|\mathbf{F}(X^K_s)\|_{\mathbf{H}_{2(\gamma-\beta)}} \big] \, \mathrm{d}s \\
&\quad + \|\mathbf{\Lambda}^{-\beta}\|^2_{L_2(\mathbf{H}_0)} \Big[\sup_{K \in \mathcal{P}_0(\mathbb{H})} \sup_{t \in [0,T]} |u^K(t,\cdot)|_{C^2_{\mathrm{b}}(\mathbf{P}_K(\mathbf{H}_0),\mathbb{R})} \Big] \sup_{K \in \mathcal{P}_0(\mathbb{H})} \int_0^T \mathbb{E}\big[\|\mathbf{B}(X^K_s)\|^2_{L(U,\mathbf{H}_\gamma)} \big] \, \mathrm{d}s \bigg) \\
&\quad \cdot \Big[\inf_{h \in \mathbb{H} \setminus I} |\lambda_h| \Big]^{\beta-\gamma} < \infty.
\end{aligned} \tag{2.68}$$

This constitutes an estimate for the first summand on the right hand side of (2.53). Inequalities (2.68), (2.53), and (2.56) show for all $J \in \mathcal{P}_0(\mathbb{H})$, $I \in \mathcal{P}(J) \setminus \{\mathbb{H}\}$ that

$$\begin{aligned}
&\big| \mathbb{E}\big[\varphi\big(X^J_T\big)\big] - \mathbb{E}\big[\varphi\big(X^I_T\big)\big] \big| \\
&\leq \bigg(\Big[\sup_{K \in \mathcal{P}_0(\mathbb{H})} \sup_{t \in [0,T]} |u^K(t,\cdot)|_{C^1_{\mathrm{b}}(\mathbf{P}_K(\mathbf{H}_0),\mathbb{R})} \Big] \Big[\mathbb{E}\big[\|\xi\|_{\mathbf{H}_{2(\gamma-\beta)}} \big] + \sup_{K \in \mathcal{P}_0(\mathbb{H})} \int_0^T \mathbb{E}\big[\|\mathbf{F}(X^K_s)\|_{\mathbf{H}_{2(\gamma-\beta)}} \big] \, \mathrm{d}s \Big] \\
&\quad + \|\mathbf{\Lambda}^{-\beta}\|^2_{L_2(\mathbf{H}_0)} \Big[\sup_{K \in \mathcal{P}_0(\mathbb{H})} \sup_{t \in [0,T]} |u^K(t,\cdot)|_{C^2_{\mathrm{b}}(\mathbf{P}_K(\mathbf{H}_0),\mathbb{R})} \Big] \sup_{K \in \mathcal{P}_0(\mathbb{H})} \int_0^T \mathbb{E}\big[\|\mathbf{B}(X^K_s)\|^2_{L(U,\mathbf{H}_\gamma)} \big] \, \mathrm{d}s \bigg) \\
&\quad \cdot \Big[\inf_{h \in \mathbb{H} \setminus I} |\lambda_h| \Big]^{\beta-\gamma} < \infty.
\end{aligned} \tag{2.69}$$

In a third step Lemma 2.9, Lemma 2.5, Lemma 2.6, the Cauchy–Schwarz inequality, and the Burkholder–Davis–Gundy-type inequality in Da Prato & Zabczyk [94, Lemma 7.7] imply for all $n \in \mathbb{N}$, $(J_k)_{k\in\mathbb{N}_0} \subseteq \mathcal{P}(\mathbb{H})$ with $\bigcup_{k=1}^{\infty} J_k = J_0$ and $\forall k \in \mathbb{N}\colon J_k \subseteq J_{k+1} \in \mathcal{P}_0(\mathbb{H})$ that

$$
\begin{aligned}
&\sup_{t\in[0,T]} \big(\mathbb{E}\big[\big\|X_t^{J_0} - X_t^{J_n}\big\|_{\mathbf{H}_0}^2\big]\big)^{1/2} \\
&\leq \sqrt{2}\bigg[\sup_{t\in[0,T]} \big(\mathbb{E}\big[\big\|\mathbf{P}_{J_0\setminus J_n} X_t^{J_0}\big\|_{\mathbf{H}_0}^2\big]\big)^{1/2}\bigg] \\
&\quad \cdot \exp\!\big(\tfrac{1}{2}\big[\sqrt{2T}\,|\mathbf{P}_{J_n}\mathbf{F}|_{\operatorname{Lip}^0(\mathbf{H}_0,\mathbf{H}_0)} + \sqrt{2T}\,|\mathbf{P}_{J_n}\mathbf{B}|_{\operatorname{Lip}^0(\mathbf{H}_0,L_2(U,\mathbf{H}_0))}\big]^2\big) \\
&\leq \sqrt{2}\exp\!\big(\tfrac{1}{2}\big[\sqrt{2T}\,|\mathbf{F}|_{\operatorname{Lip}^0(\mathbf{H}_0,\mathbf{H}_0)} + \sqrt{2T}\,|\mathbf{B}|_{\operatorname{Lip}^0(\mathbf{H}_0,L_2(U,\mathbf{H}_0))}\big]^2\big) \\
&\quad \cdot \Bigg(\big(\mathbb{E}\big[\|\mathbf{P}_{J_0\setminus J_n}\xi\|_{\mathbf{H}_0}^2\big]\big)^{1/2} + \bigg[T\int_0^T \mathbb{E}\big[\big\|\mathbf{P}_{J_0\setminus J_n}\mathbf{F}(X_s^{J_0})\big\|_{\mathbf{H}_0}^2\big]\,\mathrm{d}s\bigg]^{1/2} \\
&\qquad + \bigg[\int_0^T \mathbb{E}\big[\big\|\mathbf{P}_{J_0\setminus J_n}\mathbf{B}(X_s^{J_0})\big\|_{L_2(U,\mathbf{H}_0)}^2\big]\,\mathrm{d}s\bigg]^{1/2}\Bigg).
\end{aligned} \tag{2.70}
$$

Therefore, Lebesgue's theorem of dominated convergence proves for all $(J_k)_{k\in\mathbb{N}_0} \subseteq \mathcal{P}(\mathbb{H})$ with $\bigcup_{k=1}^{\infty} J_k = J_0$ and $\forall k \in \mathbb{N}\colon J_k \subseteq J_{k+1} \in \mathcal{P}_0(\mathbb{H})$ that

$$
\limsup_{n\to\infty} \sup_{t\in[0,T]} \big(\mathbb{E}\big[\big\|X_t^{J_0} - X_t^{J_n}\big\|_{\mathbf{H}_0}^2\big]\big)^{1/2} = 0. \tag{2.71}
$$

Moreover, observe that (2.69) ensures for all $n \in \mathbb{N}$, $I \in \mathcal{P}_0(\mathbb{H}) \setminus \{\mathbb{H}\}$, $(J_k)_{k\in\mathbb{N}} \subseteq \mathcal{P}_0(\mathbb{H})$ with $\bigcup_{k=1}^{\infty} J_k = \mathbb{H}$ and $\forall k \in \mathbb{N}\colon I \subseteq J_k \subseteq J_{k+1}$ that

$$
\begin{aligned}
&\big|\mathbb{E}\big[\varphi(X_T^{\mathbb{H}})\big] - \mathbb{E}\big[\varphi(X_T^{I})\big]\big| \leq \big|\mathbb{E}\big[\varphi(X_T^{\mathbb{H}})\big] - \mathbb{E}\big[\varphi(X_T^{J_n})\big]\big| + \big|\mathbb{E}\big[\varphi(X_T^{J_n})\big] - \mathbb{E}\big[\varphi(X_T^{I})\big]\big| \\
&\leq |\varphi|_{C_{\mathrm{b}}^1(\mathbf{H}_0,\mathbb{R})}\big(\mathbb{E}\big[\big\|X_T^{\mathbb{H}} - X_T^{J_n}\big\|_{\mathbf{H}_0}^2\big]\big)^{1/2} \\
&+ \Bigg(\bigg[\sup_{K\in\mathcal{P}_0(\mathbb{H})} \sup_{t\in[0,T]} \big|u^K(t,\cdot)\big|_{C_{\mathrm{b}}^1(\mathbf{P}_K(\mathbf{H}_0),\mathbb{R})}\bigg]\bigg[\mathbb{E}\big[\|\xi\|_{\mathbf{H}_{2(\gamma-\beta)}}\big] + \sup_{K\in\mathcal{P}_0(\mathbb{H})}\int_0^T \mathbb{E}\big[\big\|\mathbf{F}(X_s^K)\big\|_{\mathbf{H}_{2(\gamma-\beta)}}\big]\,\mathrm{d}s\bigg] \\
&\quad + \|\mathbf{\Lambda}^{-\beta}\|_{L_2(\mathbf{H}_0)}^2\bigg[\sup_{K\in\mathcal{P}_0(\mathbb{H})} \sup_{t\in[0,T]} \big|u^K(t,\cdot)\big|_{C_{\mathrm{b}}^2(\mathbf{P}_K(\mathbf{H}_0),\mathbb{R})}\bigg] \sup_{K\in\mathcal{P}_0(\mathbb{H})}\int_0^T \mathbb{E}\big[\big\|\mathbf{B}(X_s^K)\big\|_{L(U,\mathbf{H}_\gamma)}^2\big]\,\mathrm{d}s\Bigg) \\
&\cdot \bigg[\inf_{h\in\mathbb{H}\setminus I}|\lambda_h|\bigg]^{\beta-\gamma}.
\end{aligned} \tag{2.72}
$$

Note that (2.71) and letting $n \to \infty$ in (2.72) complete the proof of Theorem 2.12 in the case that $I \in \mathcal{P}_0(\mathbb{H}) \setminus \{\mathbb{H}\}$. In a last step we prove the remaining cases. Estimate (2.72) ensures for all $n \in \mathbb{N}$, $I_0 \in \mathcal{P}(\mathbb{H}) \setminus \{\mathbb{H}\}$, $(I_k)_{k\in\mathbb{N}} \subseteq \mathcal{P}_0(I_0)$ with $\bigcup_{k=1}^{\infty} I_k = I_0$ and $\forall k \in \mathbb{N}\colon I_k \subseteq I_{k+1}$ that

$$
\begin{aligned}
&\big|\mathbb{E}\big[\varphi(X_T^{\mathbb{H}})\big] - \mathbb{E}\big[\varphi(X_T^{I_0})\big]\big| \leq \big|\mathbb{E}\big[\varphi(X_T^{\mathbb{H}})\big] - \mathbb{E}\big[\varphi(X_T^{I_n})\big]\big| + \big|\mathbb{E}\big[\varphi(X_T^{I_0})\big] - \mathbb{E}\big[\varphi(X_T^{I_n})\big]\big| \\
&\leq \Bigg(\bigg[\sup_{J\in\mathcal{P}_0(\mathbb{H})} \sup_{t\in[0,T]} \big|u^J(t,\cdot)\big|_{C_{\mathrm{b}}^1(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})}\bigg]\bigg[\mathbb{E}\big[\|\xi\|_{\mathbf{H}_{2(\gamma-\beta)}}\big] + \sup_{J\in\mathcal{P}_0(\mathbb{H})}\int_0^T \mathbb{E}\big[\big\|\mathbf{F}(X_s^J)\big\|_{\mathbf{H}_{2(\gamma-\beta)}}\big]\,\mathrm{d}s\bigg] \\
&\quad + \|\mathbf{\Lambda}^{-\beta}\|_{L_2(\mathbf{H}_0)}^2\bigg[\sup_{J\in\mathcal{P}_0(\mathbb{H})} \sup_{t\in[0,T]} \big|u^J(t,\cdot)\big|_{C_{\mathrm{b}}^2(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})}\bigg] \sup_{J\in\mathcal{P}_0(\mathbb{H})}\int_0^T \mathbb{E}\big[\big\|\mathbf{B}(X_s^J)\big\|_{L(U,\mathbf{H}_\gamma)}^2\big]\,\mathrm{d}s\Bigg) \\
&\quad \cdot \bigg[\inf_{h\in\mathbb{H}\setminus I}|\lambda_h|\bigg]^{\beta-\gamma} + |\varphi|_{C_{\mathrm{b}}^1(\mathbf{H}_0,\mathbb{R})}\big(\mathbb{E}\big[\big\|X_T^{I_0} - X_T^{I_n}\big\|_{\mathbf{H}_0}^2\big]\big)^{1/2}.
\end{aligned} \tag{2.73}
$$

Equation (2.71) and Lemma 2.11 thus complete the proof of Theorem 2.12. □

The next corollary is a direct consequence of Theorem 2.12 and Lemma 2.8.

Corollary 2.13. *Assume Setting 2.2, let $X^I\colon [0,T]\times\Omega\to\mathbf{P}_I(\mathbf{H}_\rho)$, $I\in\mathcal{P}(\mathbb{H})$, and $X^{J,x}\colon [0,T]\times\Omega\to\mathbf{P}_J(\mathbf{H}_0)$, $x\in\mathbf{P}_J(\mathbf{H}_0)$, $J\in\mathcal{P}_0(\mathbb{H})$, be $(\mathbb{F}_t)_{t\in[0,T]}$-predictable stochastic processes such that for all $I\in\mathcal{P}(\mathbb{H})$, $J\in\mathcal{P}_0(\mathbb{H})$, $x\in\mathbf{P}_J(\mathbf{H}_0)$, $t\in[0,T]$ it holds that $\sup_{s\in[0,T]}\mathbb{E}\big[\|X_s^I\|_{\mathbf{H}_\rho}^2+\|X_s^{J,x}\|_{\mathbf{H}_0}^2\big]<\infty$ and $\mathbb{P}$-a.s. that*

$$X_t^I = e^{t\mathbf{A}}\mathbf{P}_I\xi + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_I\mathbf{F}(X_s^I)\,\mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_I\mathbf{B}(X_s^I)\,\mathrm{d}W_s, \tag{2.74}$$

$$X_t^{J,x} = e^{t\mathbf{A}}x + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_J\mathbf{F}(X_s^{J,x})\,\mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_J\mathbf{B}(X_s^{J,x})\,\mathrm{d}W_s, \tag{2.75}$$

let $\varphi\in C_{\mathrm{b}}^2(\mathbf{H}_0,\mathbb{R})$, and let $u^J\colon[0,T]\times\mathbf{P}_J(\mathbf{H}_0)\to\mathbb{R}$, $J\in\mathcal{P}_0(\mathbb{H})$, be the functions which satisfy for all $J\in\mathcal{P}_0(\mathbb{H})$, $(t,x)\in[0,T]\times\mathbf{P}_J(\mathbf{H}_0)$ that $u^J(t,x)=\mathbb{E}\big[\varphi\big(X_t^{J,x}\big)\big]$. Then it holds for all $I\in\mathcal{P}(\mathbb{H})\setminus\{\mathbb{H}\}$ that

$$\begin{aligned}
&\big|\mathbb{E}\big[\varphi\big(X_T^{\mathbb{H}}\big)\big]-\mathbb{E}\big[\varphi\big(X_T^I\big)\big]\big| \\
&\leq \left[\max_{i\in\{1,2\}}\sup_{J\in\mathcal{P}_0(\mathbb{H})}\sup_{t\in[0,T]}\big|u^J(t,\cdot)\big|_{C_{\mathrm{b}}^i(\mathbf{P}_J(\mathbf{H}_0),\mathbb{R})}\right]\sup_{J\in\mathcal{P}_0(\mathbb{H})}\sup_{t\in[0,T]}\max\big\{1,\mathbb{E}\big[\|X_t^J\|_{\mathbf{H}_\rho}^2\big]\big\} \\
&\quad\cdot\Big(\mathbb{E}\big[\|\xi\|_{\mathbf{H}_{2(\gamma-\beta)}}\big]+T\,\big\|\mathbf{F}|_{\mathbf{H}_\rho}\big\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,\mathbf{H}_{2(\gamma-\beta)})}+T\,\|\mathbf{\Lambda}^{-\beta}\|_{L_2(\mathbf{H}_0)}^2\big\|\mathbf{B}|_{\mathbf{H}_\rho}\big\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,L(U,\mathbf{H}_\gamma))}^2\Big) \\
&\quad\cdot\left[\inf_{h\in\mathbb{H}\setminus I}|\lambda_h|\right]^{\beta-\gamma}<\infty.
\end{aligned}\tag{2.76}$$

The last result in this subsection, Corollary 2.14 below, follows immediately from Corollary 2.13 and Lemmas 2.10 and 2.8.

Corollary 2.14. *Assume Setting 2.2 and let $X^I\colon[0,T]\times\Omega\to\mathbf{P}_I(\mathbf{H}_\rho)$, $I\in\mathcal{P}(\mathbb{H})$, be $(\mathbb{F}_t)_{t\in[0,T]}$-predictable stochastic processes such that for all $I\in\mathcal{P}(\mathbb{H})$, $t\in[0,T]$ it holds that $\sup_{s\in[0,T]}\mathbb{E}\big[\|X_s^I\|_{\mathbf{H}_\rho}^2\big]<\infty$ and $\mathbb{P}$-a.s. that*

$$X_t^I = e^{t\mathbf{A}}\mathbf{P}_I\xi + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_I\mathbf{F}(X_s^I)\,\mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_I\mathbf{B}(X_s^I)\,\mathrm{d}W_s. \tag{2.77}$$

Then it holds for all $\varphi\in C_{\mathrm{b}}^2(\mathbf{H}_0,\mathbb{R})$, $I\in\mathcal{P}(\mathbb{H})\setminus\{\mathbb{H}\}$ that

$$\begin{aligned}
&\big|\mathbb{E}\big[\varphi\big(X_T^{\mathbb{H}}\big)\big]-\mathbb{E}\big[\varphi\big(X_T^I\big)\big]\big| \leq \|\varphi\|_{C_{\mathrm{b}}^2(\mathbf{H}_0,\mathbb{R})}\max\big\{1,\mathbb{E}\big[\|\xi\|_{\mathbf{H}_\rho}^2\big]\big\} \\
&\quad\cdot\Big(\mathbb{E}\big[\|\xi\|_{\mathbf{H}_{2(\gamma-\beta)}}\big]+T\,\big\|\mathbf{F}|_{\mathbf{H}_\rho}\big\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,\mathbf{H}_{2(\gamma-\beta)})}+T\,\|\mathbf{\Lambda}^{-\beta}\|_{L_2(\mathbf{H}_0)}^2\big\|\mathbf{B}|_{\mathbf{H}_\rho}\big\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,L(U,\mathbf{H}_\gamma))}^2\Big) \\
&\quad\cdot\max\Big\{1,\,\big[T\big((C_{\mathbf{F}})^2+2\,(C_{\mathbf{B}})^2\big)\big]^{1/2}\Big\}\exp\!\big(T\big[\tfrac{1}{2}+3\,|\mathbf{F}|_{\mathrm{Lip}^0(\mathbf{H}_0,\mathbf{H}_0)}+4\,|\mathbf{B}|_{\mathrm{Lip}^0(\mathbf{H}_0,L_2(U,\mathbf{H}_0))}^2\big]\big) \\
&\quad\cdot\exp\!\Big(T\Big[2\,\big\|\mathbf{F}|_{\mathbf{H}_\rho}\big\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,\mathbf{H}_\rho)}+\big\|\mathbf{B}|_{\mathbf{H}_\rho}\big\|_{\mathrm{Lip}^0(\mathbf{H}_\rho,L_2(U,\mathbf{H}_\rho))}^2\Big]\Big)\left[\inf_{h\in\mathbb{H}\setminus I}|\lambda_h|\right]^{\beta-\gamma}<\infty.
\end{aligned}\tag{2.78}$$

2.2.3 Semi-linear stochastic wave equations and the continuous version of the hyperbolic Anderson model

Roughly speaking, the following elementary and well-known lemma provides a useful criterion for determining whether a vector belonging to an interpolation space associated to a symmetric diagonal linear operator possesses more regularity (cf., e.g., Sell & You [279, Example 37.1]).

Lemma 2.15. *Consider the notation in Subsection 2.1.1, let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, let $(H, \langle \cdot, \cdot \rangle_H, \|\cdot\|_H)$ be a $\mathbb{K}$-Hilbert space, let $\mathbb{H} \subseteq H$ be a non-empty orthonormal basis of H, let $A\colon D(A) \subseteq H \to H$ be a symmetric diagonal linear operator with $\inf(\sigma_{\mathrm{P}}(A)) > 0$, and let $(H_r, \langle \cdot, \cdot \rangle_{H_r}, \|\cdot\|_{H_r})$, $r \in \mathbb{R}$, be a family of interpolation spaces associated to A. Then*

(i) for all $v \in \bigcup_{s \in \mathbb{R}} H_s$, $r \in \mathbb{R}$ it holds that $v \in H_r$ if and only if

$$\sup_{w \in \operatorname{span}_H(\mathbb{H}) \setminus \{0\}} \frac{|\langle w, v \rangle_H|}{\|w\|_{H_{-r}}} < \infty, \tag{2.79}$$

(ii) for all $s \in \mathbb{R}$, $v \in H_{-s}$, $r \in [-s, \infty)$ it holds that $v \in H_r$ if and only if

$$\sup_{w \in H_s \setminus \{0\}} \frac{|\langle w, v \rangle_H|}{\|w\|_{H_{-r}}} < \infty, \tag{2.80}$$

and

(iii) for all $r \in \mathbb{R}$, $v \in H_r$, $s \in [-r, \infty)$ it holds that

$$\|v\|_{H_r} = \sup_{w \in \operatorname{span}_H(\mathbb{H}) \setminus \{0\}} \frac{|\langle w, v \rangle_H|}{\|w\|_{H_{-r}}} = \sup_{w \in H_s \setminus \{0\}} \frac{|\langle w, v \rangle_H|}{\|w\|_{H_{-r}}}. \tag{2.81}$$

Proof of Lemma 2.15. Note that for all $r \in \mathbb{R}$, $v \in H_r$, $w \in H_{-r}$ it holds that

$$|\langle w, v \rangle_H| = |\langle A^{-r} w, A^r v \rangle_H| \leq \|w\|_{H_{-r}} \|v\|_{H_r}. \tag{2.82}$$

This proves the "$\Rightarrow$" direction in the statement of (i). Next we consider the "$\Leftarrow$" direction in the statement of (i) and the first equality in (2.81). For this let $\mathbf{s} \in \mathbb{R}$, $\mathbf{r} \in [\mathbf{s}, \infty)$, $\mathbf{v} \in H_{\mathbf{s}}$ satisfy

$$\sup_{w \in \operatorname{span}_H(\mathbb{H}) \setminus \{0\}} \frac{|\langle w, \mathbf{v} \rangle_H|}{\|w\|_{H_{-\mathbf{r}}}} < \infty. \tag{2.83}$$

Observe that it holds that

$$\begin{aligned}
\sup_{w \in \operatorname{span}_H(\mathbb{H}) \setminus \{0\}} \frac{|\langle \mathbf{v}, w \rangle_{H_{\mathbf{r}}}|}{\|w\|_{H_{\mathbf{r}}}} &= \sup_{w \in \operatorname{span}_H(\mathbb{H}) \setminus \{0\}} \frac{|\langle A^{\mathbf{r}} \mathbf{v}, A^{\mathbf{r}} w \rangle_H|}{\|A^{2\mathbf{r}} w\|_{H_{-\mathbf{r}}}} \\
&= \sup_{w \in \operatorname{span}_H(\mathbb{H}) \setminus \{0\}} \frac{|\langle \mathbf{v}, A^{2\mathbf{r}} w \rangle_H|}{\|A^{2\mathbf{r}} w\|_{H_{-\mathbf{r}}}} = \sup_{w \in \operatorname{span}_H(\mathbb{H}) \setminus \{0\}} \frac{|\langle \mathbf{v}, w \rangle_H|}{\|w\|_{H_{-\mathbf{r}}}} \\
&= \sup_{w \in \operatorname{span}_H(\mathbb{H}) \setminus \{0\}} \frac{|\langle w, \mathbf{v} \rangle_H|}{\|w\|_{H_{-\mathbf{r}}}} < \infty.
\end{aligned} \tag{2.84}$$

This ensures that there exists $\varphi \in L(H_{\mathbf{r}}, \mathbb{K})$ such that for all $w \in \operatorname{span}_H(\mathbb{H})$ it holds that $\varphi(w) = \langle \mathbf{v}, w \rangle_{H_{\mathbf{r}}}$. The Riesz–Fréchet representation theorem hence proves that there exists a vector $\mathbf{u} \in H_{\mathbf{r}}$ such that $\forall w \in H_{\mathbf{r}} \colon \langle \mathbf{u}, w \rangle_{H_{\mathbf{r}}} = \varphi(w)$ and

$$\|\mathbf{u}\|_{H_{\mathbf{r}}} = \|\varphi\|_{L(H_{\mathbf{r}},\mathbb{K})} = \sup_{w \in \operatorname{span}_H(\mathbb{H})\setminus\{0\}} \frac{|\langle \mathbf{v}, w \rangle_{H_{\mathbf{r}}}|}{\|w\|_{H_{\mathbf{r}}}}. \tag{2.85}$$

This implies for all $w \in H_{\mathbf{s}}$ that

$$\langle \mathbf{u}, w \rangle_{H_{\mathbf{s}}} = \langle A^{\mathbf{s}-\mathbf{r}} \mathbf{u}, A^{\mathbf{s}-\mathbf{r}} w \rangle_{H_{\mathbf{r}}} = \langle \mathbf{u}, A^{2(\mathbf{s}-\mathbf{r})} w \rangle_{H_{\mathbf{r}}} = \varphi(A^{2(\mathbf{s}-\mathbf{r})} w) = \langle \mathbf{v}, w \rangle_{H_{\mathbf{s}}}. \tag{2.86}$$

Combining this with (2.84) and (2.85) demonstrates that $\mathbf{v} = \mathbf{u} \in H_{\mathbf{r}}$ and

$$\|\mathbf{v}\|_{H_{\mathbf{r}}} = \|\mathbf{u}\|_{H_{\mathbf{r}}} = \sup_{w \in \operatorname{span}_H(\mathbb{H})\setminus\{0\}} \frac{|\langle \mathbf{v}, w \rangle_{H_{\mathbf{r}}}|}{\|w\|_{H_{\mathbf{r}}}} = \sup_{w \in \operatorname{span}_H(\mathbb{H})\setminus\{0\}} \frac{|\langle w, \mathbf{v} \rangle_H|}{\|w\|_{H_{-\mathbf{r}}}}. \tag{2.87}$$

This and the fact that $\forall s \in \mathbb{R}, r \in (-\infty, s] \colon H_s \subseteq H_r$ establish (i) and the first equality in (2.81). Next note that the "$\Rightarrow$" direction in the statement of (ii) follows directly from (2.82), while the "$\Leftarrow$" direction in the statement of (ii) is a consequence of (i). Finally, (2.82) also shows for all $r \in \mathbb{R}$, $v \in H_r$, $s \in [-r, \infty)$ that

$$\begin{aligned} \|v\|_{H_r} &= \sup_{w \in \operatorname{span}_H(\mathbb{H})\setminus\{0\}} \frac{|\langle w, v \rangle_H|}{\|w\|_{H_{-r}}} \leq \sup_{w \in H_s\setminus\{0\}} \frac{|\langle w, v \rangle_H|}{\|w\|_{H_{-r}}} \\ &\leq \sup_{w \in H_{-r}\setminus\{0\}} \frac{|\langle w, v \rangle_H|}{\|w\|_{H_{-r}}} \leq \|v\|_{H_r}. \end{aligned} \tag{2.88}$$

This proves (iii). The proof of Lemma 2.15 is thus complete. □

In the next result, Corollary 2.16, we specialise Corollary 2.13 above to the case of semi-linear stochastic wave equations. Corollary 2.16 is an elementary consequence of Corollary 2.13.

Corollary 2.16. *Consider the notation in Subsection 2.1.1, let $T, \vartheta \in (0, \infty)$, $\gamma \in (1/4, 1/2)$, $\rho \in [0, 2\gamma - 1/2)$, $\varrho \in [1/6, \infty)$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a normal filtration $(\mathbb{F}_t)_{t\in[0,T]}$, let $(H, \langle\cdot,\cdot\rangle_H, \|\cdot\|_H) = \big(L^2(\mu_{(0,1)}; \mathbb{R}), \langle\cdot,\cdot\rangle_{L^2(\mu_{(0,1)};\mathbb{R})}, \|\cdot\|_{L^2(\mu_{(0,1)};\mathbb{R})}\big)$, let $(W_t)_{t\in[0,T]}$ be an id_H-cylindrical $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$-Wiener process, let $(e_n)_{n\in\mathbb{N}} \subseteq H$ satisfy for all $n \in \mathbb{N}$ and $\mu_{(0,1)}$-a.e. $x \in (0,1)$ that $e_n(x) = \sqrt{2}\sin(n\pi x)$, let $A \colon D(A) \subseteq H \to H$ be the Laplacian with Dirichlet boundary conditions on H multiplied by ϑ, let $(H_r, \langle\cdot,\cdot\rangle_{H_r}, \|\cdot\|_{H_r})$, $r \in \mathbb{R}$, be a family of interpolation spaces associated to $-A$, let $\mathbf{P}_N \in L(H \times H_{-1/2})$, $N \in \mathbb{N} \cup \{\infty\}$, be the linear operators which satisfy for all $N \in \mathbb{N} \cup \{\infty\}$, $v, w \in H$ that $\mathbf{P}_N(v, w) = \big(\sum_{n=1}^N \langle e_n, v \rangle_H e_n, \sum_{n=1}^N \langle e_n, w \rangle_H e_n\big)$, let $\mathbf{A} \colon D(\mathbf{A}) \subseteq H \times H_{-1/2} \to H \times H_{-1/2}$ be the linear operator which satisfies $D(\mathbf{A}) = H_{1/2} \times H$ and $\forall (v, w) \in H_{1/2} \times H \colon \mathbf{A}(v, w) = (w, Av)$, let $\xi \in L^2(\mathbb{P}|_{\mathbb{F}_0}; H_{1/2} \times H)$, $\varphi \in C^2_{\mathrm{b}}(H \times H_{-1/2}, \mathbb{R})$, $f \in \operatorname{Lip}^2((0,1) \times \mathbb{R}, \mathbb{R})$, $B \in \operatorname{Lip}^0(H, L_2(H, H_{-1/2}))$ satisfy $\forall v \in H_\rho, u \in H \colon B(v)u \in H_{\gamma-1/2}$, $\forall v \in H_\rho \colon (H \ni u \mapsto B(v)u \in H_{\rho-1/2}) \in L_2(H, H_{\rho-1/2})$, $(H_\rho \ni v \mapsto (H \ni u \mapsto B(v)u \in H_{\rho-1/2}) \in L_2(H, H_{\rho-1/2})) \in \operatorname{Lip}^0(H_\rho, L_2(H, H_{\rho-1/2}))$, $\forall v \in H_\rho \colon (H \ni u \mapsto B(v)u \in H_{\gamma-1/2}) \in L(H, H_{\gamma-1/2})$, $(H_\rho \ni v \mapsto (H \ni u \mapsto B(v)u \in H_{\gamma-1/2}) \in L(H, H_{\gamma-1/2})) \in \operatorname{Lip}^0(H_\rho, L(H, H_{\gamma-1/2}))$, $B|_{H_\varrho} \in C^2_{\mathrm{b}}(H_\varrho, L_2(H, H_{-1/2}))$, and $\sup_{x,v_1,v_2 \in H_\varrho,\, \max\{\|v_1\|_H, \|v_2\|_H\} \leq 1} \|B''(x)(v_1, v_2)\|_{L_2(H,H_{-1/2})} < \infty$, and let $\mathbf{F} \colon H \times H_{-1/2} \to$*

$H_{1/2} \times H$ and $\mathbf{B}\colon H \times H_{-1/2} \to L_2(H, H \times H_{-1/2})$ be the functions which satisfy for all $v, u \in H$, $w \in H_{-1/2}$ and $\mu_{(0,1)}$-a.e. $x \in (0,1)$ that $\big(\mathbf{F}(v,w)\big)(x) = \big(0, f(x, v(x))\big)$ and $\mathbf{B}(v,w)u = \big(0, B(v)u\big)$. Then

(i) it holds that $\mathbf{F} \in \mathrm{Lip}^0(H \times H_{-1/2}, H_{1/2} \times H)$, $\mathbf{F}|_{H_\varrho \times H_{\varrho-1/2}} \in \mathrm{Lip}^2(H_\varrho \times H_{\varrho-1/2}, H_{1/2} \times H)$, $\mathbf{B} \in \mathrm{Lip}^0(H \times H_{-1/2}, L_2(H, H \times H_{-1/2}))$, $\forall v \in H_\rho \times H_{\rho-1/2}, u \in H\colon \mathbf{B}(v)u \in H_\gamma \times H_{\gamma-1/2}$, $\forall v \in H_\rho \times H_{\rho-1/2}\colon (H \ni u \mapsto \mathbf{B}(v)u \in H_\rho \times H_{\rho-1/2}) \in L_2(H, H_\rho \times H_{\rho-1/2})$, $(H_\rho \times H_{\rho-1/2} \ni v \mapsto (H \ni u \mapsto \mathbf{B}(v)u \in H_\rho \times H_{\rho-1/2}) \in L_2(H, H_\rho \times H_{\rho-1/2})) \in \mathrm{Lip}^0(H_\rho \times H_{\rho-1/2}, L_2(H, H_\rho \times H_{\rho-1/2}))$, $\forall v \in H_\rho \times H_{\rho-1/2}\colon (H \ni u \mapsto \mathbf{B}(v)u \in H_\gamma \times H_{\gamma-1/2}) \in L(H, H_\gamma \times H_{\gamma-1/2})$, $(H_\rho \times H_{\rho-1/2} \ni v \mapsto (H \ni u \mapsto \mathbf{B}(v)u \in H_\gamma \times H_{\gamma-1/2}) \in L(H, H_\gamma \times H_{\gamma-1/2})) \in \mathrm{Lip}^0(H_\rho \times H_{\rho-1/2}, L(H, H_\gamma \times H_{\gamma-1/2}))$, $\mathbf{B}|_{H_\varrho \times H_{\varrho-1/2}} \in C^2_{\mathrm{b}}(H_\varrho \times H_{\varrho-1/2}, L_2(H, H \times H_{-1/2}))$, and

$$\forall \delta \in (-\infty, 1/4)\colon \quad \sup_{\substack{x \in H_\varrho \times H_{\varrho-1/2},\\ v_1, v_2 \in H_\varrho \times H_{\varrho-1/2} \setminus \{0\}}} \frac{\|\mathbf{F}''(x)(v_1,v_2)\|_{H_\delta \times H_{\delta-1/2}} + \|\mathbf{B}''(x)(v_1,v_2)\|_{L_2(H, H \times H_{-1/2})}}{\|v_1\|_{H \times H_{-1/2}} \|v_2\|_{H \times H_{-1/2}}} < \infty, \tag{2.89}$$

(ii) it holds that there exist up to modifications unique $(\mathbb{F}_t)_{t \in [0,T]}$-predictable stochastic processes $X^N\colon [0,T] \times \Omega \to \mathbf{P}_N(H_\rho \times H_{\rho-1/2})$, $N \in \mathbb{N} \cup \{\infty\}$, such that for all $N \in \mathbb{N} \cup \{\infty\}$, $t \in [0,T]$ it holds that $\sup_{s \in [0,T]} \mathbb{E}\big[\|X^N_s\|^2_{H_\rho \times H_{\rho-1/2}}\big] < \infty$ and $\mathbb{P}$-a.s. that

$$X^N_t = e^{t\mathbf{A}} \mathbf{P}_N \xi + \int_0^t e^{(t-s)\mathbf{A}} \mathbf{P}_N \mathbf{F}(X^N_s)\, \mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}} \mathbf{P}_N \mathbf{B}(X^N_s)\, \mathrm{d}W_s, \tag{2.90}$$

and

(iii) it holds for all $\varepsilon \in (4(1/2 - \gamma), \infty)$ that there exists a real number $C \in [0, \infty)$ such that for all $N \in \mathbb{N}$ it holds that

$$\big|\mathbb{E}\big[\varphi(X^\infty_T)\big] - \mathbb{E}\big[\varphi(X^N_T)\big]\big| \leq C \cdot N^{\varepsilon - 1}. \tag{2.91}$$

Proof of Corollary 2.16. Throughout this proof let $f_{k,\ell}\colon (0,1) \times \mathbb{R} \to \mathbb{R}$, $k, \ell \in \{0,1,2\}$ with $k + \ell \leq 2$, be the functions which satisfy for all $k, \ell \in \{0,1,2\}$, $(x,y) \in (0,1) \times \mathbb{R}$ with $k + \ell \leq 2$ that $f_{k,\ell}(x,y) = \big(\frac{\partial^{k+\ell}}{\partial x^k \partial y^\ell} f\big)(x,y)$ and let $F\colon H \to H$ be the function which satisfies for all $v \in H$ and $\mu_{(0,1)}$-a.e. $x \in (0,1)$ that $\big(F(v)\big)(x) = f(x, v(x))$. Then note that for all $u, v \in H$, $w \in H_{-1/2}$ it holds that $\mathbf{F}(v,w) = \big(0, F(v)\big)$ and

$$\begin{aligned} \|F(u) - F(v)\|_H &= \left(\int_0^1 |f(x,u(x)) - f(x,v(x))|^2\, \mathrm{d}x\right)^{1/2} \\ &\leq |f|_{\mathrm{Lip}^0((0,1) \times \mathbb{R}, \mathbb{R})} \|u - v\|_H. \end{aligned} \tag{2.92}$$

This proves that $F \in \mathrm{Lip}^0(H, H)$. Hence, we obtain that $\mathbf{F} \in \mathrm{Lip}^0(H \times H_{-1/2}, H_{1/2} \times H)$. Next observe that the Sobolev embedding theorem ensures for all $\delta \in [1, 6]$ that

$$\sup_{w \in H_\varrho \setminus \{0\}} \frac{\|w\|_{L^\delta(\mu_{(0,1)}; \mathbb{R})}}{\|w\|_{H_\varrho}} < \infty. \tag{2.93}$$

Moreover, note that it holds for all $v, h \in H$ and $\mu_{(0,1)}$-a.e. $x \in (0,1)$ that

$$\begin{aligned}
&|f(x, v(x) + h(x)) - f(x, v(x)) - f_{0,1}(x, v(x))h(x)| \\
&= \left| \int_0^1 \big[f_{0,1}(x, v(x) + yh(x)) - f_{0,1}(x, v(x)) \big] h(x) \, \mathrm{d}y \right| \leq |f|_{\mathrm{Lip}^1((0,1)\times\mathbb{R},\mathbb{R})} |h(x)|^2.
\end{aligned} \tag{2.94}$$

This, Hölder's inequality, and (2.93) imply for all $v \in H_\varrho$, $h \in H_\varrho \setminus \{0\}$ that

$$\begin{aligned}
&\frac{1}{\|h\|_{H_\varrho}} \left(\int_0^1 |f(x, v(x) + h(x)) - f(x, v(x)) - f_{0,1}(x, v(x))h(x)|^2 \, \mathrm{d}x \right)^{1/2} \\
&\leq |f|_{\mathrm{Lip}^1((0,1)\times\mathbb{R},\mathbb{R})} \frac{\|h\|^2_{L^4(\mu_{(0,1)};\mathbb{R})}}{\|h\|_{H_\varrho}} \leq |f|_{\mathrm{Lip}^1((0,1)\times\mathbb{R},\mathbb{R})} \left(\sup_{w \in H_\varrho \setminus \{0\}} \frac{\|w\|_{L^4(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_\varrho}} \right)^2 \|h\|_{H_\varrho} < \infty.
\end{aligned} \tag{2.95}$$

In addition, observe that it holds for all $v, h \in H_\varrho$ that

$$\begin{aligned}
&\left(\int_0^1 |f_{0,1}(x, v(x))h(x)|^2 \, \mathrm{d}x \right)^{1/2} \leq |f|_{C^1_{\mathrm{b}}((0,1)\times\mathbb{R},\mathbb{R})} \|h\|_H \\
&\leq |f|_{C^1_{\mathrm{b}}((0,1)\times\mathbb{R},\mathbb{R})} \left(\sup_{w \in H_\varrho \setminus \{0\}} \frac{\|w\|_H}{\|w\|_{H_\varrho}} \right) \|h\|_{H_\varrho} \\
&= |f|_{\mathrm{Lip}^0((0,1)\times\mathbb{R},\mathbb{R})} \left(\sup_{w \in H_\varrho \setminus \{0\}} \frac{\|w\|_H}{\|w\|_{H_\varrho}} \right) \|h\|_{H_\varrho} < \infty.
\end{aligned} \tag{2.96}$$

Inequalities (2.95)–(2.96) prove that $F|_{H_\varrho} \colon H_\varrho \to H$ is Fréchet differentiable, that for all $v, h \in H_\varrho$ and $\mu_{(0,1)}$-a.e. $x \in (0,1)$ it holds that

$$\big(F'(v)h\big)(x) = f_{0,1}(x, v(x))h(x), \tag{2.97}$$

and that $\sup_{v \in H_\varrho} \|F'(v)\|_{L(H_\varrho, H)} \leq |f|_{C^1_{\mathrm{b}}((0,1)\times\mathbb{R},\mathbb{R})} < \infty$. Furthermore, Hölder's inequality and (2.93) show for all $u, v, h \in H_\varrho$ that

$$\begin{aligned}
&\|(F'(u) - F'(v))h\|_H = \left(\int_0^1 \big| \big[f_{0,1}(x, u(x)) - f_{0,1}(x, v(x)) \big] h(x) \big|^2 \, \mathrm{d}x \right)^{1/2} \\
&\leq |f|_{\mathrm{Lip}^1((0,1)\times\mathbb{R},\mathbb{R})} \|u - v\|_{L^4(\mu_{(0,1)};\mathbb{R})} \|h\|_{L^4(\mu_{(0,1)};\mathbb{R})} \\
&\leq |f|_{\mathrm{Lip}^1((0,1)\times\mathbb{R},\mathbb{R})} \left(\sup_{w \in H_\varrho \setminus \{0\}} \frac{\|w\|_{L^4(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_\varrho}} \right)^2 \|u - v\|_{H_\varrho} \|h\|_{H_\varrho} < \infty.
\end{aligned} \tag{2.98}$$

This ensures that $F|_{H_\varrho} \in \mathrm{Lip}^1(H_\varrho, H)$. Similarly, observe that for all $v, h, g \in H$ and $\mu_{(0,1)}$-a.e. $x \in (0,1)$ it holds that

$$\begin{aligned}
&|f_{0,1}(x, v(x) + g(x))h(x) - f_{0,1}(x, v(x))h(x) - f_{0,2}(x, v(x))h(x)g(x)| \\
&= \left| \int_0^1 \big[f_{0,2}(x, v(x) + yg(x)) - f_{0,2}(x, v(x)) \big] h(x)g(x) \, \mathrm{d}y \right| \\
&\leq |f|_{\mathrm{Lip}^2((0,1)\times\mathbb{R},\mathbb{R})} |h(x)| |g(x)|^2.
\end{aligned} \tag{2.99}$$

This, Hölder's inequality, and (2.93) establish for all $v, h \in H_\varrho$, $g \in H_\varrho \setminus \{0\}$ that

$$\begin{aligned}
&\frac{1}{\|g\|_{H_\varrho}}\left(\int_0^1 |f_{0,1}(x, v(x) + g(x))h(x) - f_{0,1}(x, v(x))h(x) - f_{0,2}(x, v(x))h(x)g(x)|^2 \, \mathrm{d}x\right)^{1/2} \\
&\leq \frac{|f|_{\mathrm{Lip}^2((0,1)\times\mathbb{R},\mathbb{R})}}{\|g\|_{H_\varrho}}\left(\int_0^1 |h(x)|^2|g(x)|^4 \, \mathrm{d}x\right)^{1/2} \\
&\leq |f|_{\mathrm{Lip}^2((0,1)\times\mathbb{R},\mathbb{R})}\frac{\|h\|_{L^6(\mu_{(0,1)};\mathbb{R})}\|g\|^2_{L^6(\mu_{(0,1)};\mathbb{R})}}{\|g\|_{H_\varrho}} \\
&\leq |f|_{\mathrm{Lip}^2((0,1)\times\mathbb{R},\mathbb{R})}\left(\sup_{w\in H_\varrho\setminus\{0\}}\frac{\|w\|_{L^6(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_\varrho}}\right)^3 \|h\|_{H_\varrho}\|g\|_{H_\varrho} < \infty.
\end{aligned} \tag{2.100}$$

Furthermore, Hölder's inequality and (2.93) also prove for all $v, h, g \in H_\varrho$ that

$$\begin{aligned}
&\left(\int_0^1 |f_{0,2}(x, v(x))h(x)g(x)|^2 \, \mathrm{d}x\right)^{1/2} \leq |f|_{C^2_{\mathrm{b}}((0,1)\times\mathbb{R},\mathbb{R})}\|h\|_{L^4(\mu_{(0,1)};\mathbb{R})}\|g\|_{L^4(\mu_{(0,1)};\mathbb{R})} \\
&\leq |f|_{C^2_{\mathrm{b}}((0,1)\times\mathbb{R},\mathbb{R})}\left(\sup_{w\in H_\varrho\setminus\{0\}}\frac{\|w\|_{L^4(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_\varrho}}\right)^2 \|h\|_{H_\varrho}\|g\|_{H_\varrho} \\
&= |f|_{\mathrm{Lip}^1((0,1)\times\mathbb{R},\mathbb{R})}\left(\sup_{w\in H_\varrho\setminus\{0\}}\frac{\|w\|_{L^4(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_\varrho}}\right)^2 \|h\|_{H_\varrho}\|g\|_{H_\varrho} < \infty.
\end{aligned} \tag{2.101}$$

Combining (2.100)–(2.101) ensures that $F|_{H_\varrho} \colon H_\varrho \to H$ is twice Fréchet differentiable, that for all $v, h, g \in H_\varrho$ and $\mu_{(0,1)}$-a.e. $x \in (0,1)$ it holds that

$$\big(F''(v)(h, g)\big)(x) = f_{0,2}(x, v(x))h(x)g(x), \tag{2.102}$$

and that

$$\sup_{v\in H_\varrho}\|F''(v)\|_{L^{(2)}(H_\varrho,H)} \leq |f|_{C^2_{\mathrm{b}}((0,1)\times\mathbb{R},\mathbb{R})}\left(\sup_{w\in H_\varrho\setminus\{0\}}\frac{\|w\|_{L^4(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_\varrho}}\right)^2 < \infty. \tag{2.103}$$

In addition, Hölder's inequality and (2.93) establish for all $u, v, h, g \in H_\varrho$ that

$$\begin{aligned}
&\|(F''(u) - F''(v))(h, g)\|_H = \left(\int_0^1 \big|\big[f_{0,2}(x, u(x)) - f_{0,2}(x, v(x))\big]h(x)g(x)\big|^2 \, \mathrm{d}x\right)^{1/2} \\
&\leq |f|_{\mathrm{Lip}^2((0,1)\times\mathbb{R},\mathbb{R})}\|u - v\|_{L^6(\mu_{(0,1)};\mathbb{R})}\|h\|_{L^6(\mu_{(0,1)};\mathbb{R})}\|g\|_{L^6(\mu_{(0,1)};\mathbb{R})} \\
&\leq |f|_{\mathrm{Lip}^2((0,1)\times\mathbb{R},\mathbb{R})}\left(\sup_{w\in H_\varrho\setminus\{0\}}\frac{\|w\|_{L^6(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_\varrho}}\right)^3 \|u - v\|_{H_\varrho}\|h\|_{H_\varrho}\|g\|_{H_\varrho} < \infty.
\end{aligned} \tag{2.104}$$

This shows that $F|_{H_\varrho} \in \mathrm{Lip}^2(H_\varrho, H)$. This proves that $\mathbf{F}|_{H_\varrho\times H_{\varrho-1/2}} \in \mathrm{Lip}^2(H_\varrho \times H_{\varrho-1/2}, H_{1/2} \times H)$. Next, note that the assumptions that $B \in \mathrm{Lip}^0(H, L_2(H, H_{-1/2}))$, $\forall v \in H_\rho, u \in H \colon B(v)u \in H_{\gamma-1/2}$, $\forall v \in H_\rho \colon (H \ni u \mapsto B(v)u \in H_{\rho-1/2}) \in L_2(H, H_{\rho-1/2})$, $(H_\rho \ni v \mapsto (H \ni u \mapsto B(v)u \in H_{\rho-1/2}) \in L_2(H, H_{\rho-1/2})) \in \mathrm{Lip}^0(H_\rho, L_2(H, H_{\rho-1/2}))$, $\forall v \in H_\rho \colon (H \ni u \mapsto B(v)u \in H_{\gamma-1/2}) \in L(H, H_{\gamma-1/2})$, $(H_\rho \ni v \mapsto (H \ni u \mapsto B(v)u \in H_{\gamma-1/2}) \in L(H, H_{\gamma-1/2})) \in \mathrm{Lip}^0(H_\rho, L(H, H_{\gamma-1/2}))$, and $B|_{H_\varrho} \in C^2_{\mathrm{b}}(H_\varrho, L_2(H, H_{-1/2}))$ ensure that $\mathbf{B} \in \mathrm{Lip}^0(H \times H_{-1/2}, L_2(H, H \times H_{-1/2}))$, $\forall v \in H_\rho \times H_{\rho-1/2}, u \in H \colon \mathbf{B}(v)u \in$

$H_\gamma \times H_{\gamma-1/2}$, $\forall v \in H_\rho \times H_{\rho-1/2}\colon (H \ni u \mapsto \mathbf{B}(v)u \in H_\rho \times H_{\rho-1/2}) \in L_2(H, H_\rho \times H_{\rho-1/2})$, $(H_\rho \times H_{\rho-1/2} \ni v \mapsto (H \ni u \mapsto \mathbf{B}(v)u \in H_\rho \times H_{\rho-1/2}) \in L_2(H, H_\rho \times H_{\rho-1/2})) \in \operatorname{Lip}^0(H_\rho \times H_{\rho-1/2}, L_2(H, H_\rho \times H_{\rho-1/2}))$, $\forall v \in H_\rho \times H_{\rho-1/2}\colon (H \ni u \mapsto \mathbf{B}(v)u \in H_\gamma \times H_{\gamma-1/2}) \in L(H, H_\gamma \times H_{\gamma-1/2})$, $(H_\rho \times H_{\rho-1/2} \ni v \mapsto (H \ni u \mapsto \mathbf{B}(v)u \in H_\gamma \times H_{\gamma-1/2}) \in L(H, H_\gamma \times H_{\gamma-1/2})) \in \operatorname{Lip}^0(H_\rho \times H_{\rho-1/2}, L(H, H_\gamma \times H_{\gamma-1/2}))$, and $\mathbf{B}|_{H_\varrho \times H_{\varrho-1/2}} \in C^2_{\mathrm{b}}(H_\varrho \times H_{\varrho-1/2}, L_2(H, H \times H_{-1/2}))$. In addition, Lemma 2.15 proves for all $\delta \in (-\infty, 1/4)$, $v, h, g \in H_\varrho$ that

$$
\begin{aligned}
\|F''(v)(h,g)\|_{H_{\delta-1/2}} &= \sup_{w \in H_{1/2-\delta}\setminus\{0\}} \frac{\langle w, F''(v)(h,g)\rangle_H}{\|w\|_{H_{1/2-\delta}}} \\
&\leq |f|_{C^2_{\mathrm{b}}((0,1)\times\mathbb{R},\mathbb{R})} \left(\sup_{w \in H_{1/2-\delta}\setminus\{0\}} \frac{\|w\|_{L^\infty(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_{1/2-\delta}}} \right) \|h\|_H \|g\|_H < \infty.
\end{aligned}
\tag{2.105}
$$

This and the assumption that

$$
\sup_{x, v_1, v_2 \in H_\varrho,\, \max\{\|v_1\|_H, \|v_2\|_H\} \leq 1} \|B''(x)(v_1, v_2)\|_{L_2(H, H_{-1/2})} < \infty \tag{2.106}
$$

show (2.89). The proof of (i) is thus complete. Furthermore, observe that (ii) follows directly from (i) and Remark 2.7. It thus remains to prove (iii). For this let $\varepsilon \in (4(1/2 - \gamma), 1 - 2\rho]$, $\beta \in (1/2, 2\gamma]$ and $\lambda_n \in \mathbb{R}$, $n \in \mathbb{N}$, be real numbers which satisfy for all $n \in \mathbb{N}$ that $\beta = 1/2 + (\varepsilon - 4(1/2-\gamma))/2$ and $\lambda_n = -\vartheta \pi^2 n^2$ and let $\mathbf{\Lambda}\colon D(\mathbf{\Lambda}) \subseteq H \times H_{-1/2} \to H \times H_{-1/2}$ be the linear operator which satisfies for all $(v, w) \in H_{1/2} \times H$ that $D(\mathbf{\Lambda}) = H_{1/2} \times H$ and

$$
\mathbf{\Lambda}(v, w) = \begin{pmatrix} \sum_{n=1}^\infty |\lambda_n|^{1/2} \langle e_n, v\rangle_H e_n \\ \sum_{n=1}^\infty |\lambda_n|^{1/2} \langle e_n, w\rangle_H e_n \end{pmatrix}. \tag{2.107}
$$

Then note that for all $v \in H_1$ it holds that $Av = \sum_{n=1}^\infty \lambda_n \langle e_n, v\rangle_H e_n$ and $\|\mathbf{\Lambda}^{-\beta}\|_{L_2(H \times H_{-1/2})} < \infty$. Furthermore, observe that (i) and the fact that $2\gamma - \beta = (1-\varepsilon)/2 \leq 1/2$ imply that $(H \times H_{-1/2} \ni v \mapsto \mathbf{F}(v) \in H_{2\gamma-\beta} \times H_{2\gamma-\beta-1/2}) \in \operatorname{Lip}^0(H \times H_{-1/2}, H_{2\gamma-\beta} \times H_{2\gamma-\beta-1/2})$. This, the fact that $2\rho \leq 1 - \varepsilon = 2(2\gamma - \beta)$, and again (i) enable us to apply Corollary 2.13 to obtain that there exists a real number $C \in [0, \infty)$ such that for all $N \in \mathbb{N}$ it holds that

$$
\big|\mathbb{E}\big[\varphi(X_T^\infty)\big] - \mathbb{E}\big[\varphi(X_T^N)\big]\big| \leq C\, |\lambda_{N+1}|^{\beta - 2\gamma} \leq C\, \vartheta^{(\varepsilon-1)/2} \cdot N^{\varepsilon - 1}. \tag{2.108}
$$

The proof of Corollary 2.16 is thus complete. □

In the proof of Corollary 2.18 below we employ the following elementary and well-known result, Lemma 2.17.

Lemma 2.17. *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, let $(H, \langle\cdot,\cdot\rangle_H, \|\cdot\|_H)$ be a $\mathbb{K}$-Hilbert space, let $\mathbb{H} \subseteq H$ be a non-empty orthonormal basis of H, let $A\colon D(A) \subseteq H \to H$ be a symmetric diagonal linear operator with $\inf(\sigma_{\mathrm{P}}(A)) > 0$, let $(H_r, \langle\cdot,\cdot\rangle_{H_r}, \|\cdot\|_{H_r})$, $r \in \mathbb{R}$, be a family of interpolation spaces associated to A, and let $q, s \in \mathbb{R}$, $p \in [q, \infty)$, $r \in [s, \infty)$. Then*

(i) for all $B \in L(H_q, H_s)$ it holds that $\big(B(H_q) \subseteq H_r$ and $(H_q \ni v \mapsto Bv \in H_r) \in L(H_q, H_r)\big)$ if and only if

$$
\left(B(\operatorname{span}_H(\mathbb{H})) \subseteq H_r \quad \text{and} \quad \sup_{w \in \operatorname{span}_H(\mathbb{H})\setminus\{0\}} \frac{\|Bw\|_{H_r}}{\|w\|_{H_q}} < \infty \right), \tag{2.109}
$$

(ii) *for all* $B \in L(H_q, H_s)$ *it holds that* $\big(B(H_q) \subseteq H_r$ *and* $(H_q \ni v \mapsto Bv \in H_r) \in L(H_q, H_r)\big)$ *if and only if*

$$\left(B(H_p) \subseteq H_r \quad \text{and} \quad \sup_{w \in H_p \setminus \{0\}} \frac{\|Bw\|_{H_r}}{\|w\|_{H_q}} < \infty \right), \tag{2.110}$$

and

(iii) *for all* $B \in L(H_q, H_r)$ *it holds that*

$$\|B\|_{L(H_q,H_r)} = \sup_{w \in \operatorname{span}_H(\mathbb{H}) \setminus \{0\}} \frac{\|Bw\|_{H_r}}{\|w\|_{H_q}} = \sup_{w \in H_p \setminus \{0\}} \frac{\|Bw\|_{H_r}}{\|w\|_{H_q}}. \tag{2.111}$$

Corollary 2.18. *Consider the notation in Subsection 2.1.1, let* $T, \vartheta \in (0,\infty)$, $\alpha, \beta \in \mathbb{R}$, *let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space with a normal filtration* $(\mathbb{F}_t)_{t\in[0,T]}$, *let* $(H, \langle\cdot,\cdot\rangle_H, \|\cdot\|_H) = \big(L^2(\mu_{(0,1)};\mathbb{R}), \langle\cdot,\cdot\rangle_{L^2(\mu_{(0,1)};\mathbb{R})}, \|\cdot\|_{L^2(\mu_{(0,1)};\mathbb{R})}\big)$, *let* $(W_t)_{t\in[0,T]}$ *be an* id_H*-cylindrical* $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$*-Wiener process, let* $(e_n)_{n\in\mathbb{N}} \subseteq H$ *satisfy for all* $n \in \mathbb{N}$ *and* $\mu_{(0,1)}$*-a.e.* $x \in (0,1)$ *that* $e_n(x) = \sqrt{2}\sin(n\pi x)$, *let* $A\colon D(A) \subseteq H \to H$ *be the Laplacian with Dirichlet boundary conditions on* H *multiplied by* ϑ, *let* $(H_r, \langle\cdot,\cdot\rangle_{H_r}, \|\cdot\|_{H_r})$, $r \in \mathbb{R}$, *be a family of interpolation spaces associated to* $-A$, *let* $\mathbf{P}_N \in L(H \times H_{-1/2})$, $N \in \mathbb{N} \cup \{\infty\}$, *be the linear operators which satisfy for all* $N \in \mathbb{N} \cup \{\infty\}$, $v, w \in H$ *that* $\mathbf{P}_N(v,w) = \big(\sum_{n=1}^N \langle e_n, v\rangle_H e_n, \sum_{n=1}^N \langle e_n, w\rangle_H e_n\big)$, *let* $\mathbf{A}\colon D(\mathbf{A}) \subseteq H \times H_{-1/2} \to H \times H_{-1/2}$ *be the linear operator which satisfies* $D(\mathbf{A}) = H_{1/2} \times H$ *and* $\forall (v,w) \in H_{1/2} \times H\colon \mathbf{A}(v,w) = (w, Av)$, *let* $\xi \in L^2(\mathbb{P}|_{\mathbb{F}_0}; H_{1/2} \times H)$, $\varphi \in C^2_{\mathrm{b}}(H \times H_{-1/2}, \mathbb{R})$, $f \in \operatorname{Lip}^2((0,1) \times \mathbb{R}, \mathbb{R})$, *and let* $\mathbf{F}\colon H \times H_{-1/2} \to H_{1/2} \times H$ *and* $\mathbf{B}\colon H \times H_{-1/2} \to L_2(H, H \times H_{-1/2})$ *be the functions which satisfy for all* $(v,w) \in H \times H_{-1/2}$, $u \in H_1$ *and* $\mu_{(0,1)}$*-a.e.* $x \in (0,1)$ *that* $\big(\mathbf{F}(v,w)\big)(x) = \big(0, f(x, v(x))\big)$ *and* $\big(\mathbf{B}(v,w)u\big)(x) = \big(0, (\alpha + \beta v(x))u(x)\big)$. *Then*

(i) *it holds that there exist up to modifications unique* $(\mathbb{F}_t)_{t\in[0,T]}$*-predictable stochastic processes* $X^N\colon [0,T] \times \Omega \to \big(\bigcap_{\rho\in[0,1/4)} \mathbf{P}_N(H_\rho \times H_{\rho-1/2})\big)$, $N \in \mathbb{N} \cup \{\infty\}$, *such that for all* $\rho \in [0, 1/4)$, $N \in \mathbb{N} \cup \{\infty\}$, $t \in [0,T]$ *it holds that* $\sup_{s\in[0,T]} \mathbb{E}\big[\|X^N_s\|^2_{H_\rho \times H_{\rho-1/2}}\big] < \infty$ *and* $\mathbb{P}$*-a.s. that*

$$X^N_t = e^{t\mathbf{A}}\mathbf{P}_N\xi + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_N\mathbf{F}(X^N_s)\,\mathrm{d}s + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{P}_N\mathbf{B}(X^N_s)\,\mathrm{d}W_s \tag{2.112}$$

and

(ii) *it holds for all* $\varepsilon \in (0,\infty)$ *that there exists a real number* $C \in [0,\infty)$ *such that for all* $N \in \mathbb{N}$ *it holds that*

$$\big|\mathbb{E}\big[\varphi(X^\infty_T)\big] - \mathbb{E}\big[\varphi(X^N_T)\big]\big| \le C \cdot N^{\varepsilon-1}. \tag{2.113}$$

Proof of Corollary 2.18. Throughout this proof let $B\colon H \to L_2(H, H_{-1/2})$ be the function which satisfies for all $v \in H$, $u \in H_1$ and $\mu_{(0,1)}$-a.e. $x \in (0,1)$ that $\big(B(v)u\big)(x) = (\alpha + \beta v(x))u(x)$. Note that it holds for all $\rho \in [0, 1/4)$, $v, u \in H$ that $B(v)u \in H_{\rho-1/2}$, $(H \ni y \mapsto B(v)y \in H_{\rho-1/2}) \in L_2(H, H_{\rho-1/2})$, and $(H \ni x \mapsto (H \ni y \mapsto B(x)y \in H_{\rho-1/2}) \in L_2(H, H_{\rho-1/2})) \in \operatorname{Lip}^0(H, L_2(H, H_{\rho-1/2}))$. Remark 2.7 and (i) in Corollary 2.16

thus prove (i). Next observe that the Sobolev embedding theorem proves for all $\rho \in (0, 1/4)$ that

$$\left[\sup_{w\in H_1\backslash\{0\}} \frac{\|w\|_{L^{1/(2\rho)}(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_{1/4-\rho}}}\right] + \left[\sup_{w\in H_\rho\backslash\{0\}} \frac{\|w\|_{L^{2/(1-4\rho)}(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_\rho}}\right] < \infty. \tag{2.114}$$

This and Hölder's inequality ensure for all $\rho \in (0, 1/4)$, $v \in H_\rho$, $u \in H_1$ that

$$\begin{aligned}
&\sup_{w\in H_1\backslash\{0\}} \frac{|\langle w, B(v)u\rangle_H|}{\|w\|_{H_{1/4-\rho}}} \\
&\leq \left[\sup_{w\in H_1\backslash\{0\}} \frac{\|w\|_{L^{1/(2\rho)}(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_{1/4-\rho}}}\right] \|\alpha + \beta v\|_{L^{2/(1-4\rho)}(\mu_{(0,1)};\mathbb{R})} \|u\|_{L^2(\mu_{(0,1)};\mathbb{R})} \\
&\leq \left[\sup_{w\in H_1\backslash\{0\}} \frac{\|w\|_{L^{1/(2\rho)}(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_{1/4-\rho}}}\right]\left[\sup_{w\in H_\rho\backslash\{0\}} \frac{\|w\|_{L^{2/(1-4\rho)}(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_\rho}}\right] \|\alpha + \beta v\|_{H_\rho} \|u\|_H < \infty.
\end{aligned} \tag{2.115}$$

Lemma 2.15 hence shows for all $\rho \in (0, 1/4)$, $v \in H_\rho$, $u \in H_1$ that $B(v)u \in H_{\rho-1/4}$. In addition, (2.115), Lemma 2.15, and Lemma 2.17 prove for all $\rho \in (0, 1/4)$, $v \in H_\rho$, $u \in H$ that $B(v)u \in H_{\rho-1/4}$ and $(H \ni y \mapsto B(v)y \in H_{\rho-1/4}) \in L(H, H_{\rho-1/4})$. Furthermore, Lemma 2.15 and Hölder's inequality show for all $\rho \in (0, 1/4)$, $v_1, v_2 \in H_\rho$, $u \in H_1$ that

$$\begin{aligned}
&\|(B(v_1) - B(v_2))u\|_{H_{\rho-1/4}} = \sup_{w\in H_1\backslash\{0\}} \frac{|\langle w, (B(v_1) - B(v_2))u\rangle_H|}{\|w\|_{H_{1/4-\rho}}} \\
&\leq \left[\sup_{w\in H_1\backslash\{0\}} \frac{\|w\|_{L^{1/(2\rho)}(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_{1/4-\rho}}}\right]\left[\sup_{w\in H_\rho\backslash\{0\}} \frac{\|w\|_{L^{2/(1-4\rho)}(\mu_{(0,1)};\mathbb{R})}}{\|w\|_{H_\rho}}\right] |\beta| \|v_1 - v_2\|_{H_\rho} \|u\|_H < \infty.
\end{aligned} \tag{2.116}$$

This and Lemma 2.17 establish for all $\varepsilon \in (0, 1]$, $\gamma \in (1/2 - \varepsilon/4, 1/2)$, $\rho \in [\gamma - 1/4, \min\{2\gamma - 1/2, 1/4\})$ that $(H_\rho \ni v \mapsto (H \ni u \mapsto B(v)u \in H_{\gamma-1/2}) \in L(H, H_{\gamma-1/2})) \in \operatorname{Lip}^0(H_\rho, L(H, H_{\gamma-1/2}))$. Corollary 2.16 thus completes the proof of Corollary 2.18. □

Chapter 3

Generalised multilevel Picard approximations

The content of this chapter is a slightly modified extract of the preprint Giles, Jentzen, & Welti [134].

In this chapter we develop an abstract framework for full-history recursive multilevel Picard (MLP) approximation methods and use this framework to study variants, which essentially are generalisations, of the MLP approximations for semi-linear heat equations introduced in Hutzenthaler et al. [181] (cf. Section 1.2 in Chapter 1). In particular, the main result of this chapter, Theorem 3.14 in Subsection 3.1.6, concerns the approximation error and, in a suitable sense, the computational complexity of generalised MLP methods, while Theorem 3.33 in Subsection 3.2.3 reveals that the variants for numerically approximating semi-linear heat equations have the power to beat the curse of dimensionality. The introductory results in Section 1.2, Theorem 1.2 and Corollary 1.3, are a consequence of Theorems 3.14 and 3.33, respectively, and provide similar conclusions under simplified assumptions.

This chapter is structured in the following way. Section 3.1 is devoted to the abstract framework of generalised MLP approximations. In particular, we study several elementary but crucial properties of such approximations in Proposition 3.8 in Subsection 3.1.3. Moreover, we derive in Subsection 3.1.4 an error analysis for generalised MLP approximations, which relies on suitably generalised versions of well-known identities involving bias and variance in Hilbert spaces (cf. Corollaries 3.5 and 3.7 in Subsection 3.1.2). This error analysis is subsequently combined with the cost analysis in Subsection 3.1.5 to establish a complexity analysis for generalised MLP approximations in Subsection 3.1.6 (cf. Theorem 3.14 and Corollary 3.15 in Subsection 3.1.6). Throughout Section 3.1 the measurability results in Subsection 3.1.1 are used. In Section 3.2 we employ the abstract framework for generalised MLP approximations from Section 3.1 to analyse numerical approximations for semi-linear heat equations. Subsection 3.2.1 collects several elementary and well-known auxiliary results, which are used in Subsection 3.2.2 to verify that the main assumptions of the abstract complexity result in Corollary 3.15 are fulfilled in the case of the example setting for numerical approximations for semi-linear heat equations. Finally, in Subsection 3.2.3 we combine the results from Subsection 3.2.2 with Corollary 3.15 to obtain a complexity analysis for MLP approximations for semi-linear heat equations (cf. Proposition 3.32, Theorem 3.33, and Corollary 3.34 in Subsection 3.2.3).

3.1 Generalised MLP approximations

In this section we introduce generalised MLP approximations and provide an error analysis (cf. Subsection 3.1.4), cost analysis (cf. Subsection 3.1.5), and complexity analysis (cf. Subsection 3.1.6) for such approximations.

For the formulation of the error analysis for generalised MLP approximations we require random variables which take values in the Banach space $L(\mathcal{Y}, \mathcal{H})$ of continuous linear functions between a separable Banach space $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ and a separable Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}}, \|\cdot\|_{\mathcal{H}})$ equipped with the strong σ-algebra. Let us recall that the strong σ-algebra on $L(\mathcal{Y}, \mathcal{H})$ is nothing but the trace σ-algebra of the product σ-algebra on the set $\mathcal{H}^{\mathcal{Y}}$ of all functions from $\mathcal{Y}$ to $\mathcal{H}$. Having this in mind, Subsection 3.1.1 collects three elementary measurability results (cf. Lemmas 3.1 and 3.2 and Corollary 3.3) about functions whose domains or codomains involve a set of functions equipped with the trace σ-algebra of the product σ-algebra.

In Subsection 3.1.2 we first recall the elementary and well-known bias–variance decomposition of the mean square error for random variables that take values in a separable Hilbert space (cf. Lemma 3.4). Thereafter, we present in Corollary 3.5 a generalised bias–variance decomposition, where the mean square error, the bias, and the variance are measured in a certain randomised sense. Analogously, we also recall the elementary and well-known result that the mean square norm of the sum of independent zero mean random variables in a separable Hilbert space is equal to the sum of the individual mean square norms (cf. Lemma 3.6) and prove a randomised generalisation thereof (cf. Corollary 3.7). This generalisation as well as the generalised bias–variance decomposition in Corollary 3.5 are used in our error analysis for generalised MLP approximations (cf. (3.38), (3.39), and (3.42) in the proof of Proposition 3.9).

Subsequently, Proposition 3.8 in Subsection 3.1.3 establishes several elementary but crucial properties of generalised MLP approximations, which are a consequence of their definition.

Subsection 3.1.4 is devoted to the error analysis for generalised MLP approximations. Proposition 3.9 specifies the most general hypotheses in this thesis (cf. (3.29)–(3.32) in Proposition 3.9) under which we prove an error estimate for generalised MLP approximations. The upper bound for the error in Proposition 3.9 (cf. (3.33) in Proposition 3.9) can be much shortened by choosing a natural number $\mathfrak{M} \in \mathbb{N} = \{1, 2, 3, \ldots\}$ such that for every $n \in \mathbb{N}$, $l \in \{0, 1, \ldots, n-1\}$ the Monte Carlo sample number $M_{n,l} \in \mathbb{N}$ in the generalised MLP approximations (cf. (3.29) in Proposition 3.9) is equal to $\mathfrak{M}^{n-l}$, which is the assertion of Corollary 3.10 (cf. (3.54) in Corollary 3.10).

The subject of Subsection 3.1.5 is the cost analysis for generalised MLP approximations. The cost estimate in Proposition 3.11 follows from an application of the discrete Gronwall-type inequality in Agarwal [2, Theorem 4.1.1]. The second cost estimate in Subsection 3.1.5 (cf. Corollary 3.13), in turn, is a consequence of Proposition 3.11 and the elementary and well-known estimate in Lemma 3.12.

In Subsection 3.1.6 the error analysis from Subsection 3.1.4 and the cost analysis from Subsection 3.1.5 are combined to derive a complexity analysis for generalised MLP approximations. More precisely, the main result of this chapter, Theorem 3.14, relates the error estimate in Corollary 3.10 to the cost estimate in Corollary 3.13 in order to arrive at a complexity estimate (cf. (3.81) in Theorem 3.14). The subsequent result, Corollary 3.15, is obtained by replacing assumption (3.78) in Theorem 3.14 by the simpler assumption (3.93) and choosing for every $k \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}$ the coefficient $\mathfrak{c}_k \in (0, \infty)$ in Theorem 3.14 to be equal to $k!$. Finally, the elementary result in Lemma 3.16 shows that a strictly increasing and at most linearly growing sequence of natural numbers automatically fulfils the hypotheses on the sequence $(M_j)_{j \in \mathbb{N}} \subseteq \mathbb{N}$ in Corollary 3.15.

3.1.1 Measurability involving the strong σ-algebra

Lemma 3.1. *Let $\mathcal{E}$ be a set, let $(\mathcal{F}, \mathscr{F})$ and $(\mathcal{G}, \mathscr{G})$ be measurable spaces, let $\mathcal{S} \subseteq \mathcal{F}^{\mathcal{E}}$, let $\mathscr{S} = \sigma_{\mathcal{S}}\big(\{\{\varphi \in \mathcal{S} \colon \varphi(e) \in \mathcal{A}\} \subseteq \mathcal{S} \colon e \in \mathcal{E}, \mathcal{A} \in \mathscr{F}\}\big)$, and let $\psi \colon \mathcal{G} \to \mathcal{S}$ be a function. Then it holds that ψ is $\mathscr{G}/\mathscr{S}$-measurable if and only if it holds for all $e \in \mathcal{E}$ that $\mathcal{G} \ni \omega \mapsto [\psi(\omega)](e) \in \mathcal{F}$ is $\mathscr{G}/\mathscr{F}$-measurable.*

Proof of Lemma 3.1. Throughout this proof let $P_e \colon \mathcal{S} \to \mathcal{F}$, $e \in \mathcal{E}$, satisfy for all $e \in \mathcal{E}$, $\varphi \in \mathcal{S}$ that $P_e(\varphi) = \varphi(e)$. Observe that it holds that

$$\mathscr{S} = \{(\mathcal{A} \cap \mathcal{S}) \subseteq \mathcal{S} \colon \mathcal{A} \in (\otimes_{e \in \mathcal{E}} \mathscr{F})\} = \sigma_{\mathcal{S}}\big((P_e)_{e \in \mathcal{E}}\big). \tag{3.1}$$

This ensures for all $e \in \mathcal{E}$ that $P_e \colon \mathcal{S} \to \mathcal{F}$ is an $\mathscr{S}/\mathscr{F}$-measurable function. Equation (3.1) hence shows that $\psi \colon \mathcal{G} \to \mathcal{S}$ is $\mathscr{G}/\mathscr{S}$-measurable if and only if it holds for all $e \in \mathcal{E}$ that $P_e \circ \psi \colon \mathcal{G} \to \mathcal{F}$ is $\mathscr{G}/\mathscr{F}$-measurable. The proof of Lemma 3.1 is thus complete. □

Lemma 3.2. *Let $(\mathcal{E}, d_{\mathcal{E}})$ be a separable metric space, let $(\mathcal{F}, d_{\mathcal{F}})$ be a metric space, let $\mathcal{S} \subseteq C(\mathcal{E}, \mathcal{F})$, and let $\mathscr{S} = \sigma_{\mathcal{S}}\big(\{\{\varphi \in \mathcal{S} \colon \varphi(e) \in \mathcal{A}\} \subseteq \mathcal{S} \colon e \in \mathcal{E}, \mathcal{A} \in \mathscr{B}(\mathcal{F})\}\big)$. Then it holds that $\mathcal{S} \times \mathcal{E} \ni (\varphi, e) \mapsto \varphi(e) \in \mathcal{F}$ is an $(\mathscr{S} \otimes \mathscr{B}(\mathcal{E}))/\mathscr{B}(\mathcal{F})$-measurable function.*

Proof of Lemma 3.2. Throughout this proof let $f \colon \mathcal{S} \times \mathcal{E} \to \mathcal{F}$ satisfy for all $\varphi \in \mathcal{S}$, $e \in \mathcal{E}$ that $f(\varphi, e) = \varphi(e)$. Note that it holds for all $\varphi \in \mathcal{S}$ that

$$(\mathcal{E} \ni e \mapsto f(\varphi, e) \in \mathcal{F}) = \varphi \in \mathcal{S} \subseteq C(\mathcal{E}, \mathcal{F}). \tag{3.2}$$

In addition, observe that it holds for all $e \in \mathcal{E}$, $\mathcal{A} \in \mathscr{B}(\mathcal{F})$ that

$$\{\varphi \in \mathcal{S} \colon f(\varphi, e) \in \mathcal{A}\} = \{\varphi \in \mathcal{S} \colon \varphi(e) \in \mathcal{A}\} \in \mathscr{S}. \tag{3.3}$$

This proves for all $e \in \mathcal{E}$ that $\mathcal{S} \ni \varphi \mapsto f(\varphi, e) \in \mathcal{F}$ is an $\mathscr{S}/\mathscr{B}(\mathcal{F})$-measurable function. Combining this and (3.2) with Aliprantis & Border [4, Lemma 4.51] (see also, e.g., Beck et al. [22, Lemma 2.4]) establishes that $f \colon \mathcal{S} \times \mathcal{E} \to \mathcal{F}$ is an $(\mathscr{S} \otimes \mathscr{B}(\mathcal{E}))/\mathscr{B}(\mathcal{F})$-measurable function. The proof of Lemma 3.2 is thus complete. □

Corollary 3.3. *Let $(\Omega, \mathscr{F})$ be a measurable space, let $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ be a separable normed $\mathbb{R}$-vector space, let $Y \colon \Omega \to \mathcal{V}$ be an $\mathscr{F}/\mathscr{B}(\mathcal{V})$-measurable function, let $(\mathcal{W}, \|\cdot\|_{\mathcal{W}})$ be a normed $\mathbb{R}$-vector space, let $\mathscr{S} = \sigma_{L(\mathcal{V},\mathcal{W})}\big(\{\{\varphi \in L(\mathcal{V}, \mathcal{W}) \colon \varphi(v) \in \mathcal{B}\} \subseteq L(\mathcal{V}, \mathcal{W}) \colon v \in \mathcal{V}, \mathcal{B} \in \mathscr{B}(\mathcal{W})\}\big)$, and let $\psi \colon \Omega \to L(\mathcal{V}, \mathcal{W})$ be an $\mathscr{F}/\mathscr{S}$-measurable function. Then*

(i) it holds that $L(\mathcal{V}, \mathcal{W}) \times \mathcal{V} \ni (\varphi, v) \mapsto \varphi(v) \in \mathcal{W}$ is an $(\mathscr{S} \otimes \mathscr{B}(\mathcal{V}))/\mathscr{B}(\mathcal{W})$-measurable function and

(ii) it holds that $\psi(Y) = (\Omega \ni \omega \mapsto [\psi(\omega)](Y(\omega)) \in \mathcal{W})$ is an $\mathscr{F}/\mathscr{B}(\mathcal{W})$-measurable function.

Proof of Corollary 3.3. Observe that Lemma 3.2 (with $\mathcal{E} \leftarrow \mathcal{V}$, $\mathcal{F} \leftarrow \mathcal{W}$, $\mathcal{S} \leftarrow L(\mathcal{V}, \mathcal{W})$ in the notation of Lemma 3.2) implies (i). In addition, the fact that $\Omega \ni \omega \mapsto (\psi(\omega), Y(\omega)) \in L(\mathcal{V}, \mathcal{W}) \times \mathcal{V}$ is an $\mathscr{F}/(\mathscr{S} \otimes \mathscr{B}(\mathcal{V}))$-measurable function and (i) show (ii). The proof of Corollary 3.3 is thus complete. □

3.1.2 Identities involving bias and variance in Hilbert spaces

3.1.2.1 Bias–variance decomposition

Lemma 3.4. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}}, \|\cdot\|_{\mathcal{H}})$ be a separable $\mathbb{R}$-Hilbert space, let $h \in \mathcal{H}$, and let $X\colon \Omega \to \mathcal{H}$ be an $\mathscr{F}/\mathscr{B}(\mathcal{H})$-measurable function which satisfies $\mathbb{E}[\|X\|_{\mathcal{H}}] < \infty$. Then*

$$\mathbb{E}\big[\|X - h\|_{\mathcal{H}}^2\big] = \mathbb{E}\big[\|X - \mathbb{E}[X]\|_{\mathcal{H}}^2\big] + \|\mathbb{E}[X] - h\|_{\mathcal{H}}^2. \tag{3.4}$$

Proof of Lemma 3.4. Note that the Cauchy–Schwarz inequality implies that

$$\mathbb{E}\big[|\langle X - \mathbb{E}[X], \mathbb{E}[X] - h\rangle_{\mathcal{H}}|\big] \le \mathbb{E}\big[\|X - \mathbb{E}[X]\|_{\mathcal{H}}\big]\,\|\mathbb{E}[X] - h\|_{\mathcal{H}} < \infty. \tag{3.5}$$

This ensures that

$$\begin{aligned}
\mathbb{E}\big[\|X - h\|_{\mathcal{H}}^2\big] &= \mathbb{E}\big[\|X - \mathbb{E}[X] + \mathbb{E}[X] - h\|_{\mathcal{H}}^2\big] \\
&= \mathbb{E}\big[\|X - \mathbb{E}[X]\|_{\mathcal{H}}^2 + \|\mathbb{E}[X] - h\|_{\mathcal{H}}^2 + 2\langle X - \mathbb{E}[X], \mathbb{E}[X] - h\rangle_{\mathcal{H}}\big] \\
&= \mathbb{E}\big[\|X - \mathbb{E}[X]\|_{\mathcal{H}}^2\big] + \|\mathbb{E}[X] - h\|_{\mathcal{H}}^2 + 2\langle \mathbb{E}[X] - \mathbb{E}[X], \mathbb{E}[X] - h\rangle_{\mathcal{H}} \\
&= \mathbb{E}\big[\|X - \mathbb{E}[X]\|_{\mathcal{H}}^2\big] + \|\mathbb{E}[X] - h\|_{\mathcal{H}}^2.
\end{aligned} \tag{3.6}$$

The proof of Lemma 3.4 is thus complete. ☐

3.1.2.2 Generalised bias–variance decomposition

Corollary 3.5. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be a separable $\mathbb{R}$-Banach space, let $y \in \mathcal{Y}$, let $Y\colon \Omega \to \mathcal{Y}$ be an $\mathscr{F}/\mathscr{B}(\mathcal{Y})$-measurable function which satisfies $\mathbb{E}[\|Y\|_{\mathcal{Y}}] < \infty$, let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}}, \|\cdot\|_{\mathcal{H}})$ be a separable $\mathbb{R}$-Hilbert space, let $\mathscr{S} = \sigma_{L(\mathcal{Y},\mathcal{H})}\big(\{\{\varphi \in L(\mathcal{Y},\mathcal{H})\colon \varphi(x) \in \mathcal{B}\} \subseteq L(\mathcal{Y},\mathcal{H})\colon x \in \mathcal{Y}, \mathcal{B} \in \mathscr{B}(\mathcal{H})\}\big)$, let $\psi\colon \Omega \to L(\mathcal{Y},\mathcal{H})$ be an $\mathscr{F}/\mathscr{S}$-measurable function, and assume that Y and ψ are independent. Then*

$$\mathbb{E}\big[\|\psi(Y - y)\|_{\mathcal{H}}^2\big] = \mathbb{E}\big[\|\psi(Y - \mathbb{E}[Y])\|_{\mathcal{H}}^2\big] + \mathbb{E}\big[\|\psi(\mathbb{E}[Y] - y)\|_{\mathcal{H}}^2\big] \tag{3.7}$$

(cf. (ii) in Corollary 3.3).

Proof of Corollary 3.5. The fact that it holds for all $x \in \mathcal{Y}$ that $L(\mathcal{Y},\mathcal{H}) \times \Omega \ni (\varphi, \omega) \mapsto (\varphi, Y(\omega) - x) \in L(\mathcal{Y},\mathcal{H}) \times \mathcal{Y}$ is an $(\mathscr{S} \otimes \sigma_\Omega(Y))/(\mathscr{S} \otimes \mathscr{B}(\mathcal{Y}))$-measurable function and the fact that $L(\mathcal{Y},\mathcal{H}) \times \mathcal{Y} \ni (\varphi, x) \mapsto \varphi(x) \in \mathcal{H}$ is an $(\mathscr{S} \otimes \mathscr{B}(\mathcal{Y}))/\mathscr{B}(\mathcal{H})$-measurable function (cf. (i) in Corollary 3.3) imply for all $x \in \mathcal{Y}$ that

$$L(\mathcal{Y},\mathcal{H}) \times \Omega \ni (\varphi, \omega) \mapsto \varphi(Y(\omega) - x) \in \mathcal{H} \tag{3.8}$$

is an $(\mathscr{S} \otimes \sigma_\Omega(Y))/\mathscr{B}(\mathcal{H})$-measurable function. Lemma 2.2 in Hutzenthaler et al. [181] (with $\mathcal{G} \leftarrow \sigma_\Omega(Y)$, $(S, \mathcal{S}) \leftarrow (L(\mathcal{Y},\mathcal{H}), \mathscr{S})$, $U \leftarrow (L(\mathcal{Y},\mathcal{H}) \times \Omega \ni (\varphi, \omega) \mapsto \|\varphi(Y(\omega) - y)\|_{\mathcal{H}}^2 \in [0,\infty))$, $Y \leftarrow \psi$ in the notation of [181, Lemma 2.2]) and Lemma 3.4 hence yield that

$$\begin{aligned}
\mathbb{E}\big[\|\psi(Y - y)\|_{\mathcal{H}}^2\big] &= \int_{L(\mathcal{Y},\mathcal{H})} \mathbb{E}\big[\|\varphi(Y - y)\|_{\mathcal{H}}^2\big]\,(\psi(\mathbb{P})_{\mathscr{S}})(\mathrm{d}\varphi) \\
&= \int_{L(\mathcal{Y},\mathcal{H})} \mathbb{E}\big[\|\varphi(Y) - \varphi(y)\|_{\mathcal{H}}^2\big]\,(\psi(\mathbb{P})_{\mathscr{S}})(\mathrm{d}\varphi) \\
&= \int_{L(\mathcal{Y},\mathcal{H})} \mathbb{E}\big[\|\varphi(Y) - \mathbb{E}[\varphi(Y)]\|_{\mathcal{H}}^2\big] + \|\mathbb{E}[\varphi(Y)] - \varphi(y)\|_{\mathcal{H}}^2\,(\psi(\mathbb{P})_{\mathscr{S}})(\mathrm{d}\varphi).
\end{aligned} \tag{3.9}$$

This, the fact that $L(\mathcal{Y},\mathcal{H}) \ni \varphi \mapsto \varphi(\mathbb{E}[Y]-y) \in \mathcal{H}$ is an $\mathscr{S}/\mathscr{B}(\mathcal{H})$-measurable function (cf. Lemma 3.1), (3.8), and again [181, Lemma 2.2] (with $\mathcal{G} \leftarrow \sigma_\Omega(Y)$, $(S,\mathcal{S}) \leftarrow (L(\mathcal{Y},\mathcal{H}),\mathscr{S})$, $U \leftarrow (L(\mathcal{Y},\mathcal{H}) \times \Omega \ni (\varphi,\omega) \mapsto \|\varphi(Y(\omega)-\mathbb{E}[Y])\|_{\mathcal{H}}^2 \in [0,\infty))$, $Y \leftarrow \psi$ in the notation of [181, Lemma 2.2]) establish that

$$
\begin{split}
\mathbb{E}\big[\|\psi(Y-y)\|_{\mathcal{H}}^2\big] &= \int_{L(\mathcal{Y},\mathcal{H})} \mathbb{E}\big[\|\varphi(Y-\mathbb{E}[Y])\|_{\mathcal{H}}^2\big]\,(\psi(\mathbb{P})_{\mathscr{S}})(\mathrm{d}\varphi) \\
&\quad + \int_{L(\mathcal{Y},\mathcal{H})} \|\varphi(\mathbb{E}[Y]-y)\|_{\mathcal{H}}^2\,(\psi(\mathbb{P})_{\mathscr{S}})(\mathrm{d}\varphi) \\
&= \mathbb{E}\big[\|\psi(Y-\mathbb{E}[Y])\|_{\mathcal{H}}^2\big] + \mathbb{E}\big[\|\psi(\mathbb{E}[Y]-y)\|_{\mathcal{H}}^2\big].
\end{split}
\tag{3.10}
$$

The proof of Corollary 3.5 is thus complete. □

3.1.2.3 Variance identity

Lemma 3.6. *Let $n \in \mathbb{N}$, let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space, let $(\mathcal{H},\langle\cdot,\cdot\rangle_{\mathcal{H}},\|\cdot\|_{\mathcal{H}})$ be a separable $\mathbb{R}$-Hilbert space, and let $X_1,X_2,\ldots,X_n\colon\Omega\to\mathcal{H}$ be independent $\mathscr{F}/\mathscr{B}(\mathcal{H})$-measurable functions which satisfy for all $i\in\{1,2,\ldots,n\}$ that $\mathbb{E}[\|X_i\|_{\mathcal{H}}]<\infty$. Then*

$$
\mathbb{E}\Bigg[\bigg\|\sum_{i=1}^n (X_i-\mathbb{E}[X_i])\bigg\|_{\mathcal{H}}^2\Bigg] = \sum_{i=1}^n \mathbb{E}\big[\|X_i-\mathbb{E}[X_i]\|_{\mathcal{H}}^2\big]. \tag{3.11}
$$

Proof of Lemma 3.6. Observe that the Cauchy–Schwarz inequality, the fact that $X_1,X_2,\ldots,X_n$ are independent, and the fact that it holds for all independent random variables $Y,Z\colon\Omega\to\mathbb{R}$ with $\mathbb{E}[|Y|+|Z|]<\infty$ that $\mathbb{E}[|YZ|]<\infty$ and $\mathbb{E}[YZ]=\mathbb{E}[Y]\mathbb{E}[Z]$ (cf., e.g., Klenke [202, Theorem 5.4]) demonstrate for all $i,j\in\{1,2,\ldots,n\}$ with $i\neq j$ that

$$
\begin{split}
\mathbb{E}\big[|\langle X_i-\mathbb{E}[X_i],X_j-\mathbb{E}[X_j]\rangle_{\mathcal{H}}|\big] &\le \mathbb{E}\big[\|X_i-\mathbb{E}[X_i]\|_{\mathcal{H}}\|X_j-\mathbb{E}[X_j]\|_{\mathcal{H}}\big] \\
&= \mathbb{E}\big[\|X_i-\mathbb{E}[X_i]\|_{\mathcal{H}}\big]\mathbb{E}\big[\|X_j-\mathbb{E}[X_j]\|_{\mathcal{H}}\big] < \infty.
\end{split}
\tag{3.12}
$$

Moreover, the fact that $X_1,X_2,\ldots,X_n$ are independent ensures for all $i,j\in\{1,2,\ldots,n\}$ with $i\neq j$ that

$$
(X_i,X_j)(\mathbb{P})_{\mathscr{B}(\mathcal{H})\otimes\mathscr{B}(\mathcal{H})} = [(X_i)(\mathbb{P})_{\mathscr{B}(\mathcal{H})}]\otimes[(X_j)(\mathbb{P})_{\mathscr{B}(\mathcal{H})}]. \tag{3.13}
$$

Fubini's theorem and (3.12) hence show for all $i,j\in\{1,2,\ldots,n\}$ with $i\neq j$ that

$$
\begin{split}
&\mathbb{E}\big[\langle X_i-\mathbb{E}[X_i],X_j-\mathbb{E}[X_j]\rangle_{\mathcal{H}}\big] \\
&= \int_{\mathcal{H}\times\mathcal{H}} \langle x-\mathbb{E}[X_i],y-\mathbb{E}[X_j]\rangle_{\mathcal{H}}\,\big((X_i,X_j)(\mathbb{P})_{\mathscr{B}(\mathcal{H})\otimes\mathscr{B}(\mathcal{H})}\big)(\mathrm{d}x,\mathrm{d}y) \\
&= \int_{\mathcal{H}\times\mathcal{H}} \langle x-\mathbb{E}[X_i],y-\mathbb{E}[X_j]\rangle_{\mathcal{H}}\,\big([(X_i)(\mathbb{P})_{\mathscr{B}(\mathcal{H})}]\otimes[(X_j)(\mathbb{P})_{\mathscr{B}(\mathcal{H})}]\big)(\mathrm{d}x,\mathrm{d}y) \\
&= \int_{\mathcal{H}}\int_{\mathcal{H}} \langle x-\mathbb{E}[X_i],y-\mathbb{E}[X_j]\rangle_{\mathcal{H}}\,\big((X_i)(\mathbb{P})_{\mathscr{B}(\mathcal{H})}\big)(\mathrm{d}x)\big((X_j)(\mathbb{P})_{\mathscr{B}(\mathcal{H})}\big)(\mathrm{d}y) \\
&= \int_{\mathcal{H}} \mathbb{E}\big[\langle X_i-\mathbb{E}[X_i],y-\mathbb{E}[X_j]\rangle_{\mathcal{H}}\big]\,\big((X_j)(\mathbb{P})_{\mathscr{B}(\mathcal{H})}\big)(\mathrm{d}y) \\
&= \int_{\mathcal{H}} \langle \mathbb{E}[X_i]-\mathbb{E}[X_i],y-\mathbb{E}[X_j]\rangle_{\mathcal{H}}\,\big((X_j)(\mathbb{P})_{\mathscr{B}(\mathcal{H})}\big)(\mathrm{d}y) \\
&= \langle \mathbb{E}[X_i]-\mathbb{E}[X_i],\mathbb{E}[X_j]-\mathbb{E}[X_j]\rangle_{\mathcal{H}} = 0.
\end{split}
\tag{3.14}
$$

This and again (3.12) prove that

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{i=1}^{n}(X_i-\mathbb{E}[X_i])\right\|_{\mathcal{H}}^2\right] &= \mathbb{E}\left[\left\langle\sum_{i=1}^{n}(X_i-\mathbb{E}[X_i]),\sum_{j=1}^{n}(X_j-\mathbb{E}[X_j])\right\rangle_{\mathcal{H}}\right] \\
&= \mathbb{E}\left[\sum_{i,j=1}^{n}\langle X_i-\mathbb{E}[X_i],X_j-\mathbb{E}[X_j]\rangle_{\mathcal{H}}\right] \\
&= \left[\sum_{i=1}^{n}\mathbb{E}\big[\langle X_i-\mathbb{E}[X_i],X_i-\mathbb{E}[X_i]\rangle_{\mathcal{H}}\big]\right] + \sum_{i,j=1,\,i\neq j}^{n}\mathbb{E}\big[\langle X_i-\mathbb{E}[X_i],X_j-\mathbb{E}[X_j]\rangle_{\mathcal{H}}\big] \\
&= \sum_{i=1}^{n}\mathbb{E}\big[\|X_i-\mathbb{E}[X_i]\|_{\mathcal{H}}^2\big].
\end{aligned} \tag{3.15}$$

The proof of Lemma 3.6 is thus complete. □

3.1.2.4 Generalised variance identity

Corollary 3.7. *Let $n\in\mathbb{N}$, let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space, let $(\mathcal{Y},\|\cdot\|_{\mathcal{Y}})$ be a separable $\mathbb{R}$-Banach space, let $Y_1,Y_2,\dots,Y_n\colon\Omega\to\mathcal{Y}$ be independent $\mathscr{F}/\mathscr{B}(\mathcal{Y})$-measurable functions which satisfy for all $i\in\{1,2,\dots,n\}$ that $\mathbb{E}[\|Y_i\|_{\mathcal{Y}}]<\infty$, let $(\mathcal{H},\langle\cdot,\cdot\rangle_{\mathcal{H}},\|\cdot\|_{\mathcal{H}})$ be a separable $\mathbb{R}$-Hilbert space, let $\mathscr{S}=\sigma_{L(\mathcal{Y},\mathcal{H})}\big(\big\{\{\varphi\in L(\mathcal{Y},\mathcal{H})\colon\varphi(y)\in\mathcal{B}\}\subseteq L(\mathcal{Y},\mathcal{H})\colon y\in\mathcal{Y},\,\mathcal{B}\in\mathscr{B}(\mathcal{H})\big\}\big)$, let $\psi\colon\Omega\to L(\mathcal{Y},\mathcal{H})$ be an $\mathscr{F}/\mathscr{S}$-measurable function, and assume that $(Y_i)_{i\in\{1,2,\dots,n\}}$ and ψ are independent. Then*

$$\mathbb{E}\left[\left\|\sum_{i=1}^{n}\psi(Y_i-\mathbb{E}[Y_i])\right\|_{\mathcal{H}}^2\right]=\sum_{i=1}^{n}\mathbb{E}\big[\|\psi(Y_i-\mathbb{E}[Y_i])\|_{\mathcal{H}}^2\big] \tag{3.16}$$

(cf. (ii) in Corollary 3.3).

Proof of Corollary 3.7. The fact that it holds for all $i\in\{1,2,\dots,n\}$ that $L(\mathcal{Y},\mathcal{H})\times\Omega\ni(\varphi,\omega)\mapsto(\varphi,Y_i(\omega)-\mathbb{E}[Y_i])\in L(\mathcal{Y},\mathcal{H})\times\mathcal{Y}$ is an $(\mathscr{S}\otimes\sigma_\Omega((Y_j)_{j\in\{1,2,\dots,n\}}))/(\mathscr{S}\otimes\mathscr{B}(\mathcal{Y}))$-measurable function and the fact that $L(\mathcal{Y},\mathcal{H})\times\mathcal{Y}\ni(\varphi,y)\mapsto\varphi(y)\in\mathcal{H}$ is an $(\mathscr{S}\otimes\mathscr{B}(\mathcal{Y}))/\mathscr{B}(\mathcal{H})$-measurable function (cf. (i) in Corollary 3.3) ensure for all $i\in\{1,2,\dots,n\}$ that

$$L(\mathcal{Y},\mathcal{H})\times\Omega\ni(\varphi,\omega)\mapsto\varphi(Y_i(\omega)-\mathbb{E}[Y_i])\in\mathcal{H} \tag{3.17}$$

is an $(\mathscr{S}\otimes\sigma_\Omega((Y_j)_{j\in\{1,2,\dots,n\}}))/\mathscr{B}(\mathcal{H})$-measurable function. This and [181, Lemma 2.2] (with $\mathcal{G}\leftarrow\sigma_\Omega((Y_i)_{i\in\{1,2,\dots,n\}})$, $(S,\mathcal{S})\leftarrow(L(\mathcal{Y},\mathcal{H}),\mathscr{S})$, $U\leftarrow(L(\mathcal{Y},\mathcal{H})\times\Omega\ni(\varphi,\omega)\mapsto\|\sum_{i\leftarrow1}^{n}\varphi(Y_i(\omega)-\mathbb{E}[Y_i])\|_{\mathcal{H}}^2\in[0,\infty))$, $Y\leftarrow\psi$ in the notation of [181, Lemma 2.2]) establish that

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{i=1}^{n}\psi(Y_i-\mathbb{E}[Y_i])\right\|_{\mathcal{H}}^2\right] &= \int_{L(\mathcal{Y},\mathcal{H})}\mathbb{E}\Big[\big\|\textstyle\sum_{i=1}^{n}\varphi(Y_i-\mathbb{E}[Y_i])\big\|_{\mathcal{H}}^2\Big]\,(\psi(\mathbb{P})_{\mathscr{S}})(d\varphi) \\
&= \int_{L(\mathcal{Y},\mathcal{H})}\mathbb{E}\Big[\big\|\textstyle\sum_{i=1}^{n}(\varphi(Y_i)-\mathbb{E}[\varphi(Y_i)])\big\|_{\mathcal{H}}^2\Big]\,(\psi(\mathbb{P})_{\mathscr{S}})(d\varphi).
\end{aligned} \tag{3.18}$$

Lemma 3.6, (3.17), and [181, Lemma 2.2] (with $\mathcal{G}\leftarrow\sigma_\Omega((Y_j)_{j\in\{1,2,\dots,n\}})$, $(S,\mathcal{S})\leftarrow(L(\mathcal{Y},\mathcal{H}),\mathscr{S})$, $U\leftarrow(L(\mathcal{Y},\mathcal{H})\times\Omega\ni(\varphi,\omega)\mapsto\|\varphi(Y_i(\omega)-\mathbb{E}[Y_i])\|_{\mathcal{H}}^2\in[0,\infty))$, $Y\leftarrow\psi$ for $i\in$

$\{1, 2, \ldots, n\}$ in the notation of [181, Lemma 2.2]) hence show that

$$\begin{aligned}\mathbb{E}\left[\left\|\sum_{i=1}^{n} \psi(Y_i - \mathbb{E}[Y_i])\right\|_{\mathcal{H}}^2\right] &= \int_{L(\mathcal{Y},\mathcal{H})} \sum_{i=1}^{n} \mathbb{E}\big[\|\varphi(Y_i) - \mathbb{E}[\varphi(Y_i)]\|_{\mathcal{H}}^2\big] \, (\psi(\mathbb{P})_{\mathscr{S}})(\mathrm{d}\varphi) \\ &= \sum_{i=1}^{n} \int_{L(\mathcal{Y},\mathcal{H})} \mathbb{E}\big[\|\varphi(Y_i - \mathbb{E}[Y_i])\|_{\mathcal{H}}^2\big] \, (\psi(\mathbb{P})_{\mathscr{S}})(\mathrm{d}\varphi) \\ &= \sum_{i=1}^{n} \mathbb{E}\big[\|\psi(Y_i - \mathbb{E}[Y_i])\|_{\mathcal{H}}^2\big].\end{aligned} \tag{3.19}$$

The proof of Corollary 3.7 is thus complete. □

3.1.3 Properties of generalised MLP approximations

Proposition 3.8. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be a separable $\mathbb{R}$-Banach space, let $\Theta = \bigcup_{n=1}^{\infty} \mathbb{Z}^n$, $(M_{n,l})_{(n,l)\in\mathbb{N}\times\mathbb{N}_0} \subseteq \mathbb{N}$, let $(\mathcal{Z}, \mathscr{Z})$ be a measurable space, let $Z^\theta\colon \Omega \to \mathcal{Z}$, $\theta \in \Theta$, be i.i.d. $\mathscr{F}/\mathscr{Z}$-measurable functions, let $\Phi_l\colon \mathcal{Y}\times\mathcal{Y}\times\mathcal{Z} \to \mathcal{Y}$, $l \in \mathbb{N}_0$, be $(\mathscr{B}(\mathcal{Y}) \otimes \mathscr{B}(\mathcal{Y}) \otimes \mathscr{Z})/\mathscr{B}(\mathcal{Y})$-measurable functions, let $Y_{-1}^\theta\colon \Omega \to \mathcal{Y}$, $\theta \in \Theta$, be i.i.d. $\mathscr{F}/\mathscr{B}(\mathcal{Y})$-measurable functions, let $Y_0^\theta\colon \Omega \to \mathcal{Y}$, $\theta \in \Theta$, be i.i.d. $\mathscr{F}/\mathscr{B}(\mathcal{Y})$-measurable functions, assume that $(Y_{-1}^\theta)_{\theta\in\Theta}$, $(Y_0^\theta)_{\theta\in\Theta}$, and $(Z^\theta)_{\theta\in\Theta}$ are independent, and let $Y_n^\theta\colon \Omega \to \mathcal{Y}$, $\theta \in \Theta$, $n \in \mathbb{N}$, satisfy for all $n \in \mathbb{N}$, $\theta \in \Theta$ that*

$$Y_n^\theta = \sum_{l=0}^{n-1} \tfrac{1}{M_{n,l}} \left[\sum_{i=1}^{M_{n,l}} \Phi_l\big(Y_l^{(\theta,l,i)}, Y_{l-1}^{(\theta,-l,i)}, Z^{(\theta,l,i)}\big)\right]. \tag{3.20}$$

Then

(i) it holds for all $n \in (\mathbb{N}_0 \cup \{-1\})$, $\theta \in \Theta$ that $Y_n^\theta\colon \Omega \to \mathcal{Y}$ is an $\mathscr{F}/\mathscr{B}(\mathcal{Y})$-measurable function,

(ii) it holds for all $n \in \mathbb{N}$, $\theta \in \Theta$ that $\sigma_\Omega(Y_n^\theta) \subseteq \sigma_\Omega\big((Y_{-1}^{(\theta,\vartheta)})_{\vartheta\in\Theta}, (Y_0^{(\theta,\vartheta)})_{\vartheta\in\Theta}, (Z^{(\theta,\vartheta)})_{\vartheta\in\Theta}\big)$,

(iii) it holds for every $n, m \in (\mathbb{N}_0 \cup \{-1\})$, $k \in \mathbb{N}$, $\theta_1, \theta_2, \vartheta \in \mathbb{Z}^k$ with $\theta_1 \neq \theta_2$ that $Y_n^{\theta_1}$, $Y_m^{\theta_2}$, and Z^ϑ are independent,

(iv) it holds for every $\theta \in \Theta$ that $\big(Y_l^{(\theta,l,i)}, Y_{l-1}^{(\theta,-l,i)}, Z^{(\theta,l,i)}\big)$, $i \in \mathbb{N}$, $l \in \mathbb{N}_0$, are independent,

(v) it holds for every $n \in (\mathbb{N}_0 \cup \{-1\})$ that Y_n^θ, $\theta \in \Theta$, are identically distributed, and

(vi) it holds for every $\theta \in \Theta$, $l \in \mathbb{N}_0$, $i \in \mathbb{N}$ that $\Omega \ni \omega \mapsto \Phi_l\big(Y_l^{(\theta,l,i)}(\omega), Y_{l-1}^{(\theta,-l,i)}(\omega), Z^{(\theta,l,i)}(\omega)\big) \in \mathcal{Y}$ and $\Omega \ni \omega \mapsto \Phi_l\big(Y_l^0(\omega), Y_{l-1}^1(\omega), Z^0(\omega)\big) \in \mathcal{Y}$ are identically distributed.

Proof of Proposition 3.8. Throughout this proof let $R^{\theta,l,i}\colon \Omega \to \mathcal{Y}\times\mathcal{Y}\times\mathcal{Z}$, $i \in \mathbb{N}$, $l \in \mathbb{N}_0$, $\theta \in \Theta$, satisfy for all $\theta \in \Theta$, $l \in \mathbb{N}_0$, $i \in \mathbb{N}$ that

$$R^{\theta,l,i}(\omega) = \big(Y_l^{(\theta,l,i)}(\omega), Y_{l-1}^{(\theta,-l,i)}(\omega), Z^{(\theta,l,i)}(\omega)\big) \tag{3.21}$$

and let $\Psi_n\colon (\mathcal{Y}\times\mathcal{Y}\times\mathcal{Z})^{\{0,1,\ldots,n-1\}\times\mathbb{N}} \to \mathcal{Y}$, $n \in \mathbb{N}$, satisfy for all $n \in \mathbb{N}$, $r = (r^{l,i})_{(l,i)\in\{0,1,\ldots,n-1\}\times\mathbb{N}} \in (\mathcal{Y}\times\mathcal{Y}\times\mathcal{Z})^{\{0,1,\ldots,n-1\}\times\mathbb{N}}$ that $\Psi_n(r) = \sum_{l=0}^{n-1} \frac{1}{M_{n,l}} \big[\sum_{i=1}^{M_{n,l}} \Phi_l(r^{l,i})\big]$.

First, note that the assumption that it holds for all $\theta \in \Theta$ that $Y^\theta_{-1}\colon \Omega \to \mathcal{Y}$ and $Y^\theta_0\colon \Omega \to \mathcal{Y}$ are $\mathscr{F}/\mathscr{B}(\mathcal{Y})$-measurable functions, the assumption that it holds for all $\theta \in \Theta$ that $Z^\theta\colon \Omega \to \mathcal{Z}$ is an $\mathscr{F}/\mathscr{Z}$-measurable function, the assumption that it holds for all $l \in \mathbb{N}_0$ that $\Phi_l\colon \mathcal{Y} \times \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y}$ is an $(\mathscr{B}(\mathcal{Y}) \otimes \mathscr{B}(\mathcal{Y}) \otimes \mathscr{Z})/\mathscr{B}(\mathcal{Y})$-measurable function, the assumption that $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ is a separable $\mathbb{R}$-Banach space, and induction on $\mathbb{N}_0$ prove (i).

Second, we show (ii) by induction on $n \in \mathbb{N}$. For the base case $n = 1$ observe that it holds for all $\theta \in \Theta$ that

$$Y^\theta_1 = \tfrac{1}{M_{1,0}} \sum_{i=1}^{M_{1,0}} \Phi_0\big(Y^{(\theta,0,i)}_0, Y^{(\theta,0,i)}_{-1}, Z^{(\theta,0,i)}\big). \tag{3.22}$$

This demonstrates for all $\theta \in \Theta$ that

$$\begin{aligned} \sigma_\Omega(Y^\theta_1) &\subseteq \sigma_\Omega\big((Y^{(\theta,0,i)}_{-1})_{i\in\mathbb{N}}, (Y^{(\theta,0,i)}_0)_{i\in\mathbb{N}}, (Z^{(\theta,0,i)})_{i\in\mathbb{N}}\big) \\ &\subseteq \sigma_\Omega\big((Y^{(\theta,\vartheta)}_{-1})_{\vartheta\in\Theta}, (Y^{(\theta,\vartheta)}_0)_{\vartheta\in\Theta}, (Z^{(\theta,\vartheta)})_{\vartheta\in\Theta}\big). \end{aligned} \tag{3.23}$$

This establishes (ii) in the base case $n = 1$. For the induction step $\mathbb{N} \ni n-1 \to n \in \{2, 3, \ldots\}$ let $n \in \{2, 3, \ldots\}$ and assume for all $l \in \{1, \ldots, n-1\}$, $\theta \in \Theta$ that

$$\sigma_\Omega(Y^\theta_l) \subseteq \sigma_\Omega\big((Y^{(\theta,\vartheta)}_{-1})_{\vartheta\in\Theta}, (Y^{(\theta,\vartheta)}_0)_{\vartheta\in\Theta}, (Z^{(\theta,\vartheta)})_{\vartheta\in\Theta}\big). \tag{3.24}$$

This and (3.20) imply for all $\theta \in \Theta$ that

$$\begin{aligned} \sigma_\Omega(Y^\theta_n) &\subseteq \sigma_\Omega\big((Y^{(\theta,-l,i)}_{l-1})_{(l,i)\in\{0,1,\ldots,n-1\}\times\mathbb{N}}, (Y^{(\theta,l,i)}_l)_{(l,i)\in\{0,1,\ldots,n-1\}\times\mathbb{N}}, (Z^{(\theta,l,i)})_{(l,i)\in\mathbb{N}_0\times\mathbb{N}}\big) \\ &\subseteq \sigma_\Omega\big((Y^{(\theta,\mathfrak{z})}_l)_{(l,\mathfrak{z})\in\{1,\ldots,n-1\}\times\mathbb{Z}^2}, (Y^{(\theta,\mathfrak{z})}_{-1})_{\mathfrak{z}\in\mathbb{Z}^2}, (Y^{(\theta,\mathfrak{z})}_0)_{\mathfrak{z}\in\mathbb{Z}^2}, (Z^{(\theta,\mathfrak{z})})_{\mathfrak{z}\in\mathbb{Z}^2}\big) \\ &\subseteq \sigma_\Omega\big((Y^{(\theta,\mathfrak{z},\vartheta)}_{-1})_{(\mathfrak{z},\vartheta)\in\mathbb{Z}^2\times\Theta}, (Y^{(\theta,\mathfrak{z},\vartheta)}_0)_{(\mathfrak{z},\vartheta)\in\mathbb{Z}^2\times\Theta}, (Z^{(\theta,\mathfrak{z},\vartheta)})_{(\mathfrak{z},\vartheta)\in\mathbb{Z}^2\times\Theta}, \\ &\qquad (Y^{(\theta,\vartheta)}_{-1})_{\vartheta\in\Theta}, (Y^{(\theta,\vartheta)}_0)_{\vartheta\in\Theta}, (Z^{(\theta,\vartheta)})_{\vartheta\in\Theta}\big) \\ &= \sigma_\Omega\big((Y^{(\theta,\vartheta)}_{-1})_{\vartheta\in\Theta}, (Y^{(\theta,\vartheta)}_0)_{\vartheta\in\Theta}, (Z^{(\theta,\vartheta)})_{\vartheta\in\Theta}\big). \end{aligned} \tag{3.25}$$

Induction hence establishes (ii).

Third, observe that the assumption that $(Y^\theta_{-1})_{\theta\in\Theta}$, $(Y^\theta_0)_{\theta\in\Theta}$, and $(Z^\theta)_{\theta\in\Theta}$ are independent ensures that it holds for every $k \in \mathbb{N}$, $\theta_1, \theta_2, \vartheta \in \mathbb{Z}^k$ with $\theta_1 \neq \theta_2$ that $\sigma_\Omega\big((Y^{(\theta_1,\mathfrak{z})}_{-1})_{\mathfrak{z}\in\Theta}, (Y^{(\theta_1,\mathfrak{z})}_0)_{\mathfrak{z}\in\Theta}, (Z^{(\theta_1,\mathfrak{z})})_{\mathfrak{z}\in\Theta}\big)$, $\sigma_\Omega\big((Y^{(\theta_2,\mathfrak{z})}_{-1})_{\mathfrak{z}\in\Theta}, (Y^{(\theta_2,\mathfrak{z})}_0)_{\mathfrak{z}\in\Theta}, (Z^{(\theta_2,\mathfrak{z})})_{\mathfrak{z}\in\Theta}\big)$, and $\sigma_\Omega(Z^\vartheta)$ are independent. Combining this with (ii) proves (iii).

Fourth, note that the assumption that the family $(Y^\theta_{-1})_{\theta\in\Theta}$ is independent, the assumption that the family $(Y^\theta_0)_{\theta\in\Theta}$ is independent, the assumption that the family $(Z^\theta)_{\theta\in\Theta}$ is independent, and the assumption that $(Y^\theta_{-1})_{\theta\in\Theta}$, $(Y^\theta_0)_{\theta\in\Theta}$, and $(Z^\theta)_{\theta\in\Theta}$ are independent imply for every $\theta \in \Theta$ that the family

$$\mathbb{N}_0\times\mathbb{N} \ni (l,i) \mapsto \begin{cases} \sigma_\Omega\big(Y^{(\theta,0,i)}_{-1}, Y^{(\theta,0,i)}_0, Z^{(\theta,0,i)}\big) & : l = 0 \\ \sigma_\Omega\big((Y^{(\theta,l,i,\vartheta)}_{-1})_{\vartheta\in\Theta}, (Y^{(\theta,l,i,\vartheta)}_0)_{\vartheta\in\Theta}, (Z^{(\theta,l,i,\vartheta)})_{\vartheta\in\Theta}, (Y^{(\theta,-l,i,\vartheta)}_{-1})_{\vartheta\in\Theta}, \\ \quad (Y^{(\theta,-l,i,\vartheta)}_0)_{\vartheta\in\Theta}, (Z^{(\theta,-l,i,\vartheta)})_{\vartheta\in\Theta}, Z^{(\theta,l,i)}\big) & : l \neq 0 \end{cases}$$

is independent. This, (3.21), and (ii) ensure for every $\theta \in \Theta$ that the family

$$\big[\mathbb{N}_0 \times \mathbb{N} \ni (l,i) \mapsto \big(Y^{(\theta,l,i)}_l, Y^{(\theta,-l,i)}_{l-1}, Z^{(\theta,l,i)}\big)\big] = (R^{\theta,l,i})_{(l,i)\in\mathbb{N}_0\times\mathbb{N}} \tag{3.26}$$

is independent. This finishes the proof of (iv).

Fifth, we establish (v) by induction on $n \in \mathbb{N}$. For the base case $n = 1$ note that the assumption that Y_{-1}^θ, $\theta \in \Theta$, are identically distributed, the assumption that Y_0^θ, $\theta \in \Theta$, are identically distributed, the assumption that Z^θ, $\theta \in \Theta$, are identically distributed, the assumption that $(Y_{-1}^\theta)_{\theta\in\Theta}$, $(Y_0^\theta)_{\theta\in\Theta}$, and $(Z^\theta)_{\theta\in\Theta}$ are independent, and (3.21) establish for every $\theta \in \Theta$, $i \in \mathbb{N}$ that

$$R^{\theta,0,i} = \big(Y_0^{(\theta,0,i)}, Y_{-1}^{(\theta,0,i)}, Z^{(\theta,0,i)}\big) \quad \text{and} \quad (Y_0^0, Y_{-1}^1, Z^0) \tag{3.27}$$

are identically distributed. In particular, this shows for every $i \in \mathbb{N}$ that $R^{\theta,0,i}$, $\theta \in \Theta$, are identically distributed. Combining this with (3.26) proves that $\big(\Omega \ni \omega \mapsto (R^{\theta,0,i}(\omega))_{i\in\mathbb{N}} \in (\mathcal{Y}\times\mathcal{Y}\times\mathcal{Z})^{\mathbb{N}}\big)$, $\theta \in \Theta$, are identically distributed. The fact that $\forall\, \theta \in \Theta, \omega \in \Omega\colon Y_1^\theta(\omega) = \Psi_1\big((R^{\theta,0,i}(\omega))_{i\in\mathbb{N}}\big)$ hence implies that Y_1^θ, $\theta \in \Theta$, are identically distributed. This, the assumption that Y_{-1}^θ, $\theta \in \Theta$, are identically distributed, and the assumption that Y_0^θ, $\theta \in \Theta$, are identically distributed show (v) in the base case $n = 1$. For the induction step $\mathbb{N} \ni n-1 \to n \in \{2, 3, \ldots\}$ let $n \in \{2, 3, \ldots\}$ and assume for every $l \in \{-1, 0, 1, \ldots, n-1\}$ that Y_l^θ, $\theta \in \Theta$, are identically distributed. This, the assumption that Z^θ, $\theta \in \Theta$, are identically distributed, (iii), and (3.21) ensure for every $\theta \in \Theta$, $l \in \{1, \ldots, n-1\}$, $i \in \mathbb{N}$ that

$$R^{\theta,l,i} = \big(Y_l^{(\theta,l,i)}, Y_{l-1}^{(\theta,-l,i)}, Z^{(\theta,l,i)}\big) \quad \text{and} \quad (Y_l^0, Y_{l-1}^1, Z^0) \tag{3.28}$$

are identically distributed. Combining this and (3.27) establishes for every $l \in \{0, 1, \ldots, n-1\}$, $i \in \mathbb{N}$ that $R^{\theta,l,i}$, $\theta \in \Theta$, are identically distributed. This and (3.26) demonstrate that $\big(\Omega \ni \omega \mapsto (R^{\theta,l,i}(\omega))_{(l,i)\in\{0,1,\ldots,n-1\}\times\mathbb{N}} \in (\mathcal{Y}\times\mathcal{Y}\times\mathcal{Z})^{\{0,1,\ldots,n-1\}\times\mathbb{N}}\big)$, $\theta \in \Theta$, are identically distributed. Therefore, the fact that $\forall\, \theta \in \Theta, \omega \in \Omega\colon Y_n^\theta(\omega) = \Psi_n\big((R^{\theta,l,i}(\omega))_{(l,i)\in\{0,1,\ldots,n-1\}\times\mathbb{N}}\big)$ shows that Y_n^θ, $\theta \in \Theta$, are identically distributed. Induction hence proves (v).

Sixth, observe that (3.27) and (3.28) establish (vi). The proof of Proposition 3.8 is thus complete. □

3.1.4 Error analysis

Proposition 3.9. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be a separable $\mathbb{R}$-Banach space, let $C, c \in (0,\infty)$, $(\mathfrak{c}_k)_{k\in\mathbb{N}_0} \subseteq (0,\infty)$, $\Theta = \bigcup_{n=1}^\infty \mathbb{Z}^n$, $y \in \mathcal{Y}$, for every $n \in \mathbb{N}$ let $(M_{n,l})_{l\in\{0,1,\ldots,n\}} \subseteq \mathbb{N}$ satisfy $M_{n,1} \ge M_{n,2} \ge \ldots \ge M_{n,n}$, let $(\mathcal{Z}, \mathscr{Z})$ be a measurable space, let $Z^\theta\colon \Omega \to \mathcal{Z}$, $\theta \in \Theta$, be i.i.d. $\mathscr{F}/\mathscr{Z}$-measurable functions, let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}}, \|\cdot\|_{\mathcal{H}})$ be a separable $\mathbb{R}$-Hilbert space, let $\mathscr{S} = \sigma_{L(\mathcal{Y},\mathcal{H})}\big(\big\{\{\varphi \in L(\mathcal{Y},\mathcal{H})\colon \varphi(x) \in \mathcal{B}\} \subseteq L(\mathcal{Y},\mathcal{H})\colon x \in \mathcal{Y}, \mathcal{B} \in \mathscr{B}(\mathcal{H})\big\}\big)$, let $\psi_k\colon \Omega \to L(\mathcal{Y},\mathcal{H})$, $k \in \mathbb{N}_0$, be $\mathscr{F}/\mathscr{S}$-measurable functions, let $\Phi_l\colon \mathcal{Y}\times\mathcal{Y}\times\mathcal{Z} \to \mathcal{Y}$, $l \in \mathbb{N}_0$, be $(\mathscr{B}(\mathcal{Y})\otimes\mathscr{B}(\mathcal{Y})\otimes\mathscr{Z})/\mathscr{B}(\mathcal{Y})$-measurable functions, let $Y_{-1}^\theta\colon \Omega \to \mathcal{Y}$, $\theta \in \Theta$, be i.i.d. $\mathscr{F}/\mathscr{B}(\mathcal{Y})$-measurable functions, let $Y_0^\theta\colon \Omega \to \mathcal{Y}$, $\theta \in \Theta$, be i.i.d. $\mathscr{F}/\mathscr{B}(\mathcal{Y})$-measurable functions, assume that $(Y_{-1}^\theta)_{\theta\in\Theta}$, $(Y_0^\theta)_{\theta\in\Theta}$, $(Z^\theta)_{\theta\in\Theta}$, and $(\psi_k)_{k\in\mathbb{N}_0}$ are independent, let $Y_n^\theta\colon \Omega \to \mathcal{Y}$, $\theta \in \Theta$, $n \in \mathbb{N}$, satisfy for all $n \in \mathbb{N}$, $\theta \in \Theta$ that*

$$Y_n^\theta = \sum_{l=0}^{n-1} \tfrac{1}{M_{n,l}} \Bigg[\sum_{i=1}^{M_{n,l}} \Phi_l\big(Y_l^{(\theta,l,i)}, Y_{l-1}^{(\theta,-l,i)}, Z^{(\theta,l,i)}\big)\Bigg], \tag{3.29}$$

and assume for all $k \in \mathbb{N}_0$, $n \in \mathbb{N}$ that $\mathbb{E}\big[\|\Phi_k(Y_k^0, Y_{k-1}^1, Z^0)\|_{\mathcal{Y}}\big] < \infty$ *and*

$$\max\big\{\mathbb{E}\big[\|\psi_k(\Phi_0(Y_0^0, Y_{-1}^1, Z^0))\|_{\mathcal{H}}^2\big], \mathbb{1}_{\mathbb{N}}(k)\,\mathbb{E}\big[\|\psi_k(Y_0^0 - y)\|_{\mathcal{H}}^2\big]\big\} \leq \tfrac{C^2}{\mathfrak{c}_k}, \tag{3.30}$$

$$\mathbb{E}\big[\|\psi_k(\Phi_n(Y_n^0, Y_{n-1}^1, Z^0))\|_{\mathcal{H}}^2\big] \leq c\,\mathbb{E}\big[\|\psi_{k+1}(Y_n^0 - Y_{n-1}^1)\|_{\mathcal{H}}^2\big], \tag{3.31}$$

$$\mathbb{E}\left[\left\|\psi_k\left(y - \sum_{l=0}^{n-1}\mathbb{E}\big[\Phi_l(Y_l^0, Y_{l-1}^1, Z^0)\big]\right)\right\|_{\mathcal{H}}^2\right] \leq \tfrac{2c}{M_{n,n}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_{n-1}^0 - y)\|_{\mathcal{H}}^2\big]. \tag{3.32}$$

Then it holds for all $N \in \mathbb{N}$ that

$$\begin{aligned}
&\big(\mathbb{E}\big[\|\psi_0(Y_N^0 - y)\|_{\mathcal{H}}^2\big]\big)^{1/2} \\
&\leq C(1+4c)^{N/2}\Bigg[\min\Bigg(\Bigg\{\min\{M_{l_k,0}\mathfrak{c}_k, M_{l_k,1}\mathfrak{c}_{k+1}\}\prod_{j=1}^{k} M_{l_{j-1},l_j+1}\colon \\
&\qquad \begin{array}{l} k \in \mathbb{N}\cap[0,N-1],\ (l_i)_{i\in\{0,1,\ldots,k\}} \subseteq \\ \{1,2,\ldots,N\},\ N = l_0 > l_1 > \ldots > l_k \end{array}\Bigg\} \cup \big\{\min\{M_{N,0}\mathfrak{c}_0, M_{N,1}\mathfrak{c}_1\}\big\}\Bigg)\Bigg]^{-1/2} < \infty.
\end{aligned} \tag{3.33}$$

Proof of Proposition 3.9. First of all, note that the assumption that $\forall\, l \in \mathbb{N}_0\colon \mathbb{E}\big[\|\Phi_l(Y_l^0, Y_{l-1}^1, Z^0)\|_{\mathcal{Y}}\big] < \infty$ and (vi) in Proposition 3.8 establish for all $l \in \mathbb{N}_0$, $i \in \mathbb{N}$ that

$$\mathbb{E}\Big[\big\|\Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big)\big\|_{\mathcal{Y}}\Big] = \mathbb{E}\big[\|\Phi_l(Y_l^0, Y_{l-1}^1, Z^0)\|_{\mathcal{Y}}\big] < \infty. \tag{3.34}$$

This, (i) in Proposition 3.8, and (3.29) ensure for all $n \in \mathbb{N}$ that $Y_n^0\colon \Omega \to \mathcal{Y}$ is an $\mathscr{F}/\mathscr{B}(\mathcal{Y})$-measurable function and

$$\begin{aligned}
\mathbb{E}\big[\|Y_n^0\|_{\mathcal{Y}}\big] &= \mathbb{E}\left[\left\|\sum_{l=0}^{n-1}\tfrac{1}{M_{n,l}}\left[\sum_{i=1}^{M_{n,l}}\Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big)\right]\right\|_{\mathcal{Y}}\right] \\
&\leq \sum_{l=0}^{n-1}\tfrac{1}{M_{n,l}}\left[\sum_{i=1}^{M_{n,l}}\mathbb{E}\Big[\big\|\Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big)\big\|_{\mathcal{Y}}\Big]\right] \\
&= \sum_{l=0}^{n-1}\tfrac{1}{M_{n,l}}\left[\sum_{i=1}^{M_{n,l}}\mathbb{E}\big[\|\Phi_l(Y_l^0, Y_{l-1}^1, Z^0)\|_{\mathcal{Y}}\big]\right] \\
&= \sum_{l=0}^{n-1}\mathbb{E}\big[\|\Phi_l(Y_l^0, Y_{l-1}^1, Z^0)\|_{\mathcal{Y}}\big] < \infty.
\end{aligned} \tag{3.35}$$

In addition, (ii) in Proposition 3.8 yields for all $n \in \mathbb{N}$, $\theta \in \Theta$ that

$$\begin{aligned}
\sigma_\Omega(Y_n^\theta) &\subseteq \sigma_\Omega\big((Y_{-1}^{(\theta,\vartheta)})_{\vartheta\in\Theta}, (Y_0^{(\theta,\vartheta)})_{\vartheta\in\Theta}, (Z^{(\theta,\vartheta)})_{\vartheta\in\Theta}\big) \\
&\subseteq \sigma_\Omega\big((Y_{-1}^{\vartheta})_{\vartheta\in\Theta}, (Y_0^{\vartheta})_{\vartheta\in\Theta}, (Z^{\vartheta})_{\vartheta\in\Theta}\big).
\end{aligned} \tag{3.36}$$

Note that this implies that $\sigma_\Omega\big((Y_n^\theta)_{(n,\theta)\in(\mathbb{N}_0\cup\{-1\})\times\Theta}, (Z^\theta)_{\theta\in\Theta}\big) \subseteq \sigma_\Omega\big((Y_{-1}^\theta)_{\theta\in\Theta}, (Y_0^\theta)_{\theta\in\Theta}, (Z^\theta)_{\theta\in\Theta}\big)$. This and the assumption that $(Y_{-1}^\theta)_{\theta\in\Theta}$, $(Y_0^\theta)_{\theta\in\Theta}$, $(Z^\theta)_{\theta\in\Theta}$, and $(\psi_k)_{k\in\mathbb{N}_0}$ are independent demonstrate for every $k \in \mathbb{N}_0$ that

$$\sigma_\Omega\big((Y_n^\theta)_{(n,\theta)\in(\mathbb{N}_0\cup\{-1\})\times\Theta}, (Z^\theta)_{\theta\in\Theta}\big) \quad \text{and} \quad \psi_k \tag{3.37}$$

are independent. Corollary 3.5 and (3.35) hence show for all $k \in \mathbb{N}_0$, $n \in \mathbb{N}$ that

$$\mathbb{E}\big[\|\psi_k(Y_n^0 - y)\|_{\mathcal{H}}^2\big] = \mathbb{E}\big[\|\psi_k(Y_n^0 - \mathbb{E}[Y_n^0])\|_{\mathcal{H}}^2\big] + \mathbb{E}\big[\|\psi_k(\mathbb{E}[Y_n^0] - y)\|_{\mathcal{H}}^2\big]. \tag{3.38}$$

Next observe that (3.29), (3.34), (iv) in Proposition 3.8, (3.37), and Corollary 3.7 (with $n \leftarrow \sum_{l=0}^{\mathfrak{n}-1} M_{\mathfrak{n},l}$, $\psi \leftarrow \psi_k$ for $\mathfrak{n} \in \mathbb{N}$, $k \in \mathbb{N}_0$ in the notation of Corollary 3.7) prove for all $k \in \mathbb{N}_0$, $n \in \mathbb{N}$ that

$$\begin{aligned}
&\mathbb{E}\big[\|\psi_k(Y_n^0 - \mathbb{E}[Y_n^0])\|_{\mathcal{H}}^2\big] \\
&= \mathbb{E}\Bigg[\bigg\|\psi_k\bigg(\sum_{l=0}^{n-1} \tfrac{1}{M_{n,l}} \sum_{i=1}^{M_{n,l}} \big(\Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big) - \mathbb{E}\big[\Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big)\big]\big)\bigg)\bigg\|_{\mathcal{H}}^2\Bigg] \\
&= \mathbb{E}\Bigg[\bigg\|\sum_{l=0}^{n-1} \sum_{i=1}^{M_{n,l}} \psi_k\Big(\tfrac{1}{M_{n,l}}\big(\Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big) - \mathbb{E}\big[\Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big)\big]\big)\Big)\bigg\|_{\mathcal{H}}^2\Bigg] \\
&= \sum_{l=0}^{n-1} \sum_{i=1}^{M_{n,l}} \mathbb{E}\Big[\big\|\tfrac{1}{M_{n,l}}\psi_k\big(\Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big) - \mathbb{E}\big[\Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big)\big]\big)\big\|_{\mathcal{H}}^2\Big] \\
&= \sum_{l=0}^{n-1} \tfrac{1}{(M_{n,l})^2} \Bigg[\sum_{i=1}^{M_{n,l}} \mathbb{E}\Big[\big\|\psi_k\big(\Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big) \\
&\qquad\qquad - \mathbb{E}\big[\Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big)\big]\big)\big\|_{\mathcal{H}}^2\Big]\Bigg].
\end{aligned} \tag{3.39}$$

Moreover, the fact that it holds for all $k \in \mathbb{N}_0$, $x \in \mathcal{Y}$ that $\Omega \ni \omega \mapsto [\psi_k(\omega)](x) \in \mathcal{H}$ is an $\mathscr{F}/\mathscr{B}(\mathcal{H})$-measurable function (cf. Lemma 3.1) and the fact that it holds for all $k \in \mathbb{N}_0$, $\omega \in \Omega$ that $\mathcal{Y} \ni x \mapsto [\psi_k(\omega)](x) \in \mathcal{H}$ is a continuous function demonstrate that

$$\mathcal{Y} \times \Omega \ni (x, \omega) \mapsto [\psi_k(\omega)](x) \in \mathcal{H} \tag{3.40}$$

is a continuous random field. This, (3.37), (vi) in Proposition 3.8, and Hutzenthaler, Jentzen, & von Wurstemberger [182, Lemma 3.5] (with $S \leftarrow \mathcal{Y}$, $E \leftarrow \mathcal{H}$, $U = V \leftarrow (\mathcal{Y} \times \Omega \ni (x, \omega) \mapsto [\psi_k(\omega)](x) \in \mathcal{H})$, $X \leftarrow \Phi_l(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)})$, $Y \leftarrow \Phi_l(Y_l^0, Y_{l-1}^1, Z^0)$ for $i \in \mathbb{N}$, $l, k \in \mathbb{N}_0$ in the notation of [182, Lemma 3.5]) ensure for all $k, l \in \mathbb{N}_0$, $i \in \mathbb{N}$ that

$$\begin{aligned}
&\Omega \ni \omega \mapsto [\psi_k(\omega)]\big(\Phi_l\big(Y_l^{(0,l,i)}(\omega), Y_{l-1}^{(0,-l,i)}(\omega), Z^{(0,l,i)}(\omega)\big)\big) \in \mathcal{H} \quad \text{and} \\
&\Omega \ni \omega \mapsto [\psi_k(\omega)]\big(\Phi_l(Y_l^0(\omega), Y_{l-1}^1(\omega), Z^0(\omega))\big) \in \mathcal{H}
\end{aligned} \tag{3.41}$$

are identically distributed. This, (3.39), (3.34), (3.37), and Corollary 3.5 (with $y \leftarrow 0$, $Y \leftarrow \Phi_l(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)})$, $\psi \leftarrow \psi_k$ for $i \in \{1, 2, \dots, M_{n,l}\}$, $l \in \{0, 1, \dots, n-1\}$, $n \in \mathbb{N}$, $k \in \mathbb{N}_0$ in the notation of Corollary 3.5) imply for all $k \in \mathbb{N}_0$, $n \in \mathbb{N}$ that

$$\begin{aligned}
&\mathbb{E}\big[\|\psi_k(Y_n^0 - \mathbb{E}[Y_n^0])\|_{\mathcal{H}}^2\big] \le \sum_{l=0}^{n-1} \tfrac{1}{(M_{n,l})^2} \Bigg[\sum_{i=1}^{M_{n,l}} \mathbb{E}\Big[\big\|\psi_k\big(\Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big)\big)\big\|_{\mathcal{H}}^2\Big]\Bigg] \\
&= \sum_{l=0}^{n-1} \tfrac{1}{(M_{n,l})^2} \Bigg[\sum_{i=1}^{M_{n,l}} \mathbb{E}\big[\|\psi_k(\Phi_l(Y_l^0, Y_{l-1}^1, Z^0))\|_{\mathcal{H}}^2\big]\Bigg] \\
&= \sum_{l=0}^{n-1} \Big(\tfrac{1}{M_{n,l}}\, \mathbb{E}\big[\|\psi_k(\Phi_l(Y_l^0, Y_{l-1}^1, Z^0))\|_{\mathcal{H}}^2\big]\Big) \\
&= \tfrac{1}{M_{n,0}}\, \mathbb{E}\big[\|\psi_k(\Phi_0(Y_0^0, Y_{-1}^1, Z^0))\|_{\mathcal{H}}^2\big] + \sum_{l=1}^{n-1} \Big(\tfrac{1}{M_{n,l}}\, \mathbb{E}\big[\|\psi_k(\Phi_l(Y_l^0, Y_{l-1}^1, Z^0))\|_{\mathcal{H}}^2\big]\Big).
\end{aligned} \tag{3.42}$$

Assumptions (3.30)–(3.31), the fact that $\forall\, a,b \in \mathbb{R}\colon (a+b)^2 \le 2(a^2+b^2)$, (3.40), (3.37), (v) in Proposition 3.8, [182, Lemma 3.5] (with $S \leftarrow \mathcal{Y}$, $E \leftarrow \mathcal{H}$, $U = V \leftarrow (\mathcal{Y}\times\Omega \ni (x,\omega) \mapsto [\psi_k(\omega)](x) \in \mathcal{H})$, $X \leftarrow Y_l^1 - y$, $Y \leftarrow Y_l^0 - y$ for $l,k \in \mathbb{N}_0$ in the notation of [182, Lemma 3.5]), and the assumption that $\forall\, n \in \mathbb{N}\colon M_{n,1} \ge M_{n,2} \ge \ldots \ge M_{n,n}$ hence prove for all $k \in \mathbb{N}_0$, $n \in \mathbb{N}$ that

$$
\begin{aligned}
&\mathbb{E}\big[\|\psi_k(Y_n^0 - \mathbb{E}[Y_n^0])\|_{\mathcal{H}}^2\big] \le \tfrac{C^2}{M_{n,0}\mathfrak{c}_k} + \left[\sum_{l=1}^{n-1} \tfrac{c}{M_{n,l}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_l^0 - Y_{l-1}^1)\|_{\mathcal{H}}^2\big]\right]\\
&= \tfrac{C^2}{M_{n,0}\mathfrak{c}_k} + \left[\sum_{l=1}^{n-1} \tfrac{c}{M_{n,l}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_l^0 - y) + \psi_{k+1}(y - Y_{l-1}^1)\|_{\mathcal{H}}^2\big]\right]\\
&\le \tfrac{C^2}{M_{n,0}\mathfrak{c}_k} + \left[\sum_{l=1}^{n-1} \tfrac{c}{M_{n,l}}\,\mathbb{E}\big[(\|\psi_{k+1}(Y_l^0 - y)\|_{\mathcal{H}} + \|\psi_{k+1}(Y_{l-1}^1 - y)\|_{\mathcal{H}})^2\big]\right]\\
&\le \tfrac{C^2}{M_{n,0}\mathfrak{c}_k} + \left[\sum_{l=1}^{n-1} \tfrac{2c}{M_{n,l}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_l^0 - y)\|_{\mathcal{H}}^2\big]\right] + \left[\sum_{l=1}^{n-1} \tfrac{2c}{M_{n,l}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_{l-1}^1 - y)\|_{\mathcal{H}}^2\big]\right]\\
&= \tfrac{C^2}{M_{n,0}\mathfrak{c}_k} + \left[\sum_{l=1}^{n-1} \tfrac{2c}{M_{n,l}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_l^0 - y)\|_{\mathcal{H}}^2\big]\right] + \left[\sum_{l=0}^{n-2} \tfrac{2c}{M_{n,l+1}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_l^0 - y)\|_{\mathcal{H}}^2\big]\right]\\
&\le \tfrac{C^2}{M_{n,0}\mathfrak{c}_k} + \left[\sum_{l=1}^{n-1} \tfrac{2c}{M_{n,l+1}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_l^0 - y)\|_{\mathcal{H}}^2\big]\right] + \left[\sum_{l=0}^{n-2} \tfrac{2c}{M_{n,l+1}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_l^0 - y)\|_{\mathcal{H}}^2\big]\right]\\
&= \tfrac{C^2}{M_{n,0}\mathfrak{c}_k} + \left[\sum_{l=0}^{n-1} \tfrac{2(2-\mathbb{1}_{\{0\}}(l)-\mathbb{1}_{\{n-1\}}(l))c}{M_{n,l+1}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_l^0 - y)\|_{\mathcal{H}}^2\big]\right].
\end{aligned}
\tag{3.43}
$$

Furthermore, note that (3.35), (3.29), (3.34), and (vi) in Proposition 3.8 show for all $n \in \mathbb{N}$ that

$$
\begin{aligned}
\mathbb{E}[Y_n^0] &= \mathbb{E}\left[\sum_{l=0}^{n-1} \tfrac{1}{M_{n,l}}\left[\sum_{i=1}^{M_{n,l}} \Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big)\right]\right]\\
&= \sum_{l=0}^{n-1} \tfrac{1}{M_{n,l}}\left[\sum_{i=1}^{M_{n,l}} \mathbb{E}\big[\Phi_l\big(Y_l^{(0,l,i)}, Y_{l-1}^{(0,-l,i)}, Z^{(0,l,i)}\big)\big]\right]\\
&= \sum_{l=0}^{n-1} \tfrac{1}{M_{n,l}}\left[\sum_{i=1}^{M_{n,l}} \mathbb{E}\big[\Phi_l(Y_l^0, Y_{l-1}^1, Z^0)\big]\right] = \sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l(Y_l^0, Y_{l-1}^1, Z^0)\big].
\end{aligned}
\tag{3.44}
$$

This and assumption (3.32) establish for all $k \in \mathbb{N}_0$, $n \in \mathbb{N}$ that

$$
\begin{aligned}
\mathbb{E}\big[\|\psi_k(\mathbb{E}[Y_n^0] - y)\|_{\mathcal{H}}^2\big] &= \mathbb{E}\left[\left\|\psi_k\left(\left[\sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l(Y_l^0, Y_{l-1}^1, Z^0)\big]\right] - y\right)\right\|_{\mathcal{H}}^2\right]\\
&\le \tfrac{2c}{M_{n,n}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_{n-1}^0 - y)\|_{\mathcal{H}}^2\big].
\end{aligned}
\tag{3.45}
$$

Combining (3.38) with (3.43) and assumption (3.30) hence proves for all $k \in \mathbb{N}_0$, $n \in \mathbb{N}$ that

$$
\begin{aligned}
&\mathbb{E}\big[\|\psi_k(Y_n^0 - y)\|_{\mathcal{H}}^2\big] \le \tfrac{C^2}{M_{n,0}\mathfrak{c}_k} + \left[\sum_{l=0}^{n-1} \tfrac{2(2-\mathbb{1}_{\{0\}}(l))c}{M_{n,l+1}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_l^0 - y)\|_{\mathcal{H}}^2\big]\right]\\
&= \tfrac{C^2}{M_{n,0}\mathfrak{c}_k} + \tfrac{2c}{M_{n,1}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_0^0 - y)\|_{\mathcal{H}}^2\big] + \left[\sum_{l=1}^{n-1} \tfrac{4c}{M_{n,l+1}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_l^0 - y)\|_{\mathcal{H}}^2\big]\right]\\
&\le \tfrac{C^2}{M_{n,0}\mathfrak{c}_k} + \tfrac{2C^2c}{M_{n,1}\mathfrak{c}_{k+1}} + \left[\sum_{l=1}^{n-1} \tfrac{4c}{M_{n,l+1}}\,\mathbb{E}\big[\|\psi_{k+1}(Y_l^0 - y)\|_{\mathcal{H}}^2\big]\right].
\end{aligned}
\tag{3.46}
$$

Next we introduce some additional notation. For the remainder of this proof let $N \in \mathbb{N}$, let $\varepsilon_n \in [0,\infty]$, $n \in \{1,2,\dots,N\}$, satisfy for all $n \in \{1,2,\dots,N\}$ that

$$\varepsilon_n = \max\Bigg(\Bigg\{\bigg(M_{l_{k-1},n+1}\prod_{j=1}^{k-1} M_{l_{j-1},l_j+1}\bigg)^{-1}\mathbb{E}\big[\|\psi_k(Y_n^0-y)\|_{\mathcal{H}}^2\big]\colon \begin{array}{l} k\in\mathbb{N}\cap[0,N-n],\ (l_i)_{i\in\{0,1,\dots,k-1\}}\subseteq\{n+1,\\ n+2,\dots,N\},\ N=l_0>l_1>\ldots>l_{k-1}\end{array}\Bigg\}\cup\big\{\mathbb{1}_{\{N\}}(n)\,\mathbb{E}\big[\|\psi_0(Y_N^0-y)\|_{\mathcal{H}}^2\big]\big\}\Bigg), \tag{3.47}$$

and let $a \in [0,\infty)$ be given by

$$a = C^2(1+2c)\Bigg[\min\Bigg(\Bigg\{\min\{M_{l_k,0}\mathfrak{c}_k, M_{l_k,1}\mathfrak{c}_{k+1}\}\prod_{j=1}^{k} M_{l_{j-1},l_j+1}\colon \begin{array}{l} k\in\mathbb{N}\cap[0,N-1],\ (l_i)_{i\in\{0,1,\dots,k\}}\subseteq\\ \{1,2,\dots,N\},\ N=l_0>l_1>\ldots>l_k\end{array}\Bigg\}\cup\big\{\min\{M_{N,0}\mathfrak{c}_0, M_{N,1}\mathfrak{c}_1\}\big\}\Bigg)\Bigg]^{-1}. \tag{3.48}$$

Observe that (3.46)–(3.48) establish for all $n \in \mathbb{N}\cap[0,N-1]$, $k \in \{1,2,\dots,N-n\}$, $(l_i)_{i\in\{0,1,\dots,k-1\}} \subseteq \{n+1,n+2,\dots,N\}$ with $l_0 = N$ and $\forall\, i \in \mathbb{N}\cap[0,k-1]\colon l_{i-1} > l_i$ that

$$\begin{aligned}
&\bigg(M_{l_{k-1},n+1}\prod_{j=1}^{k-1} M_{l_{j-1},l_j+1}\bigg)^{-1}\mathbb{E}\big[\|\psi_k(Y_n^0-y)\|_{\mathcal{H}}^2\big]\\
&\le \Big(\tfrac{C^2}{M_{n,0}\mathfrak{c}_k}+\tfrac{2C^2c}{M_{n,1}\mathfrak{c}_{k+1}}\Big)\bigg(M_{l_{k-1},n+1}\prod_{j=1}^{k-1} M_{l_{j-1},l_j+1}\bigg)^{-1}\\
&\quad + 4c\sum_{\ell=1}^{n-1}\Bigg[\bigg(M_{n,\ell+1}M_{l_{k-1},n+1}\prod_{j=1}^{k-1} M_{l_{j-1},l_j+1}\bigg)^{-1}\mathbb{E}\big[\|\psi_{k+1}(Y_\ell^0-y)\|_{\mathcal{H}}^2\big]\Bigg]\\
&\le C^2(1+2c)\Bigg[\max_{l_k\in\{1,2,\dots,l_{k-1}-1\}}\bigg(\min\{M_{l_k,0}\mathfrak{c}_k, M_{l_k,1}\mathfrak{c}_{k+1}\}M_{l_{k-1},l_k+1}\prod_{j=1}^{k-1} M_{l_{j-1},l_j+1}\bigg)^{-1}\Bigg]\\
&\quad + 4c\sum_{\ell=1}^{n-1}\max_{l_k\in\{\ell+1,\ell+2,\dots,l_{k-1}-1\}}\Bigg[\bigg(M_{l_k,\ell+1}M_{l_{k-1},l_k+1}\prod_{j=1}^{k-1} M_{l_{j-1},l_j+1}\bigg)^{-1}\mathbb{E}\big[\|\psi_{k+1}(Y_\ell^0-y)\|_{\mathcal{H}}^2\big]\Bigg]\\
&= C^2(1+2c)\Bigg[\min_{l_k\in\{1,2,\dots,l_{k-1}-1\}}\bigg(\min\{M_{l_k,0}\mathfrak{c}_k, M_{l_k,1}\mathfrak{c}_{k+1}\}\prod_{j=1}^{k} M_{l_{j-1},l_j+1}\bigg)\Bigg]^{-1}\\
&\quad + 4c\sum_{\ell=1}^{n-1}\max_{l_k\in\{\ell+1,\ell+2,\dots,l_{k-1}-1\}}\Bigg[\bigg(M_{l_k,\ell+1}\prod_{j=1}^{k} M_{l_{j-1},l_j+1}\bigg)^{-1}\mathbb{E}\big[\|\psi_{k+1}(Y_\ell^0-y)\|_{\mathcal{H}}^2\big]\Bigg]\\
&\le a + 4c\sum_{\ell=1}^{n-1}\varepsilon_\ell.
\end{aligned}\tag{3.49}$$

In addition, (3.46)–(3.48) ensure that

$$\begin{aligned}
\varepsilon_N &= \mathbb{E}\big[\|\psi_0(Y_N^0-y)\|_{\mathcal{H}}^2\big] \le \tfrac{C^2}{M_{N,0}\mathfrak{c}_0}+\tfrac{2C^2c}{M_{N,1}\mathfrak{c}_1}+\Bigg[\sum_{\ell=1}^{N-1}\tfrac{4c}{M_{N,\ell+1}}\,\mathbb{E}\big[\|\psi_1(Y_\ell^0-y)\|_{\mathcal{H}}^2\big]\Bigg]\\
&\le C^2(1+2c)(\min\{M_{N,0}\mathfrak{c}_0, M_{N,1}\mathfrak{c}_1\})^{-1}+4c\Bigg[\sum_{\ell=1}^{N-1}(M_{N,\ell+1})^{-1}\,\mathbb{E}\big[\|\psi_1(Y_\ell^0-y)\|_{\mathcal{H}}^2\big]\Bigg]\\
&\le a + 4c\sum_{\ell=1}^{N-1}\varepsilon_\ell.
\end{aligned}\tag{3.50}$$

This, (3.47), and (3.49) show for all $n \in \{1, 2, \ldots, N\}$ that

$$\varepsilon_n \leq a + 4c \sum_{\ell=1}^{n-1} \varepsilon_\ell. \tag{3.51}$$

The fact that $a + c < \infty$ and the discrete Gronwall-type inequality in Agarwal [2, Corollary 4.1.2] hence establish for all $n \in \{1, 2, \ldots, N\}$ that

$$\varepsilon_n \leq a(1 + 4c)^{n-1} < \infty. \tag{3.52}$$

This and (3.47)–(3.48) imply that

$$\begin{aligned} &\mathbb{E}\big[\|\psi_0(Y_N^0 - y)\|_{\mathcal{H}}^2\big] = \varepsilon_N \leq a(1+4c)^{N-1} \\ &\leq C^2(1+4c)^N \Bigg[\min\Bigg(\Bigg\{\min\{M_{l_k,0}\mathfrak{c}_k, M_{l_k,1}\mathfrak{c}_{k+1}\} \prod_{j=1}^{k} M_{l_{j-1},l_j+1} \colon \\ &\qquad \begin{array}{l} k \in \mathbb{N} \cap [0, N-1],\ (l_i)_{i\in\{0,1,\ldots,k\}} \subseteq \\ \{1, 2, \ldots, N\},\ N = l_0 > l_1 > \ldots > l_k \end{array} \Bigg\} \cup \big\{\min\{M_{N,0}\mathfrak{c}_0, M_{N,1}\mathfrak{c}_1\}\big\}\Bigg)\Bigg]^{-1} < \infty. \end{aligned} \tag{3.53}$$

The proof of Proposition 3.9 is thus complete. □

Corollary 3.10. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be a separable $\mathbb{R}$-Banach space, let $C, c \in (0, \infty)$, $(\mathfrak{c}_k)_{k\in\mathbb{N}_0} \subseteq (0, \infty)$, $\Theta = \bigcup_{n=1}^\infty \mathbb{Z}^n$, $M \in \mathbb{N}$, $y \in \mathcal{Y}$, let $(\mathcal{Z}, \mathscr{Z})$ be a measurable space, let $Z^\theta \colon \Omega \to \mathcal{Z}$, $\theta \in \Theta$, be i.i.d. $\mathscr{F}/\mathscr{Z}$-measurable functions, let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}}, \|\cdot\|_{\mathcal{H}})$ be a separable $\mathbb{R}$-Hilbert space, let $\mathscr{S} = \sigma_{L(\mathcal{Y},\mathcal{H})}\big(\{\{\varphi \in L(\mathcal{Y}, \mathcal{H}) \colon \varphi(x) \in \mathcal{B}\} \subseteq L(\mathcal{Y}, \mathcal{H}) \colon x \in \mathcal{Y}, \mathcal{B} \in \mathscr{B}(\mathcal{H})\}\big)$, let $\psi_k \colon \Omega \to L(\mathcal{Y}, \mathcal{H})$, $k \in \mathbb{N}_0$, be $\mathscr{F}/\mathscr{S}$-measurable functions, let $\Phi_l \colon \mathcal{Y}\times\mathcal{Y}\times\mathcal{Z} \to \mathcal{Y}$, $l \in \mathbb{N}_0$, be $(\mathscr{B}(\mathcal{Y})\otimes\mathscr{B}(\mathcal{Y})\otimes\mathscr{Z})/\mathscr{B}(\mathcal{Y})$-measurable functions, let $Y_{-1}^\theta \colon \Omega \to \mathcal{Y}$, $\theta \in \Theta$, be i.i.d. $\mathscr{F}/\mathscr{B}(\mathcal{Y})$-measurable functions, let $Y_0^\theta \colon \Omega \to \mathcal{Y}$, $\theta \in \Theta$, be i.i.d. $\mathscr{F}/\mathscr{B}(\mathcal{Y})$-measurable functions, assume that $(Y_{-1}^\theta)_{\theta\in\Theta}$, $(Y_0^\theta)_{\theta\in\Theta}$, $(Z^\theta)_{\theta\in\Theta}$, and $(\psi_k)_{k\in\mathbb{N}_0}$ are independent, let $Y_n^\theta \colon \Omega \to \mathcal{Y}$, $\theta \in \Theta$, $n \in \mathbb{N}$, satisfy for all $n \in \mathbb{N}$, $\theta \in \Theta$ that*

$$Y_n^\theta = \sum_{l=0}^{n-1} \tfrac{1}{M^{n-l}} \left[\sum_{i=1}^{M^{n-l}} \Phi_l\big(Y_l^{(\theta,l,i)}, Y_{l-1}^{(\theta,-l,i)}, Z^{(\theta,l,i)}\big)\right], \tag{3.54}$$

and assume for all $k \in \mathbb{N}_0$, $n \in \mathbb{N}$ that $\mathbb{E}\big[\|\Phi_k(Y_k^0, Y_{k-1}^1, Z^0)\|_{\mathcal{Y}}\big] < \infty$ and

$$\max\big\{\mathbb{E}\big[\|\psi_k(\Phi_0(Y_0^0, Y_{-1}^1, Z^0))\|_{\mathcal{H}}^2\big], \mathbb{1}_{\mathbb{N}}(k)\,\mathbb{E}\big[\|\psi_k(Y_0^0 - y)\|_{\mathcal{H}}^2\big]\big\} \leq \tfrac{C^2}{\mathfrak{c}_k}, \tag{3.55}$$

$$\mathbb{E}\big[\|\psi_k(\Phi_n(Y_n^0, Y_{n-1}^1, Z^0))\|_{\mathcal{H}}^2\big] \leq c\,\mathbb{E}\big[\|\psi_{k+1}(Y_n^0 - Y_{n-1}^1)\|_{\mathcal{H}}^2\big], \tag{3.56}$$

$$\mathbb{E}\left[\left\|\psi_k\left(y - \sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l(Y_l^0, Y_{l-1}^1, Z^0)\big]\right)\right\|_{\mathcal{H}}^2\right] \leq 2c\,\mathbb{E}\big[\|\psi_{k+1}(Y_{n-1}^0 - y)\|_{\mathcal{H}}^2\big]. \tag{3.57}$$

Then it holds for all $N \in \mathbb{N}$ that

$$\big(\mathbb{E}\big[\|\psi_0(Y_N^0 - y)\|_{\mathcal{H}}^2\big]\big)^{1/2} \leq C\big[\tfrac{1+4c}{M}\big]^{N/2} \max_{k\in\{0,1,\ldots,N\}} \sqrt{\tfrac{M^k}{\mathfrak{c}_k}} < \infty. \tag{3.58}$$

Proof of Corollary 3.10. Throughout this proof let $\mathfrak{M}_{n,l} \in \mathbb{N}$, $l \in \{0, 1, \ldots, n\}$, $n \in \mathbb{N}$, be the natural numbers which satisfy for all $n \in \mathbb{N}$, $l \in \{0, 1, \ldots, n\}$ that $\mathfrak{M}_{n,l} = M^{n-l}$. Note that it holds for all $n \in \mathbb{N}$ that $\mathfrak{M}_{n,1} \geq \mathfrak{M}_{n,2} \geq \ldots \geq \mathfrak{M}_{n,n}$. The fact that $\forall\, n \in \mathbb{N}\colon \mathfrak{M}_{n,n} = 1$ and Proposition 3.9 hence ensure for all $N \in \mathbb{N}$ that

$$\begin{aligned}
&\big(\mathbb{E}\big[\|\psi_0(Y_N^0 - y)\|_{\mathcal{H}}^2\big]\big)^{1/2} \\
&\leq C(1+4c)^{N/2}\Bigg[\min\Bigg(\Bigg\{\min\{\mathfrak{M}_{l_k,0}\mathfrak{c}_k, \mathfrak{M}_{l_k,1}\mathfrak{c}_{k+1}\}\prod_{j=1}^{k}\mathfrak{M}_{l_{j-1},l_j+1}\colon \\
&\quad \begin{array}{l} k \in \mathbb{N}\cap[0, N-1],\ (l_i)_{i\in\{0,1,\ldots,k\}} \subseteq \\ \{1, 2, \ldots, N\},\ N = l_0 > l_1 > \ldots > l_k \end{array}\Bigg\} \cup \big\{\min\{\mathfrak{M}_{N,0}\mathfrak{c}_0, \mathfrak{M}_{N,1}\mathfrak{c}_1\}\big\}\Bigg)\Bigg]^{-1/2}.
\end{aligned} \tag{3.59}$$

Next observe that it holds for all $N \in \mathbb{N}$, $k \in \mathbb{N}\cap[0, N-1]$, $(l_i)_{i\in\{0,1,\ldots,k\}} \subseteq \{1, 2, \ldots, N\}$ with $l_0 = N$ and $\forall\, i \in \{1, 2, \ldots, k\}\colon l_{i-1} > l_i$ that

$$\begin{aligned}
&\min\{\mathfrak{M}_{l_k,0}\mathfrak{c}_k, \mathfrak{M}_{l_k,1}\mathfrak{c}_{k+1}\}\prod_{j=1}^{k}\mathfrak{M}_{l_{j-1},l_j+1} = \min\{M^{l_k}\mathfrak{c}_k, M^{l_k-1}\mathfrak{c}_{k+1}\}\prod_{j=1}^{k}M^{l_{j-1}-l_j-1} \\
&\quad = \min\{M^{l_k}\mathfrak{c}_k, M^{l_k-1}\mathfrak{c}_{k+1}\}M^{l_0-l_k-k} = \min\{M^{N-k}\mathfrak{c}_k, M^{N-(k+1)}\mathfrak{c}_{k+1}\}.
\end{aligned} \tag{3.60}$$

This and (3.59) establish for all $N \in \mathbb{N}$ that

$$\begin{aligned}
&\big(\mathbb{E}\big[\|\psi_0(Y_N^0 - y)\|_{\mathcal{H}}^2\big]\big)^{1/2} \\
&\leq C(1+4c)^{N/2}\Big[\min\Big(\big\{\min\{M^{N-k}\mathfrak{c}_k, M^{N-(k+1)}\mathfrak{c}_{k+1}\}\colon k \in \mathbb{N}\cap[0, N-1]\big\} \\
&\qquad \cup \big\{\min\{M^N\mathfrak{c}_0, M^{N-1}\mathfrak{c}_1\}\big\}\Big)\Big]^{-1/2} \\
&= C(1+4c)^{N/2}\Big[\min_{k\in\{0,1,\ldots,N-1\}}\min\{M^{N-k}\mathfrak{c}_k, M^{N-(k+1)}\mathfrak{c}_{k+1}\}\Big]^{-1/2} \\
&= C(1+4c)^{N/2}\Big[\min_{k\in\{0,1,\ldots,N\}}(M^{N-k}\mathfrak{c}_k)\Big]^{-1/2} \\
&= C\big[\tfrac{1+4c}{M}\big]^{N/2}\Big[\min_{k\in\{0,1,\ldots,N\}}\tfrac{\mathfrak{c}_k}{M^k}\Big]^{-1/2} \\
&= C\big[\tfrac{1+4c}{M}\big]^{N/2}\max_{k\in\{0,1,\ldots,N\}}\sqrt{\tfrac{M^k}{\mathfrak{c}_k}} < \infty.
\end{aligned} \tag{3.61}$$

The proof of Corollary 3.10 is thus complete. □

3.1.5 Cost analysis

Proposition 3.11. *Let $M \in (0,\infty)$, $(\alpha_l)_{l\in\mathbb{N}_0}, (\beta_l)_{l\in\mathbb{N}_0}, (\gamma_l)_{l\in\mathbb{N}_0}, (\mathrm{Cost}_n)_{n\in\mathbb{N}_0\cup\{-1\}} \subseteq [0,\infty)$ satisfy for all $n \in \mathbb{N}$ that*

$$\mathrm{Cost}_n \leq \sum_{l=0}^{n-1}\big[M^{n-l}(\alpha_l\mathrm{Cost}_l + \beta_l\mathrm{Cost}_{l-1} + \gamma_l)\big]. \tag{3.62}$$

Then it holds for all $n \in \mathbb{N}$ that

$$\mathrm{Cost}_n \leq M^n\Bigg(\beta_0\mathrm{Cost}_{-1} + \big(\alpha_0 + \tfrac{\beta_1}{M}\big)\mathrm{Cost}_0 + \sum_{l=0}^{n-1}[M^{-l}\gamma_l]\Bigg)\prod_{l=1}^{n-1}\big(1 + \alpha_l + \tfrac{\beta_{l+1}}{M}\big). \tag{3.63}$$

Proof of Proposition 3.11. Observe that it holds for all $n \in \mathbb{N}$ that

$$
\begin{aligned}
\mathrm{Cost}_n &\le M^n \sum_{l=0}^{n-1} \big[M^{-l}(\alpha_l \mathrm{Cost}_l + \beta_l \mathrm{Cost}_{l-1} + \gamma_l) \big] \\
&= M^n \left(\beta_0 \mathrm{Cost}_{-1} + \left[\sum_{l=0}^{n-1} M^{-l}(\alpha_l \mathrm{Cost}_l + \gamma_l) \right] + \left[\sum_{l=1}^{n-1} M^{-l} \beta_l \mathrm{Cost}_{l-1} \right] \right) \\
&= M^n \left(\beta_0 \mathrm{Cost}_{-1} + \left[\sum_{l=0}^{n-1} M^{-l}(\alpha_l \mathrm{Cost}_l + \gamma_l) \right] + \tfrac{1}{M} \left[\sum_{l=0}^{n-2} M^{-l} \beta_{l+1} \mathrm{Cost}_{l} \right] \right) \\
&\le M^n \left(\beta_0 \mathrm{Cost}_{-1} + \left[\sum_{l=0}^{n-1} M^{-l} \gamma_l \right] + \left[\sum_{l=0}^{n-1} M^{-l} \alpha_l \mathrm{Cost}_l \right] + \tfrac{1}{M} \left[\sum_{l=0}^{n-1} M^{-l} \beta_{l+1} \mathrm{Cost}_{l} \right] \right) \\
&= M^n \left(\beta_0 \mathrm{Cost}_{-1} + \left[\sum_{l=0}^{n-1} M^{-l} \gamma_l \right] + \left[\sum_{l=0}^{n-1} M^{-l} \big(\alpha_l + \tfrac{\beta_{l+1}}{M}\big) \mathrm{Cost}_l \right] \right) \qquad (3.64) \\
&= M^n \left(\beta_0 \mathrm{Cost}_{-1} + \big(\alpha_0 + \tfrac{\beta_1}{M}\big) \mathrm{Cost}_0 + \sum_{l=0}^{n-1} [M^{-l} \gamma_l] \right) + M^n \left[\sum_{l=1}^{n-1} M^{-l} \big(\alpha_l + \tfrac{\beta_{l+1}}{M}\big) \mathrm{Cost}_l \right].
\end{aligned}
$$

Theorem 4.1.1 in Agarwal [2] (with $a \leftarrow 1$, $u(k) \leftarrow \mathrm{Cost}_k$, $p(k) \leftarrow M^k(\beta_0 \mathrm{Cost}_{-1} + (\alpha_0 + \frac{\beta_1}{M})\mathrm{Cost}_0 + \sum_{l=0}^{k-1}[M^{-l}\gamma_l])$, $q(k) \leftarrow M^k$, $f(k) \leftarrow M^{-k}(\alpha_k + \frac{\beta_{k+1}}{M})$ for $k \in \mathbb{N}$ in the notation of [2, Theorem 4.1.1]) hence establishes for all $n \in \mathbb{N}$ that

$$
\begin{aligned}
\mathrm{Cost}_n &\le M^n \left(\beta_0 \mathrm{Cost}_{-1} + \big(\alpha_0 + \tfrac{\beta_1}{M}\big) \mathrm{Cost}_0 + \sum_{l=0}^{n-1} [M^{-l} \gamma_l] \right) \\
&\quad + M^n \sum_{l=1}^{n-1} \Bigg[M^l \left(\beta_0 \mathrm{Cost}_{-1} + \big(\alpha_0 + \tfrac{\beta_1}{M}\big) \mathrm{Cost}_0 + \sum_{i=0}^{l-1} [M^{-i} \gamma_i] \right) M^{-l} \big(\alpha_l + \tfrac{\beta_{l+1}}{M}\big) \\
&\qquad \cdot \prod_{i=l+1}^{n-1} \big[1 + M^i M^{-i} \big(\alpha_i + \tfrac{\beta_{i+1}}{M}\big) \big] \Bigg] \qquad (3.65) \\
&\le M^n \left(\beta_0 \mathrm{Cost}_{-1} + \big(\alpha_0 + \tfrac{\beta_1}{M}\big) \mathrm{Cost}_0 + \sum_{l=0}^{n-1} [M^{-l} \gamma_l] \right) \\
&\quad \cdot \left(1 + \sum_{l=1}^{n-1} \left[\big(\alpha_l + \tfrac{\beta_{l+1}}{M}\big) \prod_{i=l+1}^{n-1} \big(1 + \alpha_i + \tfrac{\beta_{i+1}}{M}\big) \right] \right).
\end{aligned}
$$

This and [2, Problem 1.9.10] show for all $n \in \mathbb{N}$ that

$$
\mathrm{Cost}_n \le M^n \left(\beta_0 \mathrm{Cost}_{-1} + \big(\alpha_0 + \tfrac{\beta_1}{M}\big) \mathrm{Cost}_0 + \sum_{l=0}^{n-1} [M^{-l} \gamma_l] \right) \prod_{l=1}^{n-1} \big(1 + \alpha_l + \tfrac{\beta_{l+1}}{M}\big). \qquad (3.66)
$$

The proof of Proposition 3.11 is thus complete. □

Lemma 3.12. *Let $a, b \in [0, \infty)$. Then it holds for all $n \in \mathbb{N}$ that*

$$
(an + b) b^{n-1} \le (a + b)^n. \qquad (3.67)
$$

Proof of Lemma 3.12. We prove (3.67) by induction on $n \in \mathbb{N}$. Note that the base case $n = 1$ is clear. For the induction step $\mathbb{N} \ni n-1 \to n \in \{2, 3, \ldots\}$ let $n \in \{2, 3, \ldots\}$ and assume that $(a(n-1)+b)b^{n-2} \leq (a+b)^{n-1}$. This ensures that

$$\begin{aligned}(an+b)b^{n-1} &= ab^{n-1} + (a(n-1)+b)b^{n-1} \leq ab^{n-1} + b(a+b)^{n-1} \\ &\leq a(a+b)^{n-1} + b(a+b)^{n-1} = (a+b)^n.\end{aligned} \tag{3.68}$$

Induction hence completes the proof of Lemma 3.12. □

Corollary 3.13. *Let $M \in [1, \infty)$, $\mathfrak{z}, \alpha, \beta, \gamma \in [0, \infty)$, $(\mathrm{Cost}_n)_{n \in \mathbb{N}_0 \cup \{-1\}} \subseteq [0, \infty)$ satisfy for all $n \in \mathbb{N}$ that $\mathrm{Cost}_{-1} = \mathrm{Cost}_0 = 0$ and*

$$\mathrm{Cost}_n \leq M^n \mathfrak{z} + \sum_{l=0}^{n-1} \bigl[M^{n-l}(\alpha \mathrm{Cost}_l + \beta \mathrm{Cost}_{l-1} + \gamma \mathfrak{z})\bigr]. \tag{3.69}$$

Then it holds for all $n \in \mathbb{N}$ that

$$\mathrm{Cost}_n \leq (1 + \alpha + \beta + \gamma)^n M^n \mathfrak{z}. \tag{3.70}$$

Proof of Corollary 3.13. Note that Proposition 3.11 demonstrates for all $n \in \mathbb{N}$ that

$$\begin{aligned}\mathrm{Cost}_n &\leq M^n \left(\mathfrak{z} + \gamma \mathfrak{z} \sum_{l=0}^{n-1} M^{-l}\right) \prod_{l=1}^{n-1} \left(1 + \alpha + \tfrac{\beta}{M}\right) \\ &\leq \left(1 + \gamma \sum_{l=0}^{n-1} M^{-l}\right) (1 + \alpha + \beta)^{n-1} M^n \mathfrak{z}.\end{aligned} \tag{3.71}$$

In addition, observe that it holds for all $n \in \mathbb{N}$ that

$$\sum_{l=0}^{n-1} M^{-l} \leq \sum_{l=0}^{n-1} 1 = n. \tag{3.72}$$

Furthermore, Lemma 3.12 implies for all $n \in \mathbb{N}$ that

$$(1 + \gamma n)(1 + \alpha + \beta)^{n-1} \leq (\gamma n + 1 + \alpha + \beta)(1 + \alpha + \beta)^{n-1} \leq (1 + \alpha + \beta + \gamma)^n. \tag{3.73}$$

This, (3.71), and (3.72) prove for all $n \in \mathbb{N}$ that

$$\mathrm{Cost}_n \leq (1 + \gamma n)(1 + \alpha + \beta)^{n-1} M^n \mathfrak{z} \leq (1 + \alpha + \beta + \gamma)^n M^n \mathfrak{z}. \tag{3.74}$$

The proof of Corollary 3.13 is thus complete. □

3.1.6 Complexity analysis

Theorem 3.14. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be a separable $\mathbb{R}$-Banach space, let $\mathfrak{z}, \gamma \in [0, \infty)$, $\mathfrak{B}, \mathfrak{b}, C \in [1, \infty)$, $c \in (0, \infty)$, $(\mathfrak{c}_k)_{k \in \mathbb{N}_0} \subseteq (0, \infty)$, $\Theta = \bigcup_{n=1}^{\infty} \mathbb{Z}^n$, $(M_j)_{j \in \mathbb{N}} \subseteq \mathbb{N}$, $y, \mathfrak{y}_{-1}, \mathfrak{y}_0 \in \mathcal{Y}$ satisfy $\liminf_{j \to \infty} M_j = \infty$, $\sup_{j \in \mathbb{N}} {}^{M_{j+1}}/_{M_j} \leq \mathfrak{B}$, and $\forall\, n \in \mathbb{N}\colon \max_{k \in \{0,1,\ldots,n\}} {}^{(M_n)^k}/_{\mathfrak{c}_k} \leq \mathfrak{b}^n$, let $(\mathcal{Z}, \mathscr{Z})$ be a measurable space, let $Z^\theta \colon \Omega \to \mathcal{Z}$, $\theta \in \Theta$, be i.i.d. $\mathscr{F}/\mathscr{Z}$-measurable functions, let $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}}, \|\cdot\|_{\mathcal{H}})$ be a separable $\mathbb{R}$-Hilbert space,* *let* $\mathscr{S} = \sigma_{L(\mathcal{Y},\mathcal{H})}(\{\{\varphi \in L(\mathcal{Y},\mathcal{H})\colon \varphi(x) \in \mathcal{B}\} \subseteq L(\mathcal{Y},\mathcal{H})\colon x \in \mathcal{Y}, \mathcal{B} \in \mathscr{B}(\mathcal{H})\})$, *let* $\psi_k\colon \Omega \to L(\mathcal{Y},\mathcal{H})$, $k \in \mathbb{N}_0$, *be* $\mathscr{F}/\mathscr{S}$*-measurable functions, assume that* $(Z^\theta)_{\theta\in\Theta}$ *and* $(\psi_k)_{k\in\mathbb{N}_0}$ *are independent, let* $\Phi_l\colon \mathcal{Y}\times\mathcal{Y}\times\mathcal{Z} \to \mathcal{Y}$, $l \in \mathbb{N}_0$, *be* $(\mathscr{B}(\mathcal{Y})\otimes\mathscr{B}(\mathcal{Y})\otimes\mathscr{Z})/\mathscr{B}(\mathcal{Y})$*-measurable functions, let* $Y^\theta_{n,j}\colon \Omega \to \mathcal{Y}$, $\theta \in \Theta$, $j \in \mathbb{N}$, $n \in (\mathbb{N}_0 \cup \{-1\})$, *satisfy for all* $n, j \in \mathbb{N}$, $\theta \in \Theta$ *that* $Y^\theta_{-1,j} = \mathfrak{y}_{-1}$, $Y^\theta_{0,j} = \mathfrak{y}_0$, *and*

$$Y^\theta_{n,j} = \sum_{l=0}^{n-1} \tfrac{1}{(M_j)^{n-l}} \left[\sum_{i=1}^{(M_j)^{n-l}} \Phi_l\big(Y^{(\theta,l,i)}_{l,j}, Y^{(\theta,-l,i)}_{l-1,j}, Z^{(\theta,l,i)}\big)\right], \tag{3.75}$$

let $(\mathrm{Cost}_{n,j})_{(n,j)\in(\mathbb{N}_0\cup\{-1\})\times\mathbb{N}} \subseteq [0,\infty)$ *satisfy for all* $n, j \in \mathbb{N}$ *that* $\mathrm{Cost}_{-1,j} = \mathrm{Cost}_{0,j} = 0$ *and*

$$\mathrm{Cost}_{n,j} \leq (M_j)^n \mathfrak{z} + \sum_{l=0}^{n-1} \big[(M_j)^{n-l}(\mathrm{Cost}_{l,j} + \mathrm{Cost}_{l-1,j} + \gamma\mathfrak{z})\big], \tag{3.76}$$

and assume for all $k \in \mathbb{N}_0$, $n, j \in \mathbb{N}$ *that* $\mathbb{E}\big[\|\Phi_k(Y^0_{k,j}, Y^1_{k-1,j}, Z^0)\|_{\mathcal{Y}}\big] < \infty$ *and*

$$\max\big\{\mathbb{E}\big[\|\psi_k(\Phi_0(\mathfrak{y}_0, \mathfrak{y}_{-1}, Z^0))\|^2_{\mathcal{H}}\big], \mathbb{1}_{\mathbb{N}}(k)\,\mathbb{E}\big[\|\psi_k(\mathfrak{y}_0 - y)\|^2_{\mathcal{H}}\big]\big\} \leq \tfrac{C^2}{\mathfrak{c}_k}, \tag{3.77}$$

$$\mathbb{E}\Big[\big\|\psi_k\big(\Phi_n\big(Y^0_{n,j}, Y^1_{n-1,j}, Z^0\big)\big)\big\|^2_{\mathcal{H}}\Big] \leq c\,\mathbb{E}\Big[\big\|\psi_{k+1}\big(Y^0_{n,j} - Y^1_{n-1,j}\big)\big\|^2_{\mathcal{H}}\Big], \tag{3.78}$$

$$\mathbb{E}\left[\left\|\psi_k\left(y - \sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l\big(Y^0_{l,j}, Y^1_{l-1,j}, Z^0\big)\big]\right)\right\|^2_{\mathcal{H}}\right] \leq 2c\,\mathbb{E}\Big[\big\|\psi_{k+1}\big(Y^0_{n-1,j} - y\big)\big\|^2_{\mathcal{H}}\Big]. \tag{3.79}$$

Then

(i) *it holds for all* $n \in \mathbb{N}$ *that*

$$\big(\mathbb{E}\big[\|\psi_0(Y^0_{n,n} - y)\|^2_{\mathcal{H}}\big]\big)^{1/2} \leq C\left[\frac{\mathfrak{b}(1+4c)}{M_n}\right]^{n/2} < \infty, \tag{3.80}$$

(ii) *it holds for all* $n \in \mathbb{N}$ *that* $\mathrm{Cost}_{n,n} \leq (3+\gamma)^n (M_n)^n \mathfrak{z}$, *and*

(iii) *there exists* $(N_\varepsilon)_{\varepsilon\in(0,1]} \subseteq \mathbb{N}$ *such that it holds for all* $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ *that* $\sup_{n\in\{N_\varepsilon,N_\varepsilon+1,\ldots\}} \big(\mathbb{E}\big[\|\psi_0(Y^0_{n,n} - y)\|^2_{\mathcal{H}}\big]\big)^{1/2} \leq \varepsilon$ *and*

$$\mathrm{Cost}_{N_\varepsilon,N_\varepsilon} \leq \mathfrak{z}\mathfrak{b}\mathfrak{c}_1(3+\gamma)C^{2(1+\delta)}\left(1 + \sup_{n\in\mathbb{N}}\left[\frac{[\mathfrak{B}\mathfrak{b}^2(3+\gamma)(1+4c)]^{(1+\delta)}}{(M_n)^\delta}\right]^n\right)\varepsilon^{-2(1+\delta)} < \infty. \tag{3.81}$$

Proof of Theorem 3.14. Throughout this proof let $(N_\varepsilon)_{\varepsilon\in(0,1]} \subseteq \mathbb{N}$ be the family of natural numbers which satisfies for all $\varepsilon \in (0,1]$ that

$$N_\varepsilon = \min\left\{\mathfrak{N} \in \mathbb{N}\colon \sup_{n\in\{\mathfrak{N},\mathfrak{N}+1,\ldots\}} C\left[\frac{\mathfrak{b}(1+4c)}{M_n}\right]^{n/2} \leq \varepsilon\right\}. \tag{3.82}$$

Observe that Corollary 3.10 and the assumption that $\forall\, n \in \mathbb{N}\colon \max_{k\in\{0,1,\ldots,n\}} {}^{(M_n)^k}/_{\mathfrak{c}_k} \leq \mathfrak{b}^n$ establish for all $n \in \mathbb{N}$ that

$$\big(\mathbb{E}\big[\|\psi_0(Y^0_{n,n} - y)\|^2_{\mathcal{H}}\big]\big)^{1/2} \leq C\left[\frac{1+4c}{M_n}\right]^{n/2} \max_{k\in\{0,1,\ldots,n\}} \sqrt{\tfrac{(M_n)^k}{\mathfrak{c}_k}} \leq C\left[\frac{\mathfrak{b}(1+4c)}{M_n}\right]^{n/2} < \infty. \tag{3.83}$$

This proves (i). In addition, (3.76) and Corollary 3.13 demonstrate for all $n \in \mathbb{N}$ that

$$\mathrm{Cost}_{n,n} \leq (3+\gamma)^n (M_n)^n \mathfrak{z}. \tag{3.84}$$

This finishes the proof of (ii). It thus remains to show (iii). Observe that (3.83) and (3.82) ensure for all $\varepsilon \in (0,1]$ that

$$\sup_{n \in \{N_\varepsilon, N_\varepsilon+1,\ldots\}} \big(\mathbb{E}\big[\|\psi_0(Y^0_{n,n} - y)\|^2_{\mathcal{H}}\big]\big)^{1/2} \leq \sup_{n \in \{N_\varepsilon, N_\varepsilon+1,\ldots\}} C \left[\frac{\mathfrak{b}(1+4c)}{M_n}\right]^{n/2} \leq \varepsilon. \tag{3.85}$$

Furthermore, note that (3.82) implies for all $\varepsilon \in (0,1]$ with $N_\varepsilon \geq 2$ that

$$C \left[\frac{\mathfrak{b}(1+4c)}{M_{N_\varepsilon - 1}}\right]^{(N_\varepsilon-1)/2} > \varepsilon. \tag{3.86}$$

This, (3.84), the assumption that $\sup_{j\in\mathbb{N}} {}^{M_{j+1}}/_{M_j} \leq \mathfrak{B}$, and the fact that $\forall\, n \in \mathbb{N}\colon {}^{M_n}/_{\mathfrak{c}_1} \leq \mathfrak{b}^n$ show for all $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ with $N_\varepsilon \geq 2$ that

$$\begin{aligned}
\mathrm{Cost}_{N_\varepsilon,N_\varepsilon} &\leq (3+\gamma)^{N_\varepsilon}(M_{N_\varepsilon})^{N_\varepsilon}\mathfrak{z} \\
&\leq (3+\gamma)^{N_\varepsilon}(M_{N_\varepsilon})^{N_\varepsilon}\mathfrak{z}\left[C\left[\frac{\mathfrak{b}(1+4c)}{M_{N_\varepsilon-1}}\right]^{(N_\varepsilon-1)/2}\varepsilon^{-1}\right]^{2(1+\delta)} \\
&= \mathfrak{z}C^{2(1+\delta)}\varepsilon^{-2(1+\delta)}\left[\frac{(3+\gamma)^{N_\varepsilon}(M_{N_\varepsilon})^{N_\varepsilon}[\mathfrak{b}(1+4c)]^{(N_\varepsilon-1)(1+\delta)}}{(M_{N_\varepsilon-1})^{(N_\varepsilon-1)(1+\delta)}}\right] \\
&\leq \mathfrak{z}C^{2(1+\delta)}\varepsilon^{-2(1+\delta)}\sup_{n\in\mathbb{N}}\left[\frac{(3+\gamma)^{n+1}(M_{n+1})^{n+1}[\mathfrak{b}(1+4c)]^{n(1+\delta)}}{(M_n)^{n(1+\delta)}}\right] \\
&\leq \mathfrak{z}(3+\gamma)C^{2(1+\delta)}\varepsilon^{-2(1+\delta)}\sup_{n\in\mathbb{N}}\left[\frac{M_{n+1}(M_{n+1})^n[\mathfrak{b}(3+\gamma)(1+4c)]^{n(1+\delta)}}{(M_n)^n(M_n)^{n\delta}}\right] \\
&\leq \mathfrak{z}\mathfrak{c}_1(3+\gamma)C^{2(1+\delta)}\varepsilon^{-2(1+\delta)}\sup_{n\in\mathbb{N}}\left[\frac{\mathfrak{b}^{n+1}\mathfrak{B}^n[\mathfrak{b}(3+\gamma)(1+4c)]^{n(1+\delta)}}{(M_n)^{n\delta}}\right] \\
&\leq \mathfrak{z}\mathfrak{b}\mathfrak{c}_1(3+\gamma)C^{2(1+\delta)}\varepsilon^{-2(1+\delta)}\sup_{n\in\mathbb{N}}\left[\frac{[\mathfrak{B}\mathfrak{b}^2(3+\gamma)(1+4c)]^{(1+\delta)}}{(M_n)^\delta}\right]^n \\
&\leq \mathfrak{z}\mathfrak{b}\mathfrak{c}_1(3+\gamma)C^{2(1+\delta)}\left(1+\sup_{n\in\mathbb{N}}\left[\frac{[\mathfrak{B}\mathfrak{b}^2(3+\gamma)(1+4c)]^{(1+\delta)}}{(M_n)^\delta}\right]^n\right)\varepsilon^{-2(1+\delta)}.
\end{aligned} \tag{3.87}$$

Moreover, (3.76), the fact that ${}^{M_1}/_{\mathfrak{c}_1} \leq \mathfrak{b}$, and the fact that $C \geq 1$ ensure for all $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ that

$$\begin{aligned}
\mathrm{Cost}_{1,1} &\leq \mathfrak{z}(1+\gamma)M_1 \leq \mathfrak{z}\mathfrak{b}\mathfrak{c}_1(3+\gamma) \\
&\leq \mathfrak{z}\mathfrak{b}\mathfrak{c}_1(3+\gamma)C^{2(1+\delta)}\left(1+\sup_{n\in\mathbb{N}}\left[\frac{[\mathfrak{B}\mathfrak{b}^2(3+\gamma)(1+4c)]^{(1+\delta)}}{(M_n)^\delta}\right]^n\right)\varepsilon^{-2(1+\delta)}.
\end{aligned} \tag{3.88}$$

Combining this with (3.87) establishes for all $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ that

$$\mathrm{Cost}_{N_\varepsilon,N_\varepsilon} \leq \mathfrak{z}\mathfrak{b}\mathfrak{c}_1(3+\gamma)C^{2(1+\delta)}\left(1+\sup_{n\in\mathbb{N}}\left[\frac{[\mathfrak{B}\mathfrak{b}^2(3+\gamma)(1+4c)]^{(1+\delta)}}{(M_n)^\delta}\right]^n\right)\varepsilon^{-2(1+\delta)} < \infty. \tag{3.89}$$

The proof of Theorem 3.14 is thus complete. □

Corollary 3.15. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be a separable $\mathbb{R}$-Banach space, let $\mathfrak{z}, \gamma \in [0,\infty)$, $\mathfrak{B}, \kappa, C \in [1,\infty)$, $c \in (0,\infty)$, $\Theta = \bigcup_{n=1}^{\infty} \mathbb{Z}^n$, $(M_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$, $y, \mathfrak{y}_{-1}, \mathfrak{y}_0 \in \mathcal{Y}$ satisfy $\liminf_{j\to\infty} M_j = \infty$, $\sup_{j\in\mathbb{N}} {}^{M_{j+1}}/{}_{M_j} \leq \mathfrak{B}$, and $\sup_{j\in\mathbb{N}} {}^{M_j}/{}_j \leq \kappa$, let $(\mathcal{Z}, \mathscr{Z})$ be a measurable space, let $Z^\theta \colon \Omega \to \mathcal{Z}$, $\theta \in \Theta$, be i.i.d. $\mathscr{F}/\mathscr{Z}$-measurable functions, let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}}, \|\cdot\|_{\mathcal{H}})$ be a separable $\mathbb{R}$-Hilbert space, let $\mathscr{S} = \sigma_{L(\mathcal{Y},\mathcal{H})}\big(\big\{\{\varphi \in L(\mathcal{Y},\mathcal{H}) \colon \varphi(x) \in \mathcal{B}\} \subseteq L(\mathcal{Y},\mathcal{H}) \colon x \in \mathcal{Y}, \mathcal{B} \in \mathscr{B}(\mathcal{H})\big\}\big)$, let $\psi_k \colon \Omega \to L(\mathcal{Y},\mathcal{H})$, $k \in \mathbb{N}_0$, be $\mathscr{F}/\mathscr{S}$-measurable functions, assume that $(Z^\theta)_{\theta\in\Theta}$ and $(\psi_k)_{k\in\mathbb{N}_0}$ are independent, let $\Phi_l \colon \mathcal{Y} \times \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y}$, $l \in \mathbb{N}_0$, be $(\mathscr{B}(\mathcal{Y}) \otimes \mathscr{B}(\mathcal{Y}) \otimes \mathscr{Z})/\mathscr{B}(\mathcal{Y})$-measurable functions, let $Y^\theta_{n,j} \colon \Omega \to \mathcal{Y}$, $\theta \in \Theta$, $j \in \mathbb{N}$, $n \in (\mathbb{N}_0 \cup \{-1\})$, satisfy for all $n, j \in \mathbb{N}$, $\theta \in \Theta$ that $Y^\theta_{-1,j} = \mathfrak{y}_{-1}$, $Y^\theta_{0,j} = \mathfrak{y}_0$, and*

$$Y^\theta_{n,j} = \sum_{l=0}^{n-1} \tfrac{1}{(M_j)^{n-l}} \left[\sum_{i=1}^{(M_j)^{n-l}} \Phi_l\big(Y^{(\theta,l,i)}_{l,j}, Y^{(\theta,-l,i)}_{l-1,j}, Z^{(\theta,l,i)}\big) \right], \tag{3.90}$$

let $(\mathrm{Cost}_{n,j})_{(n,j)\in(\mathbb{N}_0\cup\{-1\})\times\mathbb{N}} \subseteq [0,\infty)$ satisfy for all $n, j \in \mathbb{N}$ that $\mathrm{Cost}_{-1,j} = \mathrm{Cost}_{0,j} = 0$ and

$$\mathrm{Cost}_{n,j} \leq (M_j)^n \mathfrak{z} + \sum_{l=0}^{n-1} \big[(M_j)^{n-l}(\mathrm{Cost}_{l,j} + \mathrm{Cost}_{l-1,j} + \gamma\mathfrak{z})\big], \tag{3.91}$$

and assume for all $k \in \mathbb{N}_0$, $n, j \in \mathbb{N}$, $u, v \in \mathcal{Y}$ that $\mathbb{E}\big[\|\Phi_k(Y^0_{k,j}, Y^1_{k-1,j}, Z^0)\|_{\mathcal{Y}}\big] < \infty$ and

$$\max\big\{\mathbb{E}\big[\|\psi_k(\Phi_0(\mathfrak{y}_0, \mathfrak{y}_{-1}, Z^0))\|^2_{\mathcal{H}}\big], \mathbb{1}_{\mathbb{N}}(k)\, \mathbb{E}\big[\|\psi_k(\mathfrak{y}_0 - y)\|^2_{\mathcal{H}}\big]\big\} \leq \tfrac{C^2}{k!}, \tag{3.92}$$

$$\mathbb{E}\big[\|\psi_k(\Phi_n(u, v, Z^0))\|^2_{\mathcal{H}}\big] \leq c\, \mathbb{E}\big[\|\psi_{k+1}(u - v)\|^2_{\mathcal{H}}\big], \tag{3.93}$$

$$\mathbb{E}\left[\left\|\psi_k\left(y - \sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l\big(Y^0_{l,j}, Y^1_{l-1,j}, Z^0\big)\big]\right)\right\|^2_{\mathcal{H}}\right] \leq 2c\, \mathbb{E}\Big[\big\|\psi_{k+1}\big(Y^0_{n-1,j} - y\big)\big\|^2_{\mathcal{H}}\Big]. \tag{3.94}$$

Then

(i) it holds for all $n \in \mathbb{N}$ that

$$\big(\mathbb{E}\big[\|\psi_0(Y^0_{n,n} - y)\|^2_{\mathcal{H}}\big]\big)^{1/2} \leq C \left[\frac{e^\kappa(1+4c)}{M_n}\right]^{n/2} < \infty, \tag{3.95}$$

(ii) it holds for all $n \in \mathbb{N}$ that $\mathrm{Cost}_{n,n} \leq (3+\gamma)^n (M_n)^n \mathfrak{z}$, and

(iii) there exists $(N_\varepsilon)_{\varepsilon\in(0,1]} \subseteq \mathbb{N}$ such that it holds for all $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ that $\sup_{n\in\{N_\varepsilon, N_\varepsilon+1,\ldots\}} \big(\mathbb{E}\big[\|\psi_0(Y^0_{n,n} - y)\|^2_{\mathcal{H}}\big]\big)^{1/2} \leq \varepsilon$ and

$$\mathrm{Cost}_{N_\varepsilon,N_\varepsilon} \leq \mathfrak{z}(3+\gamma)e^\kappa C^{2(1+\delta)} \left(1 + \sup_{n\in\mathbb{N}} \left[\frac{[\mathfrak{B}e^{2\kappa}(3+\gamma)(1+4c)]^{(1+\delta)}}{(M_n)^\delta}\right]^n\right) \varepsilon^{-2(1+\delta)} < \infty. \tag{3.96}$$

Proof of Corollary 3.15. Throughout this proof let $(\mathfrak{c}_k)_{k\in\mathbb{N}_0} \subseteq (0,\infty)$ be the family of real numbers which satisfies for all $k \in \mathbb{N}_0$ that $\mathfrak{c}_k = k!$. Note that the assumption that $\sup_{j\in\mathbb{N}} {}^{M_j}/{}_j \leq \kappa$ ensures for all $n \in \mathbb{N}$ that

$$\max_{k\in\{0,1,\ldots,n\}} \tfrac{(M_n)^k}{\mathfrak{c}_k} = \max_{k\in\{0,1,\ldots,n\}} \tfrac{(M_n)^k}{k!} \leq \sum_{k=0}^{\infty} \tfrac{(M_n)^k}{k!} = e^{M_n} \leq e^{\kappa n} = (e^\kappa)^n. \tag{3.97}$$

Next observe that (ii) in Proposition 3.8 implies for all $n, j \in \mathbb{N}$, $\theta \in \Theta$ that $\sigma_\Omega(Y^\theta_{n,j}) \subseteq \sigma_\Omega\big((Z^{(\theta,\vartheta)})_{\vartheta\in\Theta}\big)$. This demonstrates for all $n, j \in \mathbb{N}$ that $\sigma_\Omega\big(Y^0_{n,j}, Y^1_{n-1,j}\big) \subseteq \sigma_\Omega\big((Z^{(0,\theta)})_{\theta\in\Theta}, (Z^{(1,\theta)})_{\theta\in\Theta}\big)$. The fact that it holds for every $k \in \mathbb{N}_0$ that $\sigma_\Omega\big((Z^{(0,\theta)})_{\theta\in\Theta}, (Z^{(1,\theta)})_{\theta\in\Theta}\big)$, Z^0, and ψ_k are independent hence shows for every $k \in \mathbb{N}_0$, $n, j \in \mathbb{N}$ that

$$\sigma_\Omega\big(Y^0_{n,j}, Y^1_{n-1,j}\big) \quad \text{and} \quad \sigma_\Omega(\psi_k, Z^0) \tag{3.98}$$

are independent. Furthermore, the fact that it holds for all $k \in \mathbb{N}_0$, $x \in \mathcal{Y}$ that $\Omega \ni \omega \mapsto [\psi_k(\omega)](x) \in \mathcal{H}$ is an $\mathscr{F}/\mathscr{B}(\mathcal{H})$-measurable function (cf. Lemma 3.1) and the fact that it holds for all $k \in \mathbb{N}_0$, $\omega \in \Omega$ that $\mathcal{Y} \ni x \mapsto [\psi_k(\omega)](x) \in \mathcal{H}$ is a continuous function yield that

$$\mathcal{Y} \times \mathcal{Y} \times \Omega \ni (u, v, \omega) \mapsto [\psi_k(\omega)](u - v) \in \mathcal{H} \tag{3.99}$$

is a continuous random field. Moreover, note that the fact that $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ is separable ensures that $\mathscr{B}(\mathcal{Y})\otimes\mathscr{B}(\mathcal{Y}) = \mathscr{B}(\mathcal{Y}\times\mathcal{Y})$. This, (i) in Corollary 3.3, (3.98), [181, Lemma 2.2] (with $\mathcal{G} \leftarrow \sigma_\Omega(\psi_k, Z^0)$, $(S, \mathcal{S}) \leftarrow (\mathcal{Y} \times \mathcal{Y}, \mathscr{B}(\mathcal{Y}) \otimes \mathscr{B}(\mathcal{Y}))$, $U \leftarrow (\mathcal{Y} \times \mathcal{Y} \times \Omega \ni (u, v, \omega) \mapsto \|[\psi_k(\omega)](\Phi_n(u, v, Z^0(\omega)))\|^2_{\mathcal{H}} \in [0, \infty))$, $Y \leftarrow (Y^0_{n,j}, Y^1_{n-1,j})$ for $j, n \in \mathbb{N}$, $k \in \mathbb{N}_0$ in the notation of [181, Lemma 2.2]), (3.93), (3.99), and [181, Lemma 2.3] (with $S \leftarrow \mathcal{Y} \times \mathcal{Y}$, $U \leftarrow (\mathcal{Y}\times\mathcal{Y}\times\Omega \ni (u, v, \omega) \mapsto \|[\psi_k(\omega)](u-v)\|^2_{\mathcal{H}} \in [0, \infty))$, $Y \leftarrow (Y^0_{n,j}, Y^1_{n-1,j})$ for $j, n \in \mathbb{N}$, $k \in \mathbb{N}_0$ in the notation of [181, Lemma 2.3]) establish for all $k \in \mathbb{N}_0$, $n, j \in \mathbb{N}$ that

$$\begin{aligned}
&\mathbb{E}\Big[\big\|\psi_k\big(\Phi_n\big(Y^0_{n,j}, Y^1_{n-1,j}, Z^0\big)\big)\big\|^2_{\mathcal{H}}\Big] \\
&= \int_{\mathcal{Y}\times\mathcal{Y}} \mathbb{E}\big[\|\psi_k(\Phi_n(u, v, Z^0))\|^2_{\mathcal{H}}\big] \,\big(\big(Y^0_{n,j}, Y^1_{n-1,j}\big)(\mathbb{P})_{\mathscr{B}(\mathcal{Y})\otimes\mathscr{B}(\mathcal{Y})}\big)(\mathrm{d}u, \mathrm{d}v) \\
&= \int_{\mathcal{Y}\times\mathcal{Y}} \mathbb{E}\big[\|\psi_k(\Phi_n(u, v, Z^0))\|^2_{\mathcal{H}}\big] \,\big(\big(Y^0_{n,j}, Y^1_{n-1,j}\big)(\mathbb{P})_{\mathscr{B}(\mathcal{Y}\times\mathcal{Y})}\big)(\mathrm{d}u, \mathrm{d}v) \\
&\leq c \int_{\mathcal{Y}\times\mathcal{Y}} \mathbb{E}\big[\|\psi_{k+1}(u - v)\|^2_{\mathcal{H}}\big] \,\big(\big(Y^0_{n,j}, Y^1_{n-1,j}\big)(\mathbb{P})_{\mathscr{B}(\mathcal{Y}\times\mathcal{Y})}\big)(\mathrm{d}u, \mathrm{d}v) \\
&= c\,\mathbb{E}\Big[\big\|\psi_{k+1}\big(Y^0_{n,j} - Y^1_{n-1,j}\big)\big\|^2_{\mathcal{H}}\Big].
\end{aligned} \tag{3.100}$$

Combining (3.97) and (3.100) with Theorem 3.14 shows (i)–(iii). The proof of Corollary 3.15 is thus complete. □

Lemma 3.16. *Let $\kappa \in [1, \infty)$, $(M_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ satisfy for all $j \in \mathbb{N}$ that $M_j < M_{j+1}$ and $M_j \leq \kappa j$. Then*

(i) it holds for all $j \in \mathbb{N}$ that $j \leq M_j \leq \kappa j$,

(ii) it holds that $\liminf_{j\to\infty} M_j = \infty$, and

(iii) it holds that $\sup_{j\in\mathbb{N}} {}^{M_{j+1}}/_{M_j} \leq 2\kappa$.

Proof of Lemma 3.16. Note that the assumption that $\forall\, j \in \mathbb{N}\colon M_j < M_{j+1}$ and induction show (i). Next observe that (i) implies (ii). Furthermore, the assumption that $\forall\, j \in \mathbb{N}\colon M_j \leq \kappa j$ and (i) ensure for all $j \in \mathbb{N}$ that

$$\frac{M_{j+1}}{M_j} \leq \frac{\kappa(j + 1)}{j} = \kappa + \frac{\kappa}{j} \leq 2\kappa. \tag{3.101}$$

The proof of Lemma 3.16 is thus complete. □

3.2 MLP for semi-linear heat equations

In this section we employ the abstract framework for generalised MLP approximations developed in Section 3.1 to prove that appropriate MLP approximations, which essentially are generalised versions of the MLP approximations proposed in Hutzenthaler et al. [181], are able to overcome the curse of dimensionality in the numerical approximation of semi-linear heat equations (cf. Theorem 3.33 and Corollary 3.34 in Subsection 3.2.3.2).

In the context of applying the abstract complexity result about generalised MLP approximations in Corollary 3.15 above to numerical approximations for semi-linear heat equations, the separable $\mathbb{R}$-Banach space $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ in Corollary 3.15 is chosen to be a subspace of the vector space of real-valued at most polynomially growing continuous functions defined on $[0, T] \times \mathbb{R}^d$ equipped with a suitable polynomial growth norm, where $T \in (0, \infty)$, $d \in \mathbb{N}$ (cf. (3.129)–(3.130) in Subsection 3.2.2.1). In Subsection 3.2.1 we derive several elementary and well-known properties of these and related function spaces and their elements. In particular, Subsection 3.2.1.1 deals with completeness and separability of such function spaces. Lemma 3.17 recalls that the vector space of real-valued at most polynomially growing continuous functions defined on a non-empty subset of $\mathbb{R}^d$ equipped with an appropriate polynomial growth norm is complete. Thereafter, we state in Proposition 3.18 the well-known fact that the vector space of real-valued continuous functions defined on a non-empty compact subset of $\mathbb{R}^d$ equipped with the uniform norm is a separable $\mathbb{R}$-Banach space, which follows directly from Lemma 3.17 and, e.g., Conway [84, Theorem 6.6 in Chapter V]. Using Proposition 3.18 we deduce the elementary fact that also the vector space of real-valued continuous functions with compact support defined on a non-empty closed subset of $\mathbb{R}^d$ equipped with a suitable polynomial growth norm is separable (cf. Lemma 3.19). Subsection 3.2.1.1 is concluded by the well-known result in Proposition 3.20, which establishes a characterisation of the above mentioned choice for the vector space $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ (cf. (3.129)–(3.130) in Subsection 3.2.2.1) and shows that it is indeed a separable $\mathbb{R}$-Banach space. Subsequently, we provide in Lemmas 3.21 and 3.22 and Corollary 3.23 in Subsection 3.2.1.2 three elementary results about sufficient conditions under which suitable functions and suitable compositions of functions grow strictly slower than a given polynomial order. These results are used to ensure well-definedness of certain functions introduced in (3.133) in Subsection 3.2.2.1. Furthermore, Lemma 3.24 in Subsection 3.2.1.3 offers an elementary polynomial growth estimate for suitable compositions of functions.

In Subsection 3.2.2 we specify a number of the objects appearing in Corollary 3.15 above for the example of MLP approximations for semi-linear heat equations and verify that the main assumptions of Corollary 3.15 are fulfilled in this context. In particular, we first present in Setting 3.1 in Subsection 3.2.2.1 the framework which we refer to throughout Subsection 3.2.2. In the subsection that follows, Subsection 3.2.2.2, we establish measurability properties of several of the involved functions (cf. Lemmas 3.25 and 3.26). Subsequently, Lemma 3.27 in Subsection 3.2.2.3 shows that the MLP approximations introduced in (3.128) in Setting 3.1 fit into the abstract framework for generalised MLP approximations developed in Section 3.1 (cf. (3.90) in Corollary 3.15). Moreover, Subsection 3.2.2.4 is devoted to proving certain integrability properties of the MLP approximations introduced in (3.128) in Setting 3.1 (cf. Lemma 3.28), while in Subsection 3.2.2.5 we verify that the estimates assumed in (3.92)–(3.94) in Corollary 3.15 hold true for the functions introduced in Setting 3.1 (cf. Lemmas 3.29, 3.30, and 3.31).

Finally, in Subsection 3.2.3 we combine the results from Subsection 3.2.2 with Corollary 3.15 to obtain a complexity analysis for MLP approximations for semi-linear heat equations. In Proposition 3.32 in Subsection 3.2.3.1 this is done for semi-linear heat equations of fixed space dimension $d \in \mathbb{N}$ (cf. [181, Theorem 3.8]). Thereafter, Proposition 3.32 is used to establish Theorem 3.33 in Subsection 3.2.3.2, which reveals that the MLP approximations in (3.220) overcome the curse of dimensionality in the numerical approximation of semi-linear heat equations and which essentially is a slight generalisation of [181, Theorem 1.1]. The last result in this section, Corollary 3.34, is a direct consequence of Theorem 3.33 and describes the special case of Theorem 3.33 in which the non-linearity in the semi-linear heat equations is the same for every dimension (cf. (i) in Corollary 3.34) and in which the constants in the complexity estimate are not given explicitly (cf. (ii) in Corollary 3.34).

3.2.1 Properties of spaces of at most polynomially growing continuous functions

3.2.1.1 Completeness and separability

Lemma 3.17. *Let $d \in \mathbb{N}$, $p \in [0,\infty)$, let $\mathcal{A} \subseteq \mathbb{R}^d$ be a non-empty set, let $\mathcal{V} = \big\{v \in C(\mathcal{A},\mathbb{R})\colon \sup_{x\in\mathcal{A}} {}^{|v(x)|}\!/\!_{\max\{1,\|x\|^p_{\mathbb{R}^d}\}} < \infty\big\}$, and let $\|\cdot\|_{\mathcal{V}}\colon \mathcal{V} \to [0,\infty)$ satisfy for all $v \in \mathcal{V}$ that $\|v\|_{\mathcal{V}} = \sup_{x\in\mathcal{A}} {}^{|v(x)|}\!/\!_{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}$. Then it holds that $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ is an $\mathbb{R}$-Banach space.*

Proof of Lemma 3.17. Observe that it holds that $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ is a normed $\mathbb{R}$-vector space. It thus remains to prove that $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ is complete. For this let $\mathcal{W} \subseteq C(\mathcal{A},\mathbb{R})$ be the set given by

$$\mathcal{W} = \big\{w \in C(\mathcal{A},\mathbb{R})\colon \sup_{x\in\mathcal{A}} |w(x)| < \infty\big\}, \tag{3.102}$$

let $\|\cdot\|_{\mathcal{W}}\colon \mathcal{W} \to [0,\infty)$ satisfy for all $w \in \mathcal{W}$ that $\|w\|_{\mathcal{W}} = \sup_{x\in\mathcal{A}} |w(x)|$, and let $I\colon \mathcal{V} \to \mathcal{W}$ and $J\colon \mathcal{W} \to \mathcal{V}$ satisfy for all $v \in \mathcal{V}$, $w \in \mathcal{W}$, $x \in \mathcal{A}$ that $[I(v)](x) = {}^{v(x)}\!/\!_{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}$ and $[J(w)](x) = w(x)\max\{1, \|x\|^p_{\mathbb{R}^d}\}$. Note that $(\mathcal{W}, \|\cdot\|_{\mathcal{W}})$ is a normed $\mathbb{R}$-vector space. Furthermore, Jentzen, Mazzonetto, & Salimova [192, Corollary 2.3] shows that

$$(\mathcal{W}, \|\cdot\|_{\mathcal{W}}) \tag{3.103}$$

is complete. Next observe that it holds for all $v \in \mathcal{V}$ that

$$\|I(v)\|_{\mathcal{W}} = \sup_{x\in\mathcal{A}} |[I(v)](x)| = \sup_{x\in\mathcal{A}} \left[\frac{|v(x)|}{\max\{1, \|x\|^p_{\mathbb{R}^d}\}}\right] = \|v\|_{\mathcal{V}}. \tag{3.104}$$

In addition, note that it holds for all $w \in \mathcal{W}$, $x \in \mathcal{A}$ that

$$[I(J(w))](x) = \frac{[J(w)](x)}{\max\{1, \|x\|^p_{\mathbb{R}^d}\}} = \frac{w(x)\max\{1, \|x\|^p_{\mathbb{R}^d}\}}{\max\{1, \|x\|^p_{\mathbb{R}^d}\}} = w(x). \tag{3.105}$$

Combining this with (3.104) ensures that $I\colon \mathcal{V} \to \mathcal{W}$ is a bijective linear isometry and $I^{-1} = J$. This and (3.103) establish that $(\mathcal{V}, \|\cdot\|_{\mathcal{V}}) = (I^{-1}(\mathcal{W}), \|\cdot\|_{\mathcal{V}})$ is complete and thus finish the proof of Lemma 3.17. ☐

Proposition 3.18. *Let $d \in \mathbb{N}$, let $\mathcal{A} \subseteq \mathbb{R}^d$ be a non-empty compact set, and let $\|\cdot\|_{C(\mathcal{A},\mathbb{R})}\colon C(\mathcal{A},\mathbb{R}) \to [0,\infty)$ satisfy for all $f \in C(\mathcal{A},\mathbb{R})$ that $\|f\|_{C(\mathcal{A},\mathbb{R})} = \sup_{x\in\mathcal{A}} |f(x)|$. Then it holds that $(C(\mathcal{A},\mathbb{R}), \|\cdot\|_{C(\mathcal{A},\mathbb{R})})$ is a separable $\mathbb{R}$-Banach space.*

Lemma 3.19. *Let $d \in \mathbb{N}$, $p \in [0,\infty)$, let $\mathcal{A} \subseteq \mathbb{R}^d$ be a non-empty closed set, and let $|||\cdot|||\colon C_c(\mathcal{A},\mathbb{R}) \to [0,\infty)$ satisfy for all $f \in C_c(\mathcal{A},\mathbb{R})$ that $|||f||| = \sup_{x\in\mathcal{A}} {}^{|f(x)|}/_{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}$. Then it holds that $(C_c(\mathcal{A},\mathbb{R}), |||\cdot|||)$ is a separable normed $\mathbb{R}$-vector space.*

Proof of Lemma 3.19. Throughout this proof let $\mathbf{y} \in \mathcal{A}$, let $N = \min([\|\mathbf{y}\|_{\mathbb{R}^d},\infty)\cap\mathbb{N})$, let $\mathcal{S}_n \subseteq C_c(\mathcal{A},\mathbb{R})$, $n \in \{N, N+1,\ldots\}$, be the sets which satisfy for all $n \in \{N,N+1,\ldots\}$ that

$$\mathcal{S}_n = \big\{f \in C(\mathcal{A},\mathbb{R})\colon \{x \in \mathcal{A}\colon f(x) \neq 0\} \subseteq [-n,n]^d\big\}, \tag{3.106}$$

let $[\![\cdot]\!]_n\colon C(\mathcal{A}\cap[-n,n]^d,\mathbb{R}) \to [0,\infty)$, $n \in \{N,N+1,\ldots\}$, satisfy for all $n \in \{N,N+1,\ldots\}$, $f \in C(\mathcal{A}\cap[-n,n]^d,\mathbb{R})$ that

$$[\![f]\!]_n = \sup_{x\in\mathcal{A}\cap[-n,n]^d}\left[\frac{|f(x)|}{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}\right], \tag{3.107}$$

and let $I_n\colon \mathcal{S}_n \to C(\mathcal{A}\cap[-n,n]^d,\mathbb{R})$, $n \in \{N,N+1,\ldots\}$, satisfy for all $n \in \{N,N+1,\ldots\}$, $f \in \mathcal{S}_n$ that $I_n(f) = f|_{\mathcal{A}\cap[-n,n]^d}$. Note that (3.106) proves for all $n \in \{N,N+1,\ldots\}$, $f \in \mathcal{S}_n$ that

$$[\![I_n(f)]\!]_n = \sup_{x\in\mathcal{A}\cap[-n,n]^d}\left[\frac{|f(x)|}{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}\right] = \sup_{x\in\mathcal{A}}\left[\frac{|f(x)|}{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}\right] = |||f|||. \tag{3.108}$$

This and the fact that it holds for all $n \in \{N,N+1,\ldots\}$ that $(\mathcal{S}_n, |||\cdot|||\,|_{\mathcal{S}_n})$ and $(C(\mathcal{A}\cap[-n,n]^d,\mathbb{R}), [\![\cdot]\!]_n)$ are normed $\mathbb{R}$-vector spaces ensure for all $n \in \{N,N+1,\ldots\}$ that

$$I_n\colon \mathcal{S}_n \to C(\mathcal{A}\cap[-n,n]^d,\mathbb{R}) \tag{3.109}$$

is a linear isometry. Next observe that it holds for all $n \in \{N,N+1,\ldots\}$, $f \in C(\mathcal{A}\cap[-n,n]^d,\mathbb{R})$ that

$$[\![f]\!]_n \leq \sup_{x\in\mathcal{A}\cap[-n,n]^d}|f(x)| \leq \sup_{x\in\mathcal{A}\cap[-n,n]^d}\left[\frac{|f(x)|\big(n\sqrt{d}\big)^p}{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}\right] = [\![f]\!]_n\big(n\sqrt{d}\big)^p. \tag{3.110}$$

In addition, the assumption that $\mathcal{A} \subseteq \mathbb{R}^d$ is a closed set and the fact that $\mathbf{y} \in \mathcal{A}$ ensure for all $n \in \{N,N+1,\ldots\}$ that $\mathcal{A}\cap[-n,n]^d$ is a non-empty compact set. Proposition 3.18 and (3.110) hence show for all $n \in \{N,N+1,\ldots\}$ that $(C(\mathcal{A}\cap[-n,n]^d,\mathbb{R}), [\![\cdot]\!]_n)$ is a separable $\mathbb{R}$-Banach space. This implies for all $n \in \{N,N+1,\ldots\}$ that $(I_n(\mathcal{S}_n), [\![\cdot]\!]_n|_{I_n(\mathcal{S}_n)})$ is a separable normed $\mathbb{R}$-vector space. Combining this with (3.109) hence establishes for all $n \in \{N,N+1,\ldots\}$ that

$$(\mathcal{S}_n, |||\cdot|||\,|_{\mathcal{S}_n}) \tag{3.111}$$

is a separable normed $\mathbb{R}$-vector space. Furthermore, the assumption that $\mathcal{A} \subseteq \mathbb{R}^d$ is a closed set and (3.106) demonstrate that

$$\begin{aligned}
C_c(\mathcal{A},\mathbb{R}) &= \Big\{f \in C(\mathcal{A},\mathbb{R})\colon \Big(\exists\, n \in \mathbb{N}\colon \overline{\{x \in \mathcal{A}\colon f(x) \neq 0\}}^{\mathbb{R}^d} \subseteq \mathcal{A}\cap[-n,n]^d\Big)\Big\} \\
&= \big\{f \in C(\mathcal{A},\mathbb{R})\colon \big(\exists\, n \in \mathbb{N}\colon \{x \in \mathcal{A}\colon f(x) \neq 0\} \subseteq [-n,n]^d\big)\big\} \\
&= \big\{f \in C(\mathcal{A},\mathbb{R})\colon \big(\exists\, n \in \{N,N+1,\ldots\}\colon \{x \in \mathcal{A}\colon f(x) \neq 0\} \subseteq [-n,n]^d\big)\big\} \\
&= \bigcup_{n=N}^{\infty}\big\{f \in C(\mathcal{A},\mathbb{R})\colon \{x \in \mathcal{A}\colon f(x) \neq 0\} \subseteq [-n,n]^d\big\} = \bigcup_{n=N}^{\infty}\mathcal{S}_n.
\end{aligned} \tag{3.112}$$

This and (3.111) establish that $(C_c(\mathcal{A},\mathbb{R}), |||\cdot|||)$ is a separable normed $\mathbb{R}$-vector space. The proof of Lemma 3.19 is thus complete. $\square$

Proposition 3.20. *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, $p \in [0,\infty)$, let $\mathcal{Y} = \big\{ y \in C([0,T]\times\mathbb{R}^d,\mathbb{R})\colon \limsup_{\mathbb{N}\ni n\to\infty} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} {}^{|y(t,x)|}/_{\|x\|^p_{\mathbb{R}^d}} = 0 \big\}$, and let $\|\cdot\|_{\mathcal{Y}}\colon \mathcal{Y}\to[0,\infty)$ satisfy for all $y\in\mathcal{Y}$ that $\|y\|_{\mathcal{Y}} = \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} {}^{|y(t,x)|}/_{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}$. Then*

(i) it holds that $\mathcal{Y} = \overline{C_{\mathrm{c}}([0,T]\times\mathbb{R}^d,\mathbb{R})}^{\mathcal{Y}}$ and

(ii) it holds that $(\mathcal{Y},\|\cdot\|_{\mathcal{Y}})$ is a separable $\mathbb{R}$-Banach space.

Proof of Proposition 3.20. Throughout this proof let $\tau_n \in C(\mathbb{R}^d,[0,1])$, $n\in\mathbb{N}$, satisfy for all $n\in\mathbb{N}$, $x\in\mathbb{R}^d$ that

$$\tau_n(x) = \max\{\min\{n+1-\|x\|_{\mathbb{R}^d},1\},0\} = \begin{cases} 1 & \colon \|x\|_{\mathbb{R}^d}\leq n \\ n+1-\|x\|_{\mathbb{R}^d} & \colon n\leq\|x\|_{\mathbb{R}^d}\leq n+1 \\ 0 & \colon n+1\leq\|x\|_{\mathbb{R}^d} \end{cases}, \tag{3.113}$$

let $\mathbf{y}\in\mathcal{Y}$, and let $y_n\in C([0,T]\times\mathbb{R}^d,\mathbb{R})$, $n\in\mathbb{N}$, satisfy for all $n\in\mathbb{N}$, $t\in[0,T]$, $x\in\mathbb{R}^d$ that $y_n(t,x)=\tau_n(x)\,\mathbf{y}(t,x)$. Note that it holds that $(y_n)_{n\in\mathbb{N}}\subseteq C_{\mathrm{c}}([0,T]\times\mathbb{R}^d,\mathbb{R})\subseteq\mathcal{Y}$ and

$$\begin{aligned} \limsup_{\mathbb{N}\ni n\to\infty}\|\mathbf{y}-y_n\|_{\mathcal{Y}} &= \limsup_{\mathbb{N}\ni n\to\infty}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|\mathbf{y}(t,x)-\tau_n(x)\,\mathbf{y}(t,x)|}{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}\right] \\ &= \limsup_{\mathbb{N}\ni n\to\infty}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\,\|x\|_{\mathbb{R}^d}\geq n}\left[\frac{(1-\tau_n(x))|\mathbf{y}(t,x)|}{\|x\|^p_{\mathbb{R}^d}}\right] \\ &\leq \limsup_{\mathbb{N}\ni n\to\infty}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\,\|x\|_{\mathbb{R}^d}\geq n}\frac{|\mathbf{y}(t,x)|}{\|x\|^p_{\mathbb{R}^d}} = 0. \end{aligned} \tag{3.114}$$

This proves that $\mathcal{Y}\subseteq\overline{C_{\mathrm{c}}([0,T]\times\mathbb{R}^d,\mathbb{R})}^{\mathcal{Y}}$. In addition, observe that the fact that $\mathcal{Y}\supseteq C_{\mathrm{c}}([0,T]\times\mathbb{R}^d,\mathbb{R})$ ensures that $\mathcal{Y}\supseteq\overline{C_{\mathrm{c}}([0,T]\times\mathbb{R}^d,\mathbb{R})}^{\mathcal{Y}}$. This finishes the proof of (i). It thus remains to show (ii). For this let $\mathcal{V}\subseteq C([0,T]\times\mathbb{R}^d,\mathbb{R})$ be the set given by

$$\mathcal{V} = \left\{ v\in C([0,T]\times\mathbb{R}^d,\mathbb{R})\colon \sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\tfrac{|v(t,x)|}{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}\right]<\infty\right\}, \tag{3.115}$$

let $\|\cdot\|_{\mathcal{V}}\colon\mathcal{V}\to[0,\infty)$ satisfy for all $v\in\mathcal{V}$ that

$$\|v\|_{\mathcal{V}} = \sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|v(t,x)|}{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}\right], \tag{3.116}$$

let $\mathbf{v}\in\mathcal{V}$, and let $(v_n)_{n\in\mathbb{N}}\subseteq\mathcal{Y}\subseteq\mathcal{V}$ be a sequence which satisfies $\limsup_{\mathbb{N}\ni n\to\infty}\|\mathbf{v}-v_n\|_{\mathcal{V}} = 0$. Note that this implies that

$$\begin{aligned} &\limsup_{\mathbb{N}\ni n\to\infty}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\,\|x\|_{\mathbb{R}^d}\geq n}\frac{|\mathbf{v}(t,x)|}{\|x\|^p_{\mathbb{R}^d}} \\ &\leq \limsup_{\mathbb{N}\ni m\to\infty}\limsup_{\mathbb{N}\ni n\to\infty}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\,\|x\|_{\mathbb{R}^d}\geq n}\frac{|\mathbf{v}(t,x)-v_m(t,x)|}{\|x\|^p_{\mathbb{R}^d}} \\ &\quad+\limsup_{\mathbb{N}\ni m\to\infty}\limsup_{\mathbb{N}\ni n\to\infty}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\,\|x\|_{\mathbb{R}^d}\geq n}\frac{|v_m(t,x)|}{\|x\|^p_{\mathbb{R}^d}} \\ &\leq \limsup_{\mathbb{N}\ni m\to\infty}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|\mathbf{v}(t,x)-v_m(t,x)|}{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}\right] = \limsup_{\mathbb{N}\ni m\to\infty}\|\mathbf{v}-v_m\|_{\mathcal{V}} = 0. \end{aligned} \tag{3.117}$$

This establishes that $\mathbf{v} \in \mathcal{Y}$. Therefore, it holds that $\mathcal{Y} \subseteq \mathcal{V}$ is a closed set. The fact that $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ is an $\mathbb{R}$-Banach space (cf. Lemma 3.17) hence demonstrates that $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}}) = (\mathcal{Y}, \|\cdot\|_{\mathcal{V}}|_{\mathcal{Y}})$ is an $\mathbb{R}$-Banach space. Moreover, note that the fact that $(C_c([0,T] \times \mathbb{R}^d, \mathbb{R}), \|\cdot\|_{\mathcal{Y}}|_{C_c([0,T]\times\mathbb{R}^d,\mathbb{R})})$ is a separable normed $\mathbb{R}$-vector space (cf. Lemma 3.19) and (i) assure that

$$(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}}) = \left(\overline{C_c([0,T] \times \mathbb{R}^d, \mathbb{R})}^{\mathcal{Y}}, \|\cdot\|_{\mathcal{Y}}|_{\overline{C_c([0,T]\times\mathbb{R}^d,\mathbb{R})}^{\mathcal{Y}}}\right) \tag{3.118}$$

is separable. This establishes (ii). The proof of Proposition 3.20 is thus complete. □

3.2.1.2 Sufficient conditions for strictly slower growth

Lemma 3.21. *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, $p \in [0,\infty)$, $q \in (p,\infty)$ and let $y \in C([0,T]\times\mathbb{R}^d, \mathbb{R})$ satisfy $\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} {}^{|y(t,x)|}/_{\max\{1,\|x\|_{\mathbb{R}^d}^p\}} < \infty$. Then*

$$\limsup_{\mathbb{N}\ni n\to\infty} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} \frac{|y(t,x)|}{\|x\|_{\mathbb{R}^d}^q} = 0. \tag{3.119}$$

Proof of Lemma 3.21. Throughout this proof let $C \in [0,\infty)$ be the real number which satisfies $C = \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} {}^{|y(t,x)|}/_{\max\{1,\|x\|_{\mathbb{R}^d}^p\}}$. Observe that it holds that

$$\begin{aligned}
&\limsup_{\mathbb{N}\ni n\to\infty} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} \frac{|y(t,x)|}{\|x\|_{\mathbb{R}^d}^q} \\
&= \limsup_{\mathbb{N}\ni n\to\infty} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} \left[\frac{|y(t,x)|}{\max\{1,\|x\|_{\mathbb{R}^d}^p\}} \frac{\max\{1,\|x\|_{\mathbb{R}^d}^p\}}{\|x\|_{\mathbb{R}^d}^q}\right] \\
&\leq C \limsup_{\mathbb{N}\ni n\to\infty} \sup_{x\in\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} \left[\frac{\max\{1,\|x\|_{\mathbb{R}^d}^p\}}{\|x\|_{\mathbb{R}^d}^q}\right] \\
&= C \limsup_{\mathbb{N}\ni n\to\infty} \sup_{x\in\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} \frac{1}{\|x\|_{\mathbb{R}^d}^{q-p}} = C \limsup_{\mathbb{N}\ni n\to\infty} \frac{1}{n^{q-p}} = 0.
\end{aligned} \tag{3.120}$$

The proof of Lemma 3.21 is thus complete. □

Lemma 3.22. *Let $d \in \mathbb{N}$, $T, q \in (0,\infty)$, let $\mathcal{Y} = \{y \in C([0,T]\times\mathbb{R}^d, \mathbb{R})\colon \limsup_{\mathbb{N}\ni n\to\infty} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} {}^{|y(t,x)|}/_{\|x\|_{\mathbb{R}^d}^q} = 0\}$, let $\varrho = (\varrho_1, \varrho_2) \in C([0,T]\times\mathbb{R}^d, [0,T]\times\mathbb{R}^d)$ satisfy $\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} {}^{\|\varrho_2(t,x)\|_{\mathbb{R}^d}}/_{\max\{1,\|x\|_{\mathbb{R}^d}\}} < \infty$, and let $y \in \mathcal{Y}$. Then it holds that $y \circ \varrho \in \mathcal{Y}$.*

Proof of Lemma 3.22. Throughout this proof let $\varepsilon \in (0,\infty)$ and let $L, \mathfrak{N}, N \in \mathbb{N}$ satisfy $\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} {}^{\|\varrho_2(t,x)\|_{\mathbb{R}^d}}/_{\max\{1,\|x\|_{\mathbb{R}^d}\}} \leq L$, $\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq\mathfrak{N}} {}^{|y(t,x)|}/_{\|x\|_{\mathbb{R}^d}^q} \leq \frac{\varepsilon}{L^q}$, and $\varepsilon^{-1/q} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\leq\mathfrak{N}} |y(t,x)|^{1/q} \leq N$. Observe that it holds for all $t \in [0,T]$, $x \in \mathbb{R}^d$ with $\|x\|_{\mathbb{R}^d} \geq N$ and $\|\varrho_2(t,x)\|_{\mathbb{R}^d} \leq \mathfrak{N}$ that

$$|y(\varrho(t,x))| = |y(\varrho_1(t,x), \varrho_2(t,x))| \leq \sup_{(s,\mathbf{x})\in[0,T]\times\mathbb{R}^d,\, \|\mathbf{x}\|_{\mathbb{R}^d}\leq\mathfrak{N}} |y(s,\mathbf{x})| \leq \varepsilon N^q \leq \varepsilon \|x\|_{\mathbb{R}^d}^q. \tag{3.121}$$

In addition, note that it holds for all $t \in [0,T]$, $x \in \mathbb{R}^d$ with $\|x\|_{\mathbb{R}^d} \geq 1$ and $\|\varrho_2(t,x)\|_{\mathbb{R}^d} \geq \mathfrak{N}$ that

$$\begin{aligned}
|y(\varrho(t,x))| &\leq \left[\sup_{(s,\mathbf{x})\in[0,T]\times\mathbb{R}^d,\, \|\mathbf{x}\|_{\mathbb{R}^d}\geq\mathfrak{N}} \frac{|y(s,\mathbf{x})|}{\|\mathbf{x}\|_{\mathbb{R}^d}^q}\right] \|\varrho_2(t,x)\|_{\mathbb{R}^d}^q \\
&\leq \frac{\varepsilon}{L^q} \cdot L^q \max\{1, \|x\|_{\mathbb{R}^d}^q\} = \varepsilon \|x\|_{\mathbb{R}^d}^q.
\end{aligned} \tag{3.122}$$

This and (3.121) establish for all $n \in \{N, N+1, \ldots\}$ that

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} \frac{|y(\varrho(t,x))|}{\|x\|_{\mathbb{R}^d}^q} \leq \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq N} \frac{|y(\varrho(t,x))|}{\|x\|_{\mathbb{R}^d}^q} \leq \varepsilon. \tag{3.123}$$

The fact that $y \circ \varrho \in C([0,T]\times\mathbb{R}^d, \mathbb{R})$ thus completes the proof of Lemma 3.22. □

Corollary 3.23. *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, $L, p \in [0,\infty)$, $q \in (p,\infty)$, let $\mathcal{Y} = \{y \in C([0,T]\times\mathbb{R}^d, \mathbb{R})\colon \limsup_{\mathbb{N}\ni n\to\infty} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} {}^{|y(t,x)|}/{}_{\|x\|_{\mathbb{R}^d}^q} = 0\}$, and let $\varrho = (\varrho_1, \varrho_2) \in C([0,T], [0,T]\times\mathbb{R}^d)$, $f \in C([0,T]\times\mathbb{R}^d\times\mathbb{R}, \mathbb{R})$, $y \in \mathcal{Y}$ satisfy for all $t \in [0,T]$, $x \in \mathbb{R}^d$, $v, w \in \mathbb{R}$ that $|f(t,x,0)| \leq L\max\{1, \|x\|_{\mathbb{R}^d}^p\}$ and $|f(t,x,v) - f(t,x,w)| \leq L|v-w|$. Then*

(i) it holds that $([0,T]\times\mathbb{R}^d \ni (t,x) \mapsto y(\varrho_1(t), x+\varrho_2(t)) \in \mathbb{R}) \in \mathcal{Y}$ and

(ii) it holds that $([0,T]\times\mathbb{R}^d \ni (t,x) \mapsto f(\varrho_1(t), x+\varrho_2(t), y(t,x)) \in \mathbb{R}) \in \mathcal{Y}$.

Proof of Corollary 3.23. Note that it holds that

$$\begin{aligned}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \left[\frac{\|x+\varrho_2(t)\|_{\mathbb{R}^d}}{\max\{1, \|x\|_{\mathbb{R}^d}\}}\right] &\leq \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \left[\frac{\|x\|_{\mathbb{R}^d}}{\max\{1, \|x\|_{\mathbb{R}^d}\}} + \frac{\|\varrho_2(t)\|_{\mathbb{R}^d}}{\max\{1, \|x\|_{\mathbb{R}^d}\}}\right] \\ &\leq 1 + \sup_{t\in[0,T]} \|\varrho_2(t)\|_{\mathbb{R}^d} < \infty.\end{aligned} \tag{3.124}$$

Lemma 3.22 (with $d \leftarrow d$, $T \leftarrow T$, $q \leftarrow q$, $\mathcal{Y} \leftarrow \mathcal{Y}$, $\varrho \leftarrow ([0,T]\times\mathbb{R}^d \ni (t,x) \mapsto (\varrho_1(t), x+\varrho_2(t)) \in [0,T]\times\mathbb{R}^d)$, $y \leftarrow y$ in the notation of Lemma 3.22) hence shows that $([0,T]\times\mathbb{R}^d \ni (t,x) \mapsto y(\varrho_1(t), x+\varrho_2(t)) \in \mathbb{R}) \in \mathcal{Y}$. This proves (i). Next observe that Lemma 3.21 (with $d \leftarrow d$, $T \leftarrow T$, $p \leftarrow p$, $q \leftarrow q$, $y \leftarrow ([0,T]\times\mathbb{R}^d \ni (t,x) \mapsto f(t,x,0) \in \mathbb{R})$ in the notation of Lemma 3.21) ensures that $([0,T]\times\mathbb{R}^d \ni (t,x) \mapsto f(t,x,0) \in \mathbb{R}) \in \mathcal{Y}$. Combining this with (i) implies that $([0,T]\times\mathbb{R}^d \ni (t,x) \mapsto f(\varrho_1(t), x+\varrho_2(t), 0) \in \mathbb{R}) \in \mathcal{Y}$. Therefore, we obtain that

$$\begin{aligned}&\limsup_{\mathbb{N}\ni n\to\infty} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} \left[\frac{|f(\varrho_1(t), x+\varrho_2(t), y(t,x))|}{\|x\|_{\mathbb{R}^d}^q}\right] \\ &\leq \limsup_{\mathbb{N}\ni n\to\infty} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} \left[\frac{|f(\varrho_1(t), x+\varrho_2(t), 0)|}{\|x\|_{\mathbb{R}^d}^q}\right] \\ &\quad + \limsup_{\mathbb{N}\ni n\to\infty} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} \left[\frac{|f(\varrho_1(t), x+\varrho_2(t), y(t,x)) - f(\varrho_1(t), x+\varrho_2(t), 0)|}{\|x\|_{\mathbb{R}^d}^q}\right] \\ &\leq L \limsup_{\mathbb{N}\ni n\to\infty} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} \frac{|y(t,x)|}{\|x\|_{\mathbb{R}^d}^q} = 0.\end{aligned} \tag{3.125}$$

This and the fact that $([0,T]\times\mathbb{R}^d \ni (t,x) \mapsto f(\varrho_1(t), x+\varrho_2(t), y(t,x)) \in \mathbb{R}) \in C([0,T]\times\mathbb{R}^d, \mathbb{R})$ establish (ii). The proof of Corollary 3.23 is thus complete. □

3.2.1.3 Growth estimate for compositions

Lemma 3.24. *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, $p \in [0,\infty)$, $L \in [1,\infty)$, let $[\![\cdot]\!]\colon C([0,T]\times\mathbb{R}^d, \mathbb{R}) \to [0,\infty]$ satisfy for all $v \in C([0,T]\times\mathbb{R}^d, \mathbb{R})$ that $[\![v]\!] = \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} {}^{|v(t,x)|}/{}_{\max\{1, \|x\|_{\mathbb{R}^d}^p\}}$, let*

$\varrho = (\varrho_1, \varrho_2) \in C([0,T]\times\mathbb{R}^d, [0,T]\times\mathbb{R}^d)$ *satisfy for all* $t \in [0,T]$, $x \in \mathbb{R}^d$ *that* $\|\varrho_2(t,x)\|_{\mathbb{R}^d} \leq L\max\{1, \|x\|_{\mathbb{R}^d}\}$, *and let* $v \in C([0,T]\times\mathbb{R}^d, \mathbb{R})$. *Then it holds that* $v \circ \varrho \in C([0,T]\times\mathbb{R}^d, \mathbb{R})$ *and* $[\![v \circ \varrho]\!] \leq L^p [\![v]\!]$.

Proof of Lemma 3.24. Observe that it holds that $v \circ \varrho \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$. In addition, note that it holds that

$$\begin{aligned}
[\![v \circ \varrho]\!] &= \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \left[\frac{|v(\varrho(t,x))|}{\max\{1, \|x\|_{\mathbb{R}^d}^p\}}\right] \leq [\![v]\!] \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \left[\frac{\max\{1, \|\varrho_2(t,x)\|_{\mathbb{R}^d}^p\}}{\max\{1, \|x\|_{\mathbb{R}^d}^p\}}\right] \\
&\leq [\![v]\!] \sup_{x\in\mathbb{R}^d} \left[\frac{\max\{1, L^p\max\{1, \|x\|_{\mathbb{R}^d}^p\}\}}{\max\{1, \|x\|_{\mathbb{R}^d}^p\}}\right] = L^p[\![v]\!].
\end{aligned} \tag{3.126}$$

The proof of Lemma 3.24 is thus complete. □

3.2.2 Verification of the assumed properties

3.2.2.1 Setting

Setting 3.1. *Let* $d \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $T \in (0,\infty)$, $L, p \in [0,\infty)$, $q \in (p,\infty)$, $\Theta = \bigcup_{n=1}^\infty \mathbb{Z}^n$, $(M_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$, *let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space, let* $\mathbf{U}\colon \Omega \to [0,1]$ *and* $U^\theta\colon \Omega \to [0,1]$, $\theta \in \Theta$, *be on* $[0,1]$ *uniformly distributed random variables, let* $\mathbf{W}\colon [0,T]\times\Omega \to \mathbb{R}^d$ *and* $W^\theta\colon [0,T]\times\Omega \to \mathbb{R}^d$, $\theta \in \Theta$, *be standard Brownian motions with continuous sample paths, assume that* (U^θ, W^θ), $\theta \in \Theta$, *are independent, assume that* $\mathbf{U}$, $\mathbf{W}$, $(U^\theta)_{\theta\in\Theta}$, *and* $(W^\theta)_{\theta\in\Theta}$ *are independent, let* $f \in C([0,T]\times\mathbb{R}^d\times\mathbb{R}, \mathbb{R})$, $g \in C(\mathbb{R}^d, \mathbb{R})$, $y \in C([0,T]\times\mathbb{R}^d, \mathbb{R})$ *satisfy for all* $t \in [0,T]$, $x \in \mathbb{R}^d$, $v, w \in \mathbb{R}$ *that* $\max\{|f(t,x,0)|, |g(x)|\} \leq L\max\{1, \|x\|_{\mathbb{R}^d}^p\}$, $|f(t,x,v) - f(t,x,w)| \leq L|v-w|$, $\sup_{(s,\mathbf{x})\in[0,T]\times\mathbb{R}^d} {}^{|y(s,\mathbf{x})|}/{}_{\max\{1,\|\mathbf{x}\|_{\mathbb{R}^d}^p\}} < \infty$, *and*

$$y(t,x) = \mathbb{E}\left[g(x + \mathbf{W}_{T-t}) + \int_t^T f\big(s, x + \mathbf{W}_{s-t}, y(s, x + \mathbf{W}_{s-t})\big)\,\mathrm{d}s\right], \tag{3.127}$$

let $Y_{n,j}^\theta\colon [0,T]\times\mathbb{R}^d\times\Omega \to \mathbb{R}$, $\theta \in \Theta$, $j \in \mathbb{N}$, $n \in (\mathbb{N}_0 \cup \{-1\})$, *satisfy for all* $n, j \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ *that* $Y_{-1,j}^\theta(t,x) = Y_{0,j}^\theta(t,x) = 0$ *and*

$$\begin{aligned}
Y_{n,j}^\theta(T-t,x) &= \tfrac{1}{(M_j)^n}\left[\sum_{i=1}^{(M_j)^n} g\big(x + W_t^{(\theta,0,i)}\big)\right] + \sum_{l=0}^{n-1} \tfrac{t}{(M_j)^{n-l}} \left[\sum_{i=1}^{(M_j)^{n-l}}\right. \\
&\quad \Big[f\Big(T - t + U^{(\theta,l,i)}t, x + W_{U^{(\theta,l,i)}t}^{(\theta,l,i)}, Y_{l,j}^{(\theta,l,i)}\big(T - t + U^{(\theta,l,i)}t, x + W_{U^{(\theta,l,i)}t}^{(\theta,l,i)}\big)\Big) \\
&\quad \left. - \mathbb{1}_{\mathbb{N}}(l) f\Big(T - t + U^{(\theta,l,i)}t, x + W_{U^{(\theta,l,i)}t}^{(\theta,l,i)}, Y_{l-1,j}^{(\theta,-l,i)}\big(T - t + U^{(\theta,l,i)}t, x + W_{U^{(\theta,l,i)}t}^{(\theta,l,i)}\big)\Big)\Big]\right],
\end{aligned} \tag{3.128}$$

let $\mathcal{Y} \subseteq C([0,T]\times\mathbb{R}^d, \mathbb{R})$ *be the set given by*

$$\mathcal{Y} = \left\{v \in C([0,T]\times\mathbb{R}^d, \mathbb{R})\colon \limsup_{\mathbb{N}\ni n\to\infty} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} \frac{|v(t,x)|}{\|x\|_{\mathbb{R}^d}^q} = 0\right\}, \tag{3.129}$$

let $\|\cdot\|_{\mathcal{Y}}\colon \mathcal{Y} \to [0,\infty)$ *satisfy for all* $v \in \mathcal{Y}$ *that*

$$\|v\|_{\mathcal{Y}} = \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \left[\frac{|v(t,x)|}{\max\{1, \|x\|_{\mathbb{R}^d}^q\}}\right], \tag{3.130}$$

let $\mathcal{Z} = [0,1] \times C([0,T],\mathbb{R}^d)$*, let* $\mathbf{d}_{\mathcal{Z}}\colon \mathcal{Z} \times \mathcal{Z} \to [0,\infty)$ *satisfy for all* $\mathfrak{z} = (\mathfrak{u},\mathfrak{w}), \mathfrak{Z} = (\mathfrak{U},\mathfrak{W}) \in \mathcal{Z}$ *that*

$$\mathbf{d}_{\mathcal{Z}}(\mathfrak{z},\mathfrak{Z}) = |\mathfrak{u}-\mathfrak{U}| + \|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)} = |\mathfrak{u}-\mathfrak{U}| + \sup_{t\in[0,T]} \|\mathfrak{w}(t)-\mathfrak{W}(t)\|_{\mathbb{R}^d}, \tag{3.131}$$

let $Z^\theta\colon \Omega \to \mathcal{Z}$*,* $\theta \in \Theta$*, satisfy for all* $\theta \in \Theta$ *that* $Z^\theta = (U^\theta, W^\theta)$*, let* $\psi_k\colon \Omega \to \mathcal{Y}^*$*,* $k \in \mathbb{N}_0$*, satisfy for all* $k \in \mathbb{N}_0$*,* $\omega \in \Omega$*,* $v \in \mathcal{Y}$ *that*

$$[\psi_k(\omega)](v) = \begin{cases} v(0,\xi) & : k = 0 \\ \sqrt{\frac{(\mathbf{U}(\omega))^{k-1}}{(k-1)!}}\, v\big(\mathbf{U}(\omega)T, \xi + \mathbf{W}_{\mathbf{U}(\omega)T}(\omega)\big) & : k \in \mathbb{N} \end{cases}, \tag{3.132}$$

and let $\Phi_l\colon \mathcal{Y} \times \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y}$*,* $l \in \mathbb{N}_0$*, satisfy for all* $l \in \mathbb{N}_0$*,* $v,w \in \mathcal{Y}$*,* $\mathfrak{z} = (\mathfrak{u},\mathfrak{w}) \in \mathcal{Z}$*,* $t \in [0,T]$*,* $x \in \mathbb{R}^d$ *that*

$$\begin{aligned} &[\Phi_l(v,w,\mathfrak{z})](T-t,x) \\ &= \begin{cases} g(x+\mathfrak{w}_t) + tf\big(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t}, v(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t})\big) & : l = 0 \\ t\big[f\big(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t}, v(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t})\big) & \\ \quad - f\big(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t}, w(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t})\big)\big] & : l \in \mathbb{N} \end{cases} \end{aligned} \tag{3.133}$$

(cf. Lemma 3.21 and Corollary 3.23).

3.2.2.2 Measurability

Lemma 3.25. *Assume Setting 3.1 and let* $\mathscr{S} = \sigma_{\mathcal{Y}^*}\big(\{\{\varphi \in \mathcal{Y}^*\colon \varphi(v) \in \mathcal{B}\} \subseteq \mathcal{Y}^*\colon v \in \mathcal{Y}, \mathcal{B} \in \mathscr{B}(\mathbb{R})\}\big)$*. Then it holds for all* $k \in \mathbb{N}_0$ *that* $\psi_k\colon \Omega \to \mathcal{Y}^*$ *is an* $\mathscr{F}/\mathscr{S}$*-measurable function.*

Proof of Lemma 3.25. Note that it holds for all $k \in \mathbb{N}_0$, $v \in \mathcal{Y}$ that $\Omega \ni \omega \mapsto [\psi_k(\omega)](v) \in \mathbb{R}$ is an $\mathscr{F}/\mathscr{B}(\mathbb{R})$-measurable function. Lemma 3.1 (with $\mathcal{E} \leftarrow \mathcal{Y}$, $(\mathcal{F},\mathscr{F}) \leftarrow (\mathbb{R},\mathscr{B}(\mathbb{R}))$, $(\mathcal{G},\mathscr{G}) \leftarrow (\Omega,\mathscr{F})$, $\mathcal{S} \leftarrow \mathcal{Y}^*$, $\mathscr{S} \leftarrow \mathscr{S}$, $\psi \leftarrow \psi_k$ for $k \in \mathbb{N}_0$ in the notation of Lemma 3.1) hence proves for all $k \in \mathbb{N}_0$ that $\psi_k\colon \Omega \to \mathcal{Y}^*$ is an $\mathscr{F}/\mathscr{S}$-measurable function. The proof of Lemma 3.25 is thus complete. ☐

Lemma 3.26. *Assume Setting 3.1. Then*

(i) it holds for all $l \in \mathbb{N}_0$ *that* $\Phi_l\colon \mathcal{Y} \times \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y}$ *is a continuous function and*

(ii) it holds for all $l \in \mathbb{N}_0$ *that* $\Phi_l\colon \mathcal{Y} \times \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y}$ *is a* $(\mathscr{B}(\mathcal{Y}) \otimes \mathscr{B}(\mathcal{Y}) \otimes \mathscr{Z})/\mathscr{B}(\mathcal{Y})$*-measurable function.*

Proof of Lemma 3.26. Throughout this proof let $\varphi_1\colon \mathcal{Z} \to C([0,T],\mathbb{R}^d)$, $\varphi_2, \varphi_3, F\colon \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y}$, $\varphi_4\colon \mathcal{Y} \to \mathcal{Y}$, $\Psi_1, \Psi_2\colon \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y} \times \mathcal{Z}$, $\mathfrak{g} \in \mathcal{Y}$, $G\colon \mathcal{Z} \to \mathcal{Y}$ satisfy for all $v,w \in \mathcal{Y}$, $\mathfrak{z} = (\mathfrak{u},\mathfrak{w}) \in \mathcal{Z}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$[\varphi_1(\mathfrak{z})](t) = \mathfrak{w}_{\mathfrak{u}t}, \qquad \Psi_1(v,\mathfrak{z}) = (v, \mathfrak{u}, \varphi_1(\mathfrak{z})), \tag{3.134}$$

$$[\varphi_2(v,\mathfrak{z})](t,x) = v(T-t+\mathfrak{u}t, x+\mathfrak{w}_t), \qquad \Psi_2(v,\mathfrak{z}) = (\varphi_2(v,\mathfrak{z}),\mathfrak{z}), \tag{3.135}$$

$$[\varphi_3(v,\mathfrak{z})](t,x) = tf\big(T-t+\mathfrak{u}t, x+\mathfrak{w}_t, v(t,x)\big), \qquad \mathfrak{g}(t,x) = g(x), \tag{3.136}$$

$$[\varphi_4(v)](t,x) = v(T-t,x), \qquad G(\mathfrak{z}) = \varphi_4(\varphi_2(\mathfrak{g},\mathfrak{z})), \tag{3.137}$$

and $F = \varphi_4 \circ \varphi_3 \circ \Psi_2 \circ \Psi_1$ (cf. Corollary 3.23). Note that it holds for all $v \in \mathcal{Y}$, $\mathfrak{z} = (\mathfrak{u}, \mathfrak{w}) \in \mathcal{Z}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}[G(\mathfrak{z})](T-t,x) &= [\varphi_4(\varphi_2(\mathfrak{g},\mathfrak{z}))](T-t,x) = [\varphi_2(\mathfrak{g},\mathfrak{z})](t,x)\\ &= \mathfrak{g}(T-t+\mathfrak{u}t, x+\mathfrak{w}_t) = g(x+\mathfrak{w}_t)\end{aligned} \tag{3.138}$$

and

$$\begin{aligned}[F(v,\mathfrak{z})](T-t,x) &= \big[\varphi_4\big((\varphi_3 \circ \Psi_2 \circ \Psi_1)(v,\mathfrak{z})\big)\big](T-t,x)\\ &= [(\varphi_3 \circ \Psi_2 \circ \Psi_1)(v,\mathfrak{z})](t,x) = \big[(\varphi_3 \circ \Psi_2)\big(v,\mathfrak{u},\varphi_1(\mathfrak{z})\big)\big](t,x)\\ &= \big[\varphi_3\big(\varphi_2\big(v,\mathfrak{u},\varphi_1(\mathfrak{z})\big),\mathfrak{u},\varphi_1(\mathfrak{z})\big)\big](t,x)\\ &= tf\big(T-t+\mathfrak{u}t, x+[\varphi_1(\mathfrak{z})](t), \big[\varphi_2\big(v,\mathfrak{u},\varphi_1(\mathfrak{z})\big)\big](t,x)\big)\\ &= tf\big(T-t+\mathfrak{u}t, x+[\varphi_1(\mathfrak{z})](t), v\big(T-t+\mathfrak{u}t, x+[\varphi_1(\mathfrak{z})](t)\big)\big)\\ &= tf\big(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t}, v(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t})\big).\end{aligned} \tag{3.139}$$

Combining (3.138)–(3.139) with (3.133) ensures for all $l \in \mathbb{N}$, $v, w \in \mathcal{Y}$, $\mathfrak{z} \in \mathcal{Z}$ that

$$\Phi_0(v,w,\mathfrak{z}) = G(\mathfrak{z}) + F(v,\mathfrak{z}) \qquad \text{and} \qquad \Phi_l(v,w,\mathfrak{z}) = F(v,\mathfrak{z}) - F(w,\mathfrak{z}). \tag{3.140}$$

In the following we establish that $G\colon \mathcal{Z} \to \mathcal{Y}$ and $F\colon \mathcal{Y}\times\mathcal{Z} \to \mathcal{Y}$ are continuous functions.

First, we show that $\varphi_1\colon \mathcal{Z} \to C([0,T],\mathbb{R}^d)$ is a continuous function. Throughout this paragraph let $\varepsilon \in (0,\infty)$, $\mathfrak{Z} = (\mathfrak{U},\mathfrak{W}) \in \mathcal{Z}$ and let $\Delta, \delta \in (0,\infty)$ be real numbers which satisfy $\sup_{s,t\in[0,T],\, |s-t|\le\Delta} \|\mathfrak{W}_s - \mathfrak{W}_t\|_{\mathbb{R}^d} \le \frac{\varepsilon}{2}$ and $\delta = \min\{\frac{\Delta}{T}, \frac{\varepsilon}{2}\}$. Observe that it holds for all $\mathfrak{z} = (\mathfrak{u},\mathfrak{w}) \in \mathcal{Z}$ with $\mathbf{d}_{\mathcal{Z}}(\mathfrak{z},\mathfrak{Z}) = |\mathfrak{u}-\mathfrak{U}| + \|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)} \le \delta$ that

$$\begin{aligned}\|\varphi_1(\mathfrak{z}) - \varphi_1(\mathfrak{Z})\|_{C([0,T],\mathbb{R}^d)} &= \sup_{t\in[0,T]} \|\mathfrak{w}_{\mathfrak{u}t} - \mathfrak{W}_{\mathfrak{U}t}\|_{\mathbb{R}^d}\\ &\le \left[\sup_{t\in[0,T]} \|\mathfrak{w}_{\mathfrak{u}t} - \mathfrak{W}_{\mathfrak{u}t}\|_{\mathbb{R}^d}\right] + \left[\sup_{t\in[0,T]} \|\mathfrak{W}_{\mathfrak{u}t} - \mathfrak{W}_{\mathfrak{U}t}\|_{\mathbb{R}^d}\right]\\ &\le \left[\sup_{t\in[0,T]} \|\mathfrak{w}_t - \mathfrak{W}_t\|_{\mathbb{R}^d}\right] + \left[\sup_{s,t\in[0,T],\, |s-t|\le T\delta} \|\mathfrak{W}_s - \mathfrak{W}_t\|_{\mathbb{R}^d}\right]\\ &\le \|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)} + \left[\sup_{s,t\in[0,T],\, |s-t|\le\Delta} \|\mathfrak{W}_s - \mathfrak{W}_t\|_{\mathbb{R}^d}\right]\\ &\le \delta + \tfrac{\varepsilon}{2} \le \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} = \varepsilon.\end{aligned} \tag{3.141}$$

It thus holds that $\varphi_1\colon \mathcal{Z} \to C([0,T],\mathbb{R}^d)$ is a continuous function. Note that this ensures that $\Psi_1\colon \mathcal{Y}\times\mathcal{Z} \to \mathcal{Y}\times\mathcal{Z}$ is a continuous function.

Second, we claim that $\varphi_2\colon \mathcal{Y}\times\mathcal{Z} \to \mathcal{Y}$ is a continuous function. Throughout this paragraph let $\varepsilon \in (0,\infty)$, $\mathbf{v} \in \mathcal{Y}$, $\mathfrak{Z} = (\mathfrak{U},\mathfrak{W}) \in \mathcal{Z}$ and let $N \in \mathbb{N}$, $R, \Delta, \delta \in (0,\infty)$ be real numbers which satisfy

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\ge N} \frac{|\mathbf{v}(t,x)|}{\|x\|_{\mathbb{R}^d}^q} \le \frac{\varepsilon}{12 + 6\,\|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}}, \tag{3.142}$$

$$R = 1 + \|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}, \qquad \sup_{\substack{(s,\mathbf{x}),(t,x)\in[0,T]\times\{w\in\mathbb{R}^d\colon \|w\|_{\mathbb{R}^d}\le N+2R\},\\ |s-t|+\|\mathbf{x}-x\|_{\mathbb{R}^d}\le\Delta}} |\mathbf{v}(s,\mathbf{x}) - \mathbf{v}(t,x)| \le \tfrac{\varepsilon}{3}, \tag{3.143}$$

and

$$\delta = \min\left\{1, \frac{\Delta}{\max\{1,T\}}, \frac{\varepsilon}{3\,(2+\|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)})^q}\right\}. \tag{3.144}$$

Note that it holds for all $\mathfrak{w} \in C([0,T],\mathbb{R}^d)$, $t \in [0,T]$, $x \in \mathbb{R}^d$ with $\|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)} \leq 1$ that

$$\begin{aligned}\|x+\mathfrak{w}_t\|_{\mathbb{R}^d} &\leq \|x\|_{\mathbb{R}^d} + \|\mathfrak{w}_t\|_{\mathbb{R}^d} \leq \big(1+\|\mathfrak{w}\|_{C([0,T],\mathbb{R}^d)}\big)\max\{1,\|x\|_{\mathbb{R}^d}\}\\ &\leq \big(1+\|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}+\|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}\big)\max\{1,\|x\|_{\mathbb{R}^d}\}\\ &\leq \big(2+\|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}\big)\max\{1,\|x\|_{\mathbb{R}^d}\}.\end{aligned} \tag{3.145}$$

This and Lemma 3.24 (with $d \leftarrow d$, $T \leftarrow T$, $p \leftarrow q$, $L \leftarrow 2+\|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}$, $\varrho \leftarrow ([0,T]\times\mathbb{R}^d \ni (t,x) \mapsto (T-t+\mathfrak{u}t, x+\mathfrak{w}_t) \in [0,T]\times\mathbb{R}^d)$, $v \leftarrow v-\mathbf{v}$ for $(\mathfrak{u},\mathfrak{w}) \in \mathcal{Z}$, $v \in \mathcal{Y}$ with $\|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)} \leq 1$ in the notation of Lemma 3.24) imply for all $v \in \mathcal{Y}$, $(\mathfrak{u},\mathfrak{w}) \in \mathcal{Z}$ with $\|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)} \leq 1$ that

$$\begin{aligned}&\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|v(T-t+\mathfrak{u}t,x+\mathfrak{w}_t)-\mathbf{v}(T-t+\mathfrak{u}t,x+\mathfrak{w}_t)|}{\max\{1,\|x\|^q_{\mathbb{R}^d}\}}\right]\\ &\leq \big(2+\|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}\big)^q\|v-\mathbf{v}\|_{\mathcal{Y}}.\end{aligned} \tag{3.146}$$

In addition, observe that it holds for all $\mathfrak{w} \in C([0,T],\mathbb{R}^d)$, $t \in [0,T]$, $x \in \mathbb{R}^d$ with $\|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)} \leq 1$ and $\|x\|_{\mathbb{R}^d} \geq N+R$ that

$$\|x+\mathfrak{w}_t\|_{\mathbb{R}^d} \geq \|x\|_{\mathbb{R}^d} - \|\mathfrak{w}_t\|_{\mathbb{R}^d} \geq N+1+\|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)} - \|\mathfrak{w}\|_{C([0,T],\mathbb{R}^d)} \geq N. \tag{3.147}$$

This, (3.145), and (3.142) establish for all $v \in \mathcal{Y}$, $(\mathfrak{u},\mathfrak{w}) \in \mathcal{Z}$ with $\|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)} \leq 1$ that

$$\begin{aligned}&\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq N+R}\left[\frac{|\mathbf{v}(T-t+\mathfrak{u}t,x+\mathfrak{w}_t)|}{\|x\|^q_{\mathbb{R}^d}}\right]\\ &= \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq N+R}\left[\frac{\|x+\mathfrak{w}_t\|^q_{\mathbb{R}^d}}{\max\{1,\|x\|^q_{\mathbb{R}^d}\}}\,\frac{|\mathbf{v}(T-t+\mathfrak{u}t,x+\mathfrak{w}_t)|}{\|x+\mathfrak{w}_t\|^q_{\mathbb{R}^d}}\right]\\ &\leq \left[\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\frac{\|x+\mathfrak{w}_t\|^q_{\mathbb{R}^d}}{\max\{1,\|x\|^q_{\mathbb{R}^d}\}}\right]\left[\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq N}\frac{|\mathbf{v}(t,x)|}{\|x\|^q_{\mathbb{R}^d}}\right]\\ &\leq \big(2+\|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}\big)\frac{\varepsilon}{12+6\,\|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}} = \tfrac{\varepsilon}{6}.\end{aligned} \tag{3.148}$$

Furthermore, note that it holds for all $\mathfrak{z} = (\mathfrak{u},\mathfrak{w}) \in \mathcal{Z}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ with $\mathbf{d}_{\mathcal{Z}}(\mathfrak{z},\mathfrak{Z}) = |\mathfrak{u}-\mathfrak{U}| + \|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)} \leq \delta$ and $\|x\|_{\mathbb{R}^d} \leq N+R$ that

$$\begin{aligned}\|x+\mathfrak{w}_t\|_{\mathbb{R}^d} &\leq \|x\|_{\mathbb{R}^d} + \|\mathfrak{w}_t\|_{\mathbb{R}^d} \leq N+R+\|\mathfrak{w}\|_{C([0,T],\mathbb{R}^d)}\\ &\leq N+R+\|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}+\|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}\\ &\leq N+R+1+\|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)} = N+2R\end{aligned} \tag{3.149}$$

and

$$\begin{aligned}&|T-t+\mathfrak{u}t-(T-t+\mathfrak{U}t)| + \|x+\mathfrak{w}_t-(x+\mathfrak{W}_t)\|_{\mathbb{R}^d} = |\mathfrak{u}-\mathfrak{U}|t + \|\mathfrak{w}_t-\mathfrak{W}_t\|_{\mathbb{R}^d}\\ &\leq |\mathfrak{u}-\mathfrak{U}|T + \|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)} \leq \max\{1,T\}\delta.\end{aligned} \tag{3.150}$$

Combining (3.149)–(3.150) with (3.146), (3.148), (3.144), and (3.143) ensures for all $v \in \mathcal{Y}$, $\mathfrak{z} = (\mathfrak{u}, \mathfrak{w}) \in \mathcal{Z}$ with $\|v - \mathbf{v}\|_{\mathcal{Y}} + \mathbf{d}_{\mathcal{Z}}(\mathfrak{z}, \mathfrak{Z}) \leq \delta$ that

$$\begin{aligned}
&\|\varphi_2(v, \mathfrak{z}) - \varphi_2(\mathbf{v}, \mathfrak{Z})\|_{\mathcal{Y}} \\
&= \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \left[\frac{|v(T-t+\mathfrak{u}t, x+\mathfrak{w}_t) - \mathbf{v}(T-t+\mathfrak{U}t, x+\mathfrak{W}_t)|}{\max\{1, \|x\|^q_{\mathbb{R}^d}\}} \right] \\
&\leq \left[\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|v(T-t+\mathfrak{u}t, x+\mathfrak{w}_t) - \mathbf{v}(T-t+\mathfrak{u}t, x+\mathfrak{w}_t)|}{\max\{1, \|x\|^q_{\mathbb{R}^d}\}} \right] \\
&\quad + \left[\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\leq N+R} \frac{|\mathbf{v}(T-t+\mathfrak{u}t, x+\mathfrak{w}_t) - \mathbf{v}(T-t+\mathfrak{U}t, x+\mathfrak{W}_t)|}{\max\{1, \|x\|^q_{\mathbb{R}^d}\}} \right] \\
&\quad + \left[\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq N+R} \frac{|\mathbf{v}(T-t+\mathfrak{u}t, x+\mathfrak{w}_t) - \mathbf{v}(T-t+\mathfrak{U}t, x+\mathfrak{W}_t)|}{\max\{1, \|x\|^q_{\mathbb{R}^d}\}} \right] \qquad (3.151)\\
&\leq \big(2 + \|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}\big)^q \|v - \mathbf{v}\|_{\mathcal{Y}} + \sup_{\substack{(s,\mathbf{x}),(t,x)\in[0,T]\times\{w\in\mathbb{R}^d\colon \|w\|_{\mathbb{R}^d}\leq N+2R\},\\ |s-t|+\|\mathbf{x}-x\|_{\mathbb{R}^d}\leq \max\{1,T\}\delta}} |\mathbf{v}(s,\mathbf{x}) - \mathbf{v}(t,x)| \\
&\quad + \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq N+R} \left[\frac{|\mathbf{v}(T-t+\mathfrak{u}t, x+\mathfrak{w}_t)|}{\|x\|^q_{\mathbb{R}^d}} + \frac{|\mathbf{v}(T-t+\mathfrak{U}t, x+\mathfrak{W}_t)|}{\|x\|^q_{\mathbb{R}^d}} \right] \\
&\leq \big(2 + \|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}\big)^q \delta + \tfrac{2\varepsilon}{6} + \sup_{\substack{(s,\mathbf{x}),(t,x)\in[0,T]\times\{w\in\mathbb{R}^d\colon \|w\|_{\mathbb{R}^d}\leq N+2R\},\\ |s-t|+\|\mathbf{x}-x\|_{\mathbb{R}^d}\leq \Delta}} |\mathbf{v}(s,\mathbf{x}) - \mathbf{v}(t,x)| \\
&\leq \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} = \varepsilon.
\end{aligned}$$

This proves that $\varphi_2 \colon \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y}$ is a continuous function. Observe that this implies that $\Psi_2 \colon \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y} \times \mathcal{Z}$ is a continuous function.

Third, we establish that $\varphi_3 \colon \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y}$ is a continuous function. Throughout this paragraph let $\varepsilon \in (0, \infty)$, $\mathbf{v} \in \mathcal{Y}$, $\mathfrak{Z} = (\mathfrak{U}, \mathfrak{W}) \in \mathcal{Z}$ and let $N \in \mathbb{N}$, $R, \Delta, \delta \in (0, \infty)$ be real numbers which satisfy $N \geq (6LT(2 + \|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)})^p \varepsilon^{-1})^{1/(q-p)}$ and

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq N} \left[\frac{L|\mathbf{v}(t,x)| + |f(T-t+\mathfrak{U}t, x+\mathfrak{W}_t, \mathbf{v}(t,x))|}{\|x\|^q_{\mathbb{R}^d}} \right] \leq \tfrac{\varepsilon}{6T}, \qquad (3.152)$$

$$R = 1 + \|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}, \quad \sup_{\substack{(s,\mathbf{x}),(t,x)\in[0,T]\times\{w\in\mathbb{R}^d\colon \|w\|_{\mathbb{R}^d}\leq N+R\},\\ \mathfrak{v}\in\mathbb{R},\, |\mathfrak{v}|\leq\|\mathbf{v}\|_{\mathcal{Y}}N^q,\, |s-t|+\|\mathbf{x}-x\|_{\mathbb{R}^d}\leq\Delta}} |f(s,\mathbf{x},\mathfrak{v}) - f(t,x,\mathfrak{v})| \leq \tfrac{\varepsilon}{3T}, \qquad (3.153)$$

and

$$\delta = \min\left\{1, \frac{\Delta}{\max\{1,T\}}, \frac{\varepsilon}{3\max\{1,LT\}}\right\} \qquad (3.154)$$

(cf. (ii) in Corollary 3.23). Note that it holds for all $v \in \mathcal{Y}$, $\mathfrak{z} = (\mathfrak{u}, \mathfrak{w}) \in \mathcal{Z}$ that

$$\begin{aligned}
&\|\varphi_3(v, \mathfrak{z}) - \varphi_3(\mathbf{v}, \mathfrak{Z})\|_{\mathcal{Y}} \\
&= \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \left[\frac{t|f(T-t+\mathfrak{u}t, x+\mathfrak{w}_t, v(t,x)) - f(T-t+\mathfrak{U}t, x+\mathfrak{W}_t, \mathbf{v}(t,x))|}{\max\{1, \|x\|^q_{\mathbb{R}^d}\}} \right]
\end{aligned}$$

$$
\begin{aligned}
&\leq T \sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|f(T-t+\mathfrak{u}t,x+\mathfrak{w}_t,v(t,x))-f(T-t+\mathfrak{u}t,x+\mathfrak{w}_t,\mathbf{v}(t,x))|}{\max\{1,\|x\|_{\mathbb{R}^d}^q\}}\right] \\
&\quad + T \sup_{\substack{(t,x)\in[0,T]\times\mathbb{R}^d,\\ \|x\|_{\mathbb{R}^d}\leq N}}\left[\frac{|f(T-t+\mathfrak{u}t,x+\mathfrak{w}_t,\mathbf{v}(t,x))-f(T-t+\mathfrak{U}t,x+\mathfrak{W}_t,\mathbf{v}(t,x))|}{\max\{1,\|x\|_{\mathbb{R}^d}^q\}}\right] \\
&\quad + T \sup_{\substack{(t,x)\in[0,T]\times\mathbb{R}^d,\\ \|x\|_{\mathbb{R}^d}\geq N}}\left[\frac{|f(T-t+\mathfrak{u}t,x+\mathfrak{w}_t,\mathbf{v}(t,x))-f(T-t+\mathfrak{U}t,x+\mathfrak{W}_t,\mathbf{v}(t,x))|}{\|x\|_{\mathbb{R}^d}^q}\right].
\end{aligned} \tag{3.155}
$$

Next observe that it holds for all $\mathfrak{z}=(\mathfrak{u},\mathfrak{w})\in\mathcal{Z}$, $t\in[0,T]$, $x\in\mathbb{R}^d$ with $\mathbf{d}_{\mathcal{Z}}(\mathfrak{z},\mathfrak{Z})=|\mathfrak{u}-\mathfrak{U}|+\|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}\leq\delta$ and $\|x\|_{\mathbb{R}^d}\leq N$ that $\|x+\mathfrak{w}_t\|_{\mathbb{R}^d}\leq N+R$, $|\mathbf{v}(t,x)|\leq\|\mathbf{v}\|_{\mathcal{Y}}\max\{1,\|x\|_{\mathbb{R}^d}^q\}\leq\|\mathbf{v}\|_{\mathcal{Y}}N^q$, and

$$
|T-t+\mathfrak{u}t-(T-t+\mathfrak{U}t)|+\|x+\mathfrak{w}_t-(x+\mathfrak{W}_t)\|_{\mathbb{R}^d}\leq\max\{1,T\}\delta. \tag{3.156}
$$

This and (3.153) show for all $\mathfrak{z}=(\mathfrak{u},\mathfrak{w})\in\mathcal{Z}$ with $\mathbf{d}_{\mathcal{Z}}(\mathfrak{z},\mathfrak{Z})\leq\delta$ that

$$
\begin{aligned}
&\sup_{\substack{(t,x)\in[0,T]\times\mathbb{R}^d,\\ \|x\|_{\mathbb{R}^d}\leq N}}\left[\frac{|f(T-t+\mathfrak{u}t,x+\mathfrak{w}_t,\mathbf{v}(t,x))-f(T-t+\mathfrak{U}t,x+\mathfrak{W}_t,\mathbf{v}(t,x))|}{\max\{1,\|x\|_{\mathbb{R}^d}^q\}}\right] \\
&\leq \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\,\|x\|_{\mathbb{R}^d}\leq N}|f(T-t+\mathfrak{u}t,x+\mathfrak{w}_t,\mathbf{v}(t,x))-f(T-t+\mathfrak{U}t,x+\mathfrak{W}_t,\mathbf{v}(t,x))| \\
&\leq \sup_{\substack{(s,\mathbf{x}),(t,x)\in[0,T]\times\{w\in\mathbb{R}^d\colon\|w\|_{\mathbb{R}^d}\leq N+R\},\\ \mathfrak{v}\in\mathbb{R},\,|\mathfrak{v}|\leq\|\mathbf{v}\|_{\mathcal{Y}}N^q,\,|s-t|+\|\mathbf{x}-x\|_{\mathbb{R}^d}\leq\max\{1,T\}\delta}}|f(s,\mathbf{x},\mathfrak{v})-f(t,x,\mathfrak{v})| \\
&\leq \sup_{\substack{(s,\mathbf{x}),(t,x)\in[0,T]\times\{w\in\mathbb{R}^d\colon\|w\|_{\mathbb{R}^d}\leq N+R\},\\ \mathfrak{v}\in\mathbb{R},\,|\mathfrak{v}|\leq\|\mathbf{v}\|_{\mathcal{Y}}N^q,\,|s-t|+\|\mathbf{x}-x\|_{\mathbb{R}^d}\leq\Delta}}|f(s,\mathbf{x},\mathfrak{v})-f(t,x,\mathfrak{v})|\leq\tfrac{\varepsilon}{3T}.
\end{aligned} \tag{3.157}
$$

Furthermore, (3.152) and (3.145) ensure for all $(\mathfrak{u},\mathfrak{w})\in\mathcal{Z}$ with $\|\mathfrak{w}-\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}\leq 1$ that

$$
\begin{aligned}
&\sup_{\substack{(t,x)\in[0,T]\times\mathbb{R}^d,\\ \|x\|_{\mathbb{R}^d}\geq N}}\left[\frac{|f(T-t+\mathfrak{u}t,x+\mathfrak{w}_t,\mathbf{v}(t,x))-f(T-t+\mathfrak{U}t,x+\mathfrak{W}_t,\mathbf{v}(t,x))|}{\|x\|_{\mathbb{R}^d}^q}\right] \\
&\leq \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\,\|x\|_{\mathbb{R}^d}\geq N}\left[\frac{|f(T-t+\mathfrak{u}t,x+\mathfrak{w}_t,\mathbf{v}(t,x))-f(T-t+\mathfrak{u}t,x+\mathfrak{w}_t,0)|}{\|x\|_{\mathbb{R}^d}^q}\right. \\
&\qquad\qquad \left.+\frac{|f(T-t+\mathfrak{u}t,x+\mathfrak{w}_t,0)-f(T-t+\mathfrak{U}t,x+\mathfrak{W}_t,\mathbf{v}(t,x))|}{\|x\|_{\mathbb{R}^d}^q}\right] \\
&\leq \left[\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\,\|x\|_{\mathbb{R}^d}\geq N}\frac{L|\mathbf{v}(t,x)|+|f(T-t+\mathfrak{U}t,x+\mathfrak{W}_t,\mathbf{v}(t,x))|}{\|x\|_{\mathbb{R}^d}^q}\right] \\
&\quad + \left[\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\,\|x\|_{\mathbb{R}^d}\geq N}\frac{|f(T-t+\mathfrak{u}t,x+\mathfrak{w}_t,0)|}{\|x\|_{\mathbb{R}^d}^q}\right] \\
&\leq \tfrac{\varepsilon}{6T}+L\left[\sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\,\|x\|_{\mathbb{R}^d}\geq N}\frac{\max\{1,\|x+\mathfrak{w}_t\|_{\mathbb{R}^d}^p\}}{\|x\|_{\mathbb{R}^d}^q}\right]
\end{aligned} \tag{3.158}
$$

$$\le \tfrac{\varepsilon}{6T} + L\big(2 + \|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}\big)^p \left[\sup_{x\in\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\ge N} \frac{1}{\|x\|_{\mathbb{R}^d}^{q-p}}\right]$$
$$= \tfrac{\varepsilon}{6T} + L\big(2 + \|\mathfrak{W}\|_{C([0,T],\mathbb{R}^d)}\big)^p \frac{1}{N^{q-p}} \le \tfrac{\varepsilon}{6T} + \tfrac{\varepsilon}{6T} = \tfrac{\varepsilon}{3T}.$$

Combining (3.155) with (3.157), (3.158), and (3.154) establishes for all $v \in \mathcal{Y}$, $\mathfrak{z} = (\mathfrak{u}, \mathfrak{w}) \in \mathcal{Z}$ with $\|v - \mathbf{v}\|_{\mathcal{Y}} + \mathbf{d}_{\mathcal{Z}}(\mathfrak{z}, \mathfrak{Z}) \le \delta$ that

$$\|\varphi_3(v, \mathfrak{z}) - \varphi_3(\mathbf{v}, \mathfrak{Z})\|_{\mathcal{Y}} \le LT\|v - \mathbf{v}\|_{\mathcal{Y}} + \tfrac{2\varepsilon}{3} \le LT\delta + \tfrac{2\varepsilon}{3} \le \tfrac{\varepsilon}{3} + \tfrac{2\varepsilon}{3} = \varepsilon. \tag{3.159}$$

From this it follows that $\varphi_3 \colon \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y}$ is a continuous function.

As a next step observe that the fact that φ_2, Ψ_1, Ψ_2, and φ_3 are continuous functions, the fact that $\mathcal{Z} \ni \mathfrak{z} \mapsto (\mathfrak{g}, \mathfrak{z}) \in \mathcal{Y}\times\mathcal{Z}$ is a continuous function, and the fact that $\varphi_4 \colon \mathcal{Y} \to \mathcal{Y}$ is a linear isometry demonstrate that $G\colon \mathcal{Z} \to \mathcal{Y}$ and $F\colon \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y}$ are continuous functions. Combining this with (3.140) proves (i). Finally, the fact that $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ is a separable $\mathbb{R}$-Banach space (cf. (ii) in Proposition 3.20), the fact that $(\mathcal{Z}, \mathbf{d}_{\mathcal{Z}})$ is a separable metric space, and (i) establish (ii). The proof of Lemma 3.26 is thus complete. □

3.2.2.3 Recursive formulation

Lemma 3.27. *Assume Setting 3.1. Then*

(i) *it holds for all $n \in (\mathbb{N}_0 \cup \{-1\})$, $j \in \mathbb{N}$, $\theta \in \Theta$ that $Y^\theta_{n,j}(\Omega) \subseteq \mathcal{Y}$,*

(ii) *it holds for all $n, j \in \mathbb{N}$, $\theta \in \Theta$ that $Y^\theta_{-1,j} = Y^\theta_{0,j} = 0$ and*

$$Y^\theta_{n,j} = \sum_{l=0}^{n-1} \tfrac{1}{(M_j)^{n-l}} \left[\sum_{i=1}^{(M_j)^{n-l}} \Phi_l\big(Y^{(\theta,l,i)}_{l,j}, Y^{(\theta,-l,i)}_{l-1,j}, Z^{(\theta,l,i)}\big)\right], \tag{3.160}$$

and

(iii) *it holds for all $n \in (\mathbb{N}_0 \cup \{-1\})$, $j \in \mathbb{N}$, $\theta \in \Theta$ that $\Omega \ni \omega \mapsto Y^\theta_{n,j}(\omega) \in \mathcal{Y}$ is an $\mathscr{F}/\mathscr{B}(\mathcal{Y})$-measurable function.*

Proof of Lemma 3.27. We show (i)–(ii) by induction on $n \in \mathbb{N}$. For the base case $n = 1$ note that the fact that $\forall\, j \in \mathbb{N}, \theta \in \Theta\colon Y^\theta_{-1,j} = Y^\theta_{0,j} = 0$ implies for all $j \in \mathbb{N}$, $\theta \in \Theta$ that

$$Y^\theta_{-1,j}, Y^\theta_{0,j} \in \mathcal{Y}. \tag{3.161}$$

Next observe that (3.128) and (3.133) ensure for all $j \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0, T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} Y^\theta_{1,j}(T-t, x) &= \tfrac{1}{M_j}\left[\sum_{i=1}^{M_j} g\big(x + W^{(\theta,0,i)}_t\big)\right] \\ &\quad + \tfrac{t}{M_j}\left[\sum_{i=1}^{M_j} f\Big(T - t + U^{(\theta,0,i)}t, x + W^{(\theta,0,i)}_{U^{(\theta,0,i)}t}, Y^{(\theta,0,i)}_{0,j}\big(T - t + U^{(\theta,0,i)}t, x + W^{(\theta,0,i)}_{U^{(\theta,0,i)}t}\big)\Big)\right] \\ &= \tfrac{1}{M_j}\left[\sum_{i=1}^{M_j} \big[\Phi_0\big(Y^{(\theta,0,i)}_{0,j}, Y^{(\theta,0,i)}_{-1,j}, Z^{(\theta,0,i)}\big)\big](T-t, x)\right]. \end{aligned} \tag{3.162}$$

This and (3.161) prove (i)–(ii) in the base case $n = 1$. For the induction step $\mathbb{N} \ni n-1 \to n \in \{2, 3, \ldots\}$ let $n \in \{2, 3, \ldots\}$ and assume for all $l \in \{-1, 0, 1, \ldots, n-1\}$, $j \in \mathbb{N}$, $\theta \in \Theta$ that $Y^{\theta}_{l,j}(\Omega) \subseteq \mathcal{Y}$. Equations (3.128) and (3.133) hence demonstrate for all $j \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0, T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&Y^{\theta}_{n,j}(T-t,x) = \tfrac{1}{(M_j)^n}\left[\sum_{i=1}^{(M_j)^n} g\big(x + W^{(\theta,0,i)}_t\big)\right] + \sum_{l=0}^{n-1} \tfrac{t}{(M_j)^{n-l}}\left[\sum_{i=1}^{(M_j)^{n-l}}\right.\\
&\quad \Big[f\Big(T-t+U^{(\theta,l,i)}t, x + W^{(\theta,l,i)}_{U^{(\theta,l,i)}t}, Y^{(\theta,l,i)}_{l,j}\big(T-t+U^{(\theta,l,i)}t, x+W^{(\theta,l,i)}_{U^{(\theta,l,i)}t}\big)\Big)\\
&\quad \left.- \mathbb{1}_{\mathbb{N}}(l) f\Big(T-t+U^{(\theta,l,i)}t, x + W^{(\theta,l,i)}_{U^{(\theta,l,i)}t}, Y^{(\theta,-l,i)}_{l-1,j}\big(T-t+U^{(\theta,l,i)}t, x+W^{(\theta,l,i)}_{U^{(\theta,l,i)}t}\big)\Big)\Big]\right]\\
&= \tfrac{1}{(M_j)^n}\left[\sum_{i=1}^{(M_j)^n} g\big(x + W^{(\theta,0,i)}_t\big)\right]\\
&\quad + \tfrac{t}{(M_j)^n}\left[\sum_{i=1}^{(M_j)^n} f\Big(T-t+U^{(\theta,0,i)}t, x + W^{(\theta,0,i)}_{U^{(\theta,0,i)}t}, Y^{(\theta,0,i)}_{0,j}\big(T-t+U^{(\theta,0,i)}t, x+W^{(\theta,0,i)}_{U^{(\theta,0,i)}t}\big)\Big)\right]\\
&\quad + \sum_{l=1}^{n-1} \tfrac{t}{(M_j)^{n-l}}\left[\sum_{i=1}^{(M_j)^{n-l}}\right.\\
&\qquad \Big[f\Big(T-t+U^{(\theta,l,i)}t, x + W^{(\theta,l,i)}_{U^{(\theta,l,i)}t}, Y^{(\theta,l,i)}_{l,j}\big(T-t+U^{(\theta,l,i)}t, x+W^{(\theta,l,i)}_{U^{(\theta,l,i)}t}\big)\Big)\\
&\qquad \left.- f\Big(T-t+U^{(\theta,l,i)}t, x + W^{(\theta,l,i)}_{U^{(\theta,l,i)}t}, Y^{(\theta,-l,i)}_{l-1,j}\big(T-t+U^{(\theta,l,i)}t, x+W^{(\theta,l,i)}_{U^{(\theta,l,i)}t}\big)\Big)\Big]\right]\\
&= \tfrac{1}{(M_j)^n}\left[\sum_{i=1}^{(M_j)^n} \big[\Phi_0\big(Y^{(\theta,0,i)}_{0,j}, Y^{(\theta,0,i)}_{-1,j}, Z^{(\theta,0,i)}\big)\big](T-t,x)\right]\\
&\quad + \sum_{l=1}^{n-1} \tfrac{1}{(M_j)^{n-l}}\left[\sum_{i=1}^{(M_j)^{n-l}} \big[\Phi_l\big(Y^{(\theta,l,i)}_{l,j}, Y^{(\theta,-l,i)}_{l-1,j}, Z^{(\theta,l,i)}\big)\big](T-t,x)\right]\\
&= \sum_{l=0}^{n-1} \tfrac{1}{(M_j)^{n-l}}\left[\sum_{i=1}^{(M_j)^{n-l}} \big[\Phi_l\big(Y^{(\theta,l,i)}_{l,j}, Y^{(\theta,-l,i)}_{l-1,j}, Z^{(\theta,l,i)}\big)\big](T-t,x)\right].
\end{aligned} \tag{3.163}$$

Induction hence establishes (i)–(ii).

Furthermore, combining (i)–(ii) with (ii) in Lemma 3.26 and (i) in Proposition 3.8 shows (iii). The proof of Lemma 3.27 is thus complete. □

3.2.2.4 Integrability

Lemma 3.28. *Assume Setting 3.1. Then it holds for all $l \in \mathbb{N}_0$, $j \in \mathbb{N}$, $r \in [0, \infty)$ that*

$$\mathbb{E}\Big[\big\|\Phi_l\big(Y^0_{l,j}, Y^1_{l-1,j}, Z^0\big)\big\|^r_{\mathcal{Y}} + \big\|Y^0_{l-1,j}\big\|^r_{\mathcal{Y}}\Big] < \infty \tag{3.164}$$

(cf. (iii) in Lemma 3.27).

Proof of Lemma 3.28. First of all, note that it holds for all $(\mathfrak{u},\mathfrak{w})\in\mathcal{Z}$, $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$\|x+\mathfrak{w}_{\mathfrak{u}t}\|_{\mathbb{R}^d}\le\|x\|_{\mathbb{R}^d}+\|\mathfrak{w}_{\mathfrak{u}t}\|_{\mathbb{R}^d}\le\left(1+\sup_{s\in[0,T]}\|\mathfrak{w}_s\|_{\mathbb{R}^d}\right)\max\{1,\|x\|_{\mathbb{R}^d}\}. \tag{3.165}$$

This and Lemma 3.24 (with $d\leftarrow d$, $T\leftarrow T$, $p\leftarrow p$, $L\leftarrow 1+\sup_{t\in[0,T]}\|\mathfrak{w}_t\|_{\mathbb{R}^d}$, $\varrho\leftarrow([0,T]\times\mathbb{R}^d\ni(t,x)\mapsto(t,x+\mathfrak{w}_t)\in[0,T]\times\mathbb{R}^d)$, $v\leftarrow([0,T]\times\mathbb{R}^d\ni(t,x)\mapsto g(x)\in\mathbb{R})$ for $\mathfrak{w}\in C([0,T],\mathbb{R}^d)$ in the notation of Lemma 3.24) show for all $\mathfrak{w}\in C([0,T],\mathbb{R}^d)$ that

$$\begin{aligned}&\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|g(x+\mathfrak{w}_t)|}{\max\{1,\|x\|^q_{\mathbb{R}^d}\}}\right]\le\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|g(x+\mathfrak{w}_t)|}{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}\right]\\&\le\left(1+\sup_{t\in[0,T]}\|\mathfrak{w}_t\|_{\mathbb{R}^d}\right)^p\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|g(x)|}{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}\right]\le L\left(1+\sup_{t\in[0,T]}\|\mathfrak{w}_t\|_{\mathbb{R}^d}\right)^p.\end{aligned}\tag{3.166}$$

Similarly, (3.165) and Lemma 3.24 (with $d\leftarrow d$, $T\leftarrow T$, $p\leftarrow p$, $L\leftarrow 1+\sup_{t\in[0,T]}\|\mathfrak{w}_t\|_{\mathbb{R}^d}$, $\varrho\leftarrow([0,T]\times\mathbb{R}^d\ni(t,x)\mapsto(T-t+\mathfrak{u}t,x+\mathfrak{w}_{\mathfrak{u}t})\in[0,T]\times\mathbb{R}^d)$, $v\leftarrow([0,T]\times\mathbb{R}^d\ni(t,x)\mapsto tf(t,x,0)\in\mathbb{R})$ for $(\mathfrak{u},\mathfrak{w})\in\mathcal{Z}$ in the notation of Lemma 3.24) ensure for all $(\mathfrak{u},\mathfrak{w})\in\mathcal{Z}$ that

$$\begin{aligned}&\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|tf(T-t+\mathfrak{u}t,x+\mathfrak{w}_{\mathfrak{u}t},0)|}{\max\{1,\|x\|^q_{\mathbb{R}^d}\}}\right]\le\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|tf(T-t+\mathfrak{u}t,x+\mathfrak{w}_{\mathfrak{u}t},0)|}{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}\right]\\&\le\left(1+\sup_{t\in[0,T]}\|\mathfrak{w}_t\|_{\mathbb{R}^d}\right)^p\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|tf(t,x,0)|}{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}\right]\\&\le T\left(1+\sup_{t\in[0,T]}\|\mathfrak{w}_t\|_{\mathbb{R}^d}\right)^p\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|f(t,x,0)|}{\max\{1,\|x\|^p_{\mathbb{R}^d}\}}\right]\le LT\left(1+\sup_{t\in[0,T]}\|\mathfrak{w}_t\|_{\mathbb{R}^d}\right)^p.\end{aligned}\tag{3.167}$$

Combining (3.133), (3.166), and (3.167) implies for all $w\in\mathcal{Y}$, $\mathfrak{z}=(\mathfrak{u},\mathfrak{w})\in\mathcal{Z}$ that

$$\begin{aligned}\|\Phi_0(0,w,\mathfrak{z})\|_{\mathcal{Y}}&=\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|[\Phi_0(0,w,\mathfrak{z})](T-t,x)|}{\max\{1,\|x\|^q_{\mathbb{R}^d}\}}\right]\\&\le\left[\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\frac{|g(x+\mathfrak{w}_t)|}{\max\{1,\|x\|^q_{\mathbb{R}^d}\}}\right]+\left[\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\frac{|tf(T-t+\mathfrak{u}t,x+\mathfrak{w}_{\mathfrak{u}t},0)|}{\max\{1,\|x\|^q_{\mathbb{R}^d}\}}\right]\\&\le L(T+1)\left(1+\sup_{t\in[0,T]}\|\mathfrak{w}_t\|_{\mathbb{R}^d}\right)^p.\end{aligned}\tag{3.168}$$

In addition, (3.133), (3.165) and Lemma 3.24 (with $d\leftarrow d$, $T\leftarrow T$, $p\leftarrow q$, $L\leftarrow 1+\sup_{t\in[0,T]}\|\mathfrak{w}_t\|_{\mathbb{R}^d}$, $\varrho\leftarrow([0,T]\times\mathbb{R}^d\ni(t,x)\mapsto(T-t+\mathfrak{u}t,x+\mathfrak{w}_{\mathfrak{u}t})\in[0,T]\times\mathbb{R}^d)$, $v\leftarrow v-w$ for $(\mathfrak{u},\mathfrak{w})\in\mathcal{Z}$, $v,w\in\mathcal{Y}$ in the notation of Lemma 3.24) prove for all $l\in\mathbb{N}$, $v,w\in\mathcal{Y}$, $\mathfrak{z}=(\mathfrak{u},\mathfrak{w})\in\mathcal{Z}$ that

$$\begin{aligned}\|\Phi_l(v,w,\mathfrak{z})\|_{\mathcal{Y}}&=\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|[\Phi_l(v,w,\mathfrak{z})](T-t,x)|}{\max\{1,\|x\|^q_{\mathbb{R}^d}\}}\right]\\&=\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\Bigg[\frac{t}{\max\{1,\|x\|^q_{\mathbb{R}^d}\}}\Big|f\big(T-t+\mathfrak{u}t,x+\mathfrak{w}_{\mathfrak{u}t},v(T-t+\mathfrak{u}t,x+\mathfrak{w}_{\mathfrak{u}t})\big)\\&\qquad-f\big(T-t+\mathfrak{u}t,x+\mathfrak{w}_{\mathfrak{u}t},w(T-t+\mathfrak{u}t,x+\mathfrak{w}_{\mathfrak{u}t})\big)\Big|\Bigg]\end{aligned}$$

$$\begin{aligned}
&\leq LT \sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|v(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t}) - w(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t})|}{\max\{1,\|x\|_{\mathbb{R}^d}^q\}}\right] \\
&\leq LT\left(1+\sup_{t\in[0,T]}\|\mathfrak{w}_t\|_{\mathbb{R}^d}\right)^q \sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[\frac{|v(t,x)-w(t,x)|}{\max\{1,\|x\|_{\mathbb{R}^d}^q\}}\right] \\
&= LT\left(1+\sup_{t\in[0,T]}\|\mathfrak{w}_t\|_{\mathbb{R}^d}\right)^q \|v-w\|_{\mathcal{Y}}.
\end{aligned} \tag{3.169}$$

Next we claim that it holds for all $l \in \mathbb{N}_0$, $j \in \mathbb{N}$, $r \in [0,\infty)$ that

$$\mathbb{E}\Big[\big\|\Phi_l\big(Y_{l,j}^0, Y_{l-1,j}^1, Z^0\big)\big\|_{\mathcal{Y}}^r + \big\|Y_{l,j}^0\big\|_{\mathcal{Y}}^r + \big\|Y_{l-1,j}^0\big\|_{\mathcal{Y}}^r\Big] < \infty. \tag{3.170}$$

We establish (3.170) by induction on $l \in \mathbb{N}_0$. For the base case $l = 0$ observe that (3.168) and the fact that $\forall\, a,b,r \in [0,\infty)\colon (a+b)^r \leq 2^{\max\{r-1,0\}}(a^r+b^r)$ show for all $j \in \mathbb{N}$, $r \in [0,\infty)$ that

$$\begin{aligned}
\mathbb{E}\Big[\big\|\Phi_0\big(Y_{0,j}^0, Y_{-1,j}^1, Z^0\big)\big\|_{\mathcal{Y}}^r\Big] &= \mathbb{E}\Big[\big\|\Phi_0\big(0,0,U^0,W^0\big)\big\|_{\mathcal{Y}}^r\Big] \\
&\leq L^r(T+1)^r\,\mathbb{E}\left[\left(1+\sup_{t\in[0,T]}\|W_t^0\|_{\mathbb{R}^d}\right)^{pr}\right] \\
&\leq 2^{\max\{pr-1,0\}}L^r(T+1)^r\left(1+\mathbb{E}\left[\sup_{t\in[0,T]}\|W_t^0\|_{\mathbb{R}^d}^{pr}\right]\right) < \infty.
\end{aligned} \tag{3.171}$$

This and the fact that $\forall\, j \in \mathbb{N}, r \in [0,\infty)\colon \mathbb{E}\big[\|Y_{0,j}^0\|_{\mathcal{Y}}^r + \|Y_{-1,j}^0\|_{\mathcal{Y}}^r\big] = 0 < \infty$ prove (3.170) in the base case $l = 0$. For the induction step $\mathbb{N}_0 \ni l-1 \to l \in \mathbb{N}$ let $l \in \mathbb{N}$ and assume that it holds for all $k \in \{0,1,\ldots,l-1\}$, $j \in \mathbb{N}$, $r \in [0,\infty)$ that

$$\mathbb{E}\Big[\big\|\Phi_k\big(Y_{k,j}^0, Y_{k-1,j}^1, Z^0\big)\big\|_{\mathcal{Y}}^r + \big\|Y_{k,j}^0\big\|_{\mathcal{Y}}^r + \big\|Y_{k-1,j}^0\big\|_{\mathcal{Y}}^r\Big] < \infty. \tag{3.172}$$

Note that this, (ii) in Lemma 3.27, and (vi) in Proposition 3.8 ensure for all $j \in \mathbb{N}$, $r \in [1,\infty)$ that

$$\begin{aligned}
\left(\mathbb{E}\Big[\big\|Y_{l,j}^0\big\|_{\mathcal{Y}}^r\Big]\right)^{1/r} &= \left(\mathbb{E}\left[\left\|\sum_{k=0}^{l-1}\tfrac{1}{(M_j)^{l-k}}\left[\sum_{i=1}^{(M_j)^{l-k}}\Phi_k\big(Y_{k,j}^{(0,k,i)}, Y_{k-1,j}^{(0,-k,i)}, Z^{(0,k,i)}\big)\right]\right\|_{\mathcal{Y}}^r\right]\right)^{1/r} \\
&\leq \sum_{k=0}^{l-1}\tfrac{1}{(M_j)^{l-k}}\left[\sum_{i=1}^{(M_j)^{l-k}}\left(\mathbb{E}\Big[\big\|\Phi_k\big(Y_{k,j}^{(0,k,i)}, Y_{k-1,j}^{(0,-k,i)}, Z^{(0,k,i)}\big)\big\|_{\mathcal{Y}}^r\Big]\right)^{1/r}\right] \\
&= \sum_{k=0}^{l-1}\left(\mathbb{E}\Big[\big\|\Phi_k\big(Y_{k,j}^0, Y_{k-1,j}^1, Z^0\big)\big\|_{\mathcal{Y}}^r\Big]\right)^{1/r} < \infty.
\end{aligned} \tag{3.173}$$

Hölder's inequality, (3.169), the fact that $\forall\, a,b,r \in [0,\infty)\colon (a+b)^r \leq 2^{\max\{r-1,0\}}(a^r+b^r)$, (v) in Proposition 3.8, and (3.172) hence demonstrate for all $j \in \mathbb{N}$, $r \in [1,\infty)$ that

$$\begin{aligned}
&\left(\mathbb{E}\Big[\big\|\Phi_l\big(Y_{l,j}^0, Y_{l-1,j}^1, Z^0\big)\big\|_{\mathcal{Y}}^r\Big]\right)^{1/r} \leq LT\left(\mathbb{E}\left[\left(1+\sup_{t\in[0,T]}\|W_t^0\|_{\mathbb{R}^d}\right)^{qr}\big\|Y_{l,j}^0 - Y_{l-1,j}^1\big\|_{\mathcal{Y}}^r\right]\right)^{1/r} \\
&\leq LT\left(\mathbb{E}\left[\left(1+\sup_{t\in[0,T]}\|W_t^0\|_{\mathbb{R}^d}\right)^{2qr}\right]\right)^{1/(2r)}\left(\mathbb{E}\Big[\big\|Y_{l,j}^0 - Y_{l-1,j}^1\big\|_{\mathcal{Y}}^{2r}\Big]\right)^{1/(2r)}
\end{aligned} \tag{3.174}$$

$$\leq 2^{\max\{q-1/(2r),0\}} LT \left(1+\mathbb{E}\left[\sup_{t\in[0,T]} \|W_t^0\|_{\mathbb{R}^d}^{2qr}\right]\right)^{1/(2r)} \cdot \left[\left(\mathbb{E}\left[\|Y_{l,j}^0\|_{\mathcal{Y}}^{2r}\right]\right)^{1/(2r)} + \left(\mathbb{E}\left[\|Y_{l-1,j}^0\|_{\mathcal{Y}}^{2r}\right]\right)^{1/(2r)}\right] < \infty.$$

Combining this with (3.173) and (3.172) establishes for all $j\in\mathbb{N}$, $r\in[0,\infty)$ that

$$\mathbb{E}\left[\|\Phi_l(Y_{l,j}^0, Y_{l-1,j}^1, Z^0)\|_{\mathcal{Y}}^r + \|Y_{l,j}^0\|_{\mathcal{Y}}^r + \|Y_{l-1,j}^0\|_{\mathcal{Y}}^r\right] < \infty. \tag{3.175}$$

Induction hence proves (3.170). The proof of Lemma 3.28 is thus complete. □

3.2.2.5 Estimates

Lemma 3.29. *Assume Setting 3.1 and let $C\in[0,\infty)$ be given by*

$$C = e^{LT}\left[\left(\mathbb{E}\left[|g(\xi+W_T^0)|^2\right]\right)^{1/2} + \sqrt{T}\left(\int_0^T \mathbb{E}\left[|f(t,\xi+W_t^0,0)|^2\right] \mathrm{d}t\right)^{1/2}\right]. \tag{3.176}$$

Then it holds for all $k\in\mathbb{N}_0$ that

$$\max\{\mathbb{E}\left[|\psi_k(\Phi_0(0,0,Z^0))|^2\right], \mathbb{E}\left[|\psi_k(y)|^2\right]\} \leq \tfrac{C^2}{k!}. \tag{3.177}$$

Proof of Lemma 3.29. Throughout this proof let $F\colon C([0,T]\times\mathbb{R}^d,\mathbb{R}) \to C([0,T]\times\mathbb{R}^d,\mathbb{R})$ satisfy for all $v\in C([0,T]\times\mathbb{R}^d,\mathbb{R})$, $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$[F(v)](t,x) = f(t,x,v(t,x)). \tag{3.178}$$

Observe that (3.132), the fact that $\mathbf{U}$ and $\mathbf{W}$ are independent, and Hutzenthaler et al. [181, Lemma 2.3] (with $S \leftarrow [0,1]$, $U \leftarrow ([0,1]\times\Omega \ni (s,\omega) \mapsto s^{k-1}/(k-1)!|y(sT,\xi+\mathbf{W}_{sT}(\omega))|^2 \in [0,\infty))$, $Y\leftarrow \mathbf{U}$ for $k\in\mathbb{N}$ in the notation of [181, Lemma 2.3]) imply for all $k\in\mathbb{N}$ that

$$\begin{aligned}\mathbb{E}\left[|\psi_k(y)|^2\right] &= \mathbb{E}\left[\tfrac{\mathbf{U}^{k-1}}{(k-1)!}|y(\mathbf{U}T,\xi+\mathbf{W}_{\mathbf{U}T})|^2\right] \\ &= \tfrac{1}{(k-1)!}\int_0^1 s^{k-1}\,\mathbb{E}\left[|y(sT,\xi+\mathbf{W}_{sT})|^2\right]\mathrm{d}s \leq \tfrac{1}{k!}\left[\sup_{t\in[0,T]}\mathbb{E}\left[|y(t,\xi+\mathbf{W}_t)|^2\right]\right].\end{aligned} \tag{3.179}$$

This, the fact that $\mathbb{E}[|\psi_0(y)|^2] = |y(0,\xi)|^2 = \mathbb{E}[|y(0,\xi+\mathbf{W}_0)|^2]$, and [181, Lemma 3.4] establish for all $k\in\mathbb{N}_0$ that

$$\begin{aligned}\mathbb{E}\left[|\psi_k(y)|^2\right] &\leq \tfrac{1}{k!}\left[\sup_{t\in[0,T]}\mathbb{E}\left[|y(t,\xi+\mathbf{W}_t)|^2\right]\right] \\ &\leq \tfrac{e^{2LT}}{k!}\left[\left(\mathbb{E}\left[|g(\xi+\mathbf{W}_T)|^2\right]\right)^{1/2} + \sqrt{T}\left(\int_0^T \mathbb{E}\left[|[F(0)](t,\xi+\mathbf{W}_t)|^2\right]\mathrm{d}t\right)^{1/2}\right]^2 \\ &= \tfrac{e^{2LT}}{k!}\left[\left(\mathbb{E}\left[|g(\xi+W_T^0)|^2\right]\right)^{1/2} + \sqrt{T}\left(\int_0^T \mathbb{E}\left[|f(t,\xi+W_t^0,0)|^2\right]\mathrm{d}t\right)^{1/2}\right]^2 = \tfrac{C^2}{k!}.\end{aligned} \tag{3.180}$$

Next note that (3.133) shows for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$[\Phi_0(0,0,Z^0)](t,x) = g(x+W_{T-t}^0) + (T-t)f\big(t+(T-t)U^0, x+W_{(T-t)U^0}^0, 0\big). \tag{3.181}$$

This, (3.132), and Hölder's inequality demonstrate for all $k \in \mathbb{N}$ that

$$
\begin{aligned}
&\left(\mathbb{E}\left[|\psi_k(\Phi_0(0,0,Z^0))|^2\right]\right)^{1/2} = \mathbb{E}\left[\tfrac{\mathbf{U}^{k-1}}{(k-1)!}|[\Phi_0(0,0,Z^0)](\mathbf{U}T, \xi + \mathbf{W}_{\mathbf{U}T})|^2\right] \\
&= \bigg(\mathbb{E}\bigg[\tfrac{\mathbf{U}^{k-1}}{(k-1)!}\big|g(\xi + \mathbf{W}_{\mathbf{U}T} + W^0_{(1-\mathbf{U})T}) \\
&\qquad + (1-\mathbf{U})Tf\big(\mathbf{U}T + (1-\mathbf{U})U^0T, \xi + \mathbf{W}_{\mathbf{U}T} + W^0_{(1-\mathbf{U})U^0T}, 0\big)\big|^2\bigg]\bigg)^{1/2} \\
&\leq \left(\mathbb{E}\left[\tfrac{\mathbf{U}^{k-1}}{(k-1)!}|g(\xi + \mathbf{W}_{\mathbf{U}T} + W^0_{(1-\mathbf{U})T})|^2\right]\right)^{1/2} \\
&\quad + \left(\mathbb{E}\left[\tfrac{\mathbf{U}^{k-1}}{(k-1)!}\big|(1-\mathbf{U})Tf\big(\mathbf{U}T + (1-\mathbf{U})U^0T, \xi + \mathbf{W}_{\mathbf{U}T} + W^0_{(1-\mathbf{U})U^0T}, 0\big)\big|^2\right]\right)^{1/2}.
\end{aligned}
\tag{3.182}
$$

The fact that $\mathbf{U}$, $\mathbf{W}$, and W^0 are independent and [181, Lemma 2.3] (with $S \leftarrow [0,1]$, $U \leftarrow ([0,1] \times \Omega \ni (s,\omega) \mapsto {}^{s^{k-1}}/_{(k-1)!}|g(\xi + \mathbf{W}_{sT}(\omega) + W^0_{(1-s)T}(\omega))|^2 \in [0,\infty))$, $Y \leftarrow \mathbf{U}$ for $k \in \mathbb{N}$ in the notation of [181, Lemma 2.3]) ensure for all $k \in \mathbb{N}$ that

$$
\begin{aligned}
&\mathbb{E}\left[\tfrac{\mathbf{U}^{k-1}}{(k-1)!}|g(\xi + \mathbf{W}_{\mathbf{U}T} + W^0_{(1-\mathbf{U})T})|^2\right] = \int_0^1 \tfrac{s^{k-1}}{(k-1)!}\,\mathbb{E}\left[|g(\xi + \mathbf{W}_{sT} + W^0_{(1-s)T})|^2\right] \mathrm{d}s \\
&= \tfrac{1}{(k-1)!}\left[\int_0^1 s^{k-1}\,\mathrm{d}s\right]\mathbb{E}\left[|g(\xi + W^0_T)|^2\right] = \tfrac{1}{k!}\,\mathbb{E}\left[|g(\xi + W^0_T)|^2\right].
\end{aligned}
\tag{3.183}
$$

In addition, the fact that $\mathbf{U}$, U^0, $\mathbf{W}$, and W^0 are independent, [181, Lemma 2.3] (with $S \leftarrow [0,1]$, $U \leftarrow ([0,1]\times\Omega \ni (s,\omega) \mapsto {}^{s^{k-1}}/_{(k-1)!}|(1-s)Tf(sT+(1-s)U^0(\omega)T, \xi+\mathbf{W}_{sT}(\omega)+W^0_{(1-s)U^0(\omega)T}(\omega), 0)|^2 \in [0,\infty))$, $Y \leftarrow \mathbf{U}$ for $k \in \mathbb{N}$ in the notation of [181, Lemma 2.3]), and [181, Lemma 2.10] (with $k \leftarrow k$, $U \leftarrow ([0,T] \times \mathbb{R}^d \times \Omega \ni (t,x,\omega) \mapsto f(t,x,0) \in \mathbb{R})$, $\mathfrak{r} \leftarrow U^0$, $\mathbb{W} \leftarrow W^0$ for $k \in \mathbb{N}$ in the notation of [181, Lemma 2.10]) establish for all $k \in \mathbb{N}$ that

$$
\begin{aligned}
&\mathbb{E}\left[\tfrac{\mathbf{U}^{k-1}}{(k-1)!}\big|(1-\mathbf{U})Tf\big(\mathbf{U}T + (1-\mathbf{U})U^0T, \xi + \mathbf{W}_{\mathbf{U}T} + W^0_{(1-\mathbf{U})U^0T}, 0\big)\big|^2\right] \\
&= \int_0^1 \tfrac{s^{k-1}}{(k-1)!}\,\mathbb{E}\left[\big|(1-s)Tf\big(sT + (1-s)U^0T, \xi + \mathbf{W}_{sT} + W^0_{(1-s)U^0T}, 0\big)\big|^2\right] \mathrm{d}s \\
&= \tfrac{1}{T^k}\int_0^T \tfrac{t^{k-1}}{(k-1)!}\,\mathbb{E}\left[\big|(T-t)f\big(t + (T-t)U^0, \xi + \mathbf{W}_t + W^0_{(T-t)U^0}, 0\big)\big|^2\right] \mathrm{d}t \\
&= \tfrac{1}{T^k}\int_0^T \tfrac{t^{k-1}}{(k-1)!}\,\mathbb{E}\left[\big|(T-t)f\big(t + (T-t)U^0, \xi + \mathbf{W}_t + W^0_{t+(T-t)U^0} - W^0_t, 0\big)\big|^2\right] \mathrm{d}t \\
&\leq \tfrac{T^2}{T^{k+1}}\int_0^T \tfrac{t^k}{k!}\,\mathbb{E}\left[|f(t, \xi + \mathbf{W}_t, 0)|^2\right] \mathrm{d}t \leq \tfrac{T}{k!}\int_0^T \mathbb{E}\left[|f(t, \xi + \mathbf{W}_t, 0)|^2\right] \mathrm{d}t \\
&= \tfrac{T}{k!}\int_0^T \mathbb{E}\left[|f(t, \xi + W^0_t, 0)|^2\right] \mathrm{d}t.
\end{aligned}
\tag{3.184}
$$

Combining (3.182) with (3.183)–(3.184) yields for all $k \in \mathbb{N}$ that

$$
\begin{aligned}
&\mathbb{E}\left[|\psi_k(\Phi_0(0,0,Z^0))|^2\right] \\
&\leq \tfrac{1}{k!}\left[\left(\mathbb{E}\left[|g(\xi + W^0_T)|^2\right]\right)^{1/2} + \sqrt{T}\left(\textstyle\int_0^T \mathbb{E}\left[|f(t, \xi + W^0_t, 0)|^2\right] \mathrm{d}t\right)^{1/2}\right]^2 \\
&\leq \tfrac{e^{2LT}}{k!}\left[\left(\mathbb{E}\left[|g(\xi + W^0_T)|^2\right]\right)^{1/2} + \sqrt{T}\left(\textstyle\int_0^T \mathbb{E}\left[|f(t, \xi + W^0_t, 0)|^2\right] \mathrm{d}t\right)^{1/2}\right]^2 = \tfrac{C^2}{k!}.
\end{aligned}
\tag{3.185}
$$

Moreover, (3.181), (3.132), Hölder's inequality, the fact that U^0 and W^0 are independent, and [181, Lemma 2.3] imply that

$$\begin{aligned}
&\mathbb{E}\big[|\psi_0(\Phi_0(0,0,Z^0))|^2\big] = \mathbb{E}\big[|g(\xi+W^0_T)+Tf(U^0T,\xi+W^0_{U^0T},0)|^2\big]\\
&\leq \Big[\big(\mathbb{E}\big[|g(\xi+W^0_T)|^2\big]\big)^{1/2}+\big(T^2\,\mathbb{E}\big[|f(U^0T,\xi+W^0_{U^0T},0)|^2\big]\big)^{1/2}\Big]^2\\
&= \Big[\big(\mathbb{E}\big[|g(\xi+W^0_T)|^2\big]\big)^{1/2}+\big(T^2\textstyle\int_0^1\mathbb{E}\big[|f(sT,\xi+W^0_{sT},0)|^2\big]\,\mathrm{d}s\big)^{1/2}\Big]^2\\
&= \Big[\big(\mathbb{E}\big[|g(\xi+W^0_T)|^2\big]\big)^{1/2}+\big(T\textstyle\int_0^T\mathbb{E}\big[|f(t,\xi+W^0_{t},0)|^2\big]\,\mathrm{d}t\big)^{1/2}\Big]^2\\
&\leq e^{2LT}\Big[\big(\mathbb{E}\big[|g(\xi+W^0_T)|^2\big]\big)^{1/2}+\sqrt{T}\big(\textstyle\int_0^T\mathbb{E}\big[|f(t,\xi+W^0_{t},0)|^2\big]\,\mathrm{d}t\big)^{1/2}\Big]^2 = \tfrac{C^2}{0!}.
\end{aligned} \tag{3.186}$$

The proof of Lemma 3.29 is thus complete. □

Lemma 3.30. *Assume Setting 3.1. Then it holds for all $k\in\mathbb{N}_0$, $n\in\mathbb{N}$, $u,v\in\mathcal{Y}$ that*

$$\mathbb{E}\big[|\psi_k(\Phi_n(u,v,Z^0))|^2\big]\leq (LT)^2\,\mathbb{E}\big[|\psi_{k+1}(u-v)|^2\big]. \tag{3.187}$$

Proof of Lemma 3.30. Throughout this proof let $u,v\in\mathcal{Y}$. Observe that (3.133) shows for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$\begin{aligned}
&|[\Phi_1(u,v,Z^0)](t,x)|\\
&= (T-t)\big|f\big(t+(T-t)U^0,x+W^0_{(T-t)U^0},u\big(t+(T-t)U^0,x+W^0_{(T-t)U^0}\big)\big)\\
&\qquad - f\big(t+(T-t)U^0,x+W^0_{(T-t)U^0},v\big(t+(T-t)U^0,x+W^0_{(T-t)U^0}\big)\big)\big|\\
&\leq L(T-t)\big|u\big(t+(T-t)U^0,x+W^0_{(T-t)U^0}\big)-v\big(t+(T-t)U^0,x+W^0_{(T-t)U^0}\big)\big|\\
&= L\big|(T-t)\cdot[u-v]\big(t+(T-t)U^0,x+W^0_{(T-t)U^0}\big)\big|.
\end{aligned} \tag{3.188}$$

Equation (3.132), the fact that $\mathbf{U}$, U^0, $\mathbf{W}$, and W^0 are independent, [181, Lemma 2.3], and [181, Lemma 2.10] (with $k\leftarrow k$, $U\leftarrow([0,T]\times\mathbb{R}^d\times\Omega\ni(t,x,\omega)\mapsto u(t,x)-v(t,x)\in\mathbb{R})$, $\mathfrak{r}\leftarrow U^0$, $\mathbb{W}\leftarrow W^0$ for $k\in\mathbb{N}$ in the notation of [181, Lemma 2.10]) hence prove for all $k\in\mathbb{N}$ that

$$\begin{aligned}
&\mathbb{E}\big[|\psi_k(\Phi_1(u,v,Z^0))|^2\big] = \mathbb{E}\Big[\tfrac{\mathbf{U}^{k-1}}{(k-1)!}|[\Phi_1(u,v,Z^0)](\mathbf{U}T,\xi+\mathbf{W}_{\mathbf{U}T})|^2\Big]\\
&\leq L^2\,\mathbb{E}\Big[\tfrac{\mathbf{U}^{k-1}}{(k-1)!}\big|(1-\mathbf{U})T\cdot[u-v]\big(\mathbf{U}T+(1-\mathbf{U})U^0T,\xi+\mathbf{W}_{\mathbf{U}T}+W^0_{(1-\mathbf{U})U^0T}\big)\big|^2\Big]\\
&= L^2\int_0^1\tfrac{s^{k-1}}{(k-1)!}\,\mathbb{E}\Big[\big|(1-s)T\cdot[u-v]\big(sT+(1-s)U^0T,\xi+\mathbf{W}_{sT}+W^0_{(1-s)U^0T}\big)\big|^2\Big]\,\mathrm{d}s\\
&= \tfrac{L^2}{T^k}\int_0^T\tfrac{t^{k-1}}{(k-1)!}\,\mathbb{E}\Big[\big|(T-t)\cdot[u-v]\big(t+(T-t)U^0,\xi+\mathbf{W}_t+W^0_{(T-t)U^0}\big)\big|^2\Big]\,\mathrm{d}t\\
&= \tfrac{L^2}{T^k}\int_0^T\tfrac{t^{k-1}}{(k-1)!}\,\mathbb{E}\Big[\big|(T-t)\cdot[u-v]\big(t+(T-t)U^0,\xi+\mathbf{W}_t+W^0_{t+(T-t)U^0}-W^0_t\big)\big|^2\Big]\,\mathrm{d}t\\
&\leq \tfrac{(LT)^2}{T^{k+1}}\int_0^T\tfrac{t^k}{k!}\,\mathbb{E}\big[|[u-v](t,\xi+\mathbf{W}_t)|^2\big]\,\mathrm{d}t\\
&= (LT)^2\int_0^1\tfrac{s^k}{k!}\,\mathbb{E}\big[|[u-v](sT,\xi+\mathbf{W}_{sT})|^2\big]\,\mathrm{d}s\\
&= (LT)^2\,\mathbb{E}\Big[\tfrac{\mathbf{U}^k}{k!}|[u-v](\mathbf{U}T,\xi+\mathbf{W}_{\mathbf{U}T})|^2\Big] = (LT)^2\,\mathbb{E}\big[|\psi_{k+1}(u-v)|^2\big].
\end{aligned} \tag{3.189}$$

In addition, (3.132), (3.188) and the fact that (U^0, W^0) and $(\mathbf{U}, \mathbf{W})$ are identically distributed ensure that

$$\begin{aligned}&\mathbb{E}\big[|\psi_0(\Phi_1(u,v,Z^0))|^2\big] = \mathbb{E}\big[|[\Phi_1(u,v,Z^0)](0,\xi)|^2\big] \\ &\leq (LT)^2\,\mathbb{E}\big[|[u-v](U^0T,\xi+W^0_{U^0T})|^2\big] \\ &= (LT)^2\,\mathbb{E}\big[|[u-v](\mathbf{U}T,\xi+\mathbf{W}_{\mathbf{U}T})|^2\big] = (LT)^2\,\mathbb{E}\big[|\psi_1(u-v)|^2\big].\end{aligned} \tag{3.190}$$

This, (3.189), and the fact that $\forall\, n\in\mathbb{N}\colon \Phi_n = \Phi_1$ complete the proof of Lemma 3.30. □

Lemma 3.31. *Assume Setting 3.1. Then it holds for all $k\in\mathbb{N}_0$, $n,j\in\mathbb{N}$ that*

$$\mathbb{E}\Bigg[\bigg|\psi_k\bigg(y-\sum_{l=0}^{n-1}\mathbb{E}\big[\Phi_l\big(Y^0_{l,j},Y^1_{l-1,j},Z^0\big)\big]\bigg)\bigg|^2\Bigg] \leq (LT)^2\,\mathbb{E}\Big[\big|\psi_{k+1}\big(Y^0_{n-1,j}-y\big)\big|^2\Big]. \tag{3.191}$$

Proof of Lemma 3.31. Throughout this proof let $\Psi_{n,j}\colon [0,T]\times\mathbb{R}^d\to[0,\infty)$, $j\in\mathbb{N}$, $n\in\mathbb{N}_0$, satisfy for all $n,j\in\mathbb{N}$, $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$\Psi_{n-1,j}(t,x) = \mathbb{E}\Big[\big|(T-t)\cdot\big[Y^0_{n-1,j}-y\big]\big(t+(T-t)U^0, x+W^0_{(T-t)U^0}\big)\big|^2\Big] \tag{3.192}$$

(cf. Lemma 3.28). To start with, observe that (3.133), (i)–(ii) in Lemma 3.27, (ii) in Lemma 3.26, and (iii) and (v) in Proposition 3.8 show for all $l,j\in\mathbb{N}$, $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$\begin{aligned}&\mathbb{E}\big[\big[\Phi_l\big(Y^0_{l,j},Y^1_{l-1,j},Z^0\big)\big](T-t,x)\big] \\ &= t\,\mathbb{E}\big[f\big(T-t+U^0t,x+W^0_{U^0t},Y^0_{l,j}\big(T-t+U^0t,x+W^0_{U^0t}\big)\big) \\ &\qquad - f\big(T-t+U^0t,x+W^0_{U^0t},Y^1_{l-1,j}\big(T-t+U^0t,x+W^0_{U^0t}\big)\big)\big] \\ &= t\,\mathbb{E}\big[f\big(T-t+U^0t,x+W^0_{U^0t},Y^0_{l,j}\big(T-t+U^0t,x+W^0_{U^0t}\big)\big) \\ &\qquad - f\big(T-t+U^0t,x+W^0_{U^0t},Y^0_{l-1,j}\big(T-t+U^0t,x+W^0_{U^0t}\big)\big)\big].\end{aligned} \tag{3.193}$$

Again (3.133) hence ensures for all $n,j\in\mathbb{N}$, $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$\begin{aligned}&\Bigg[\sum_{l=0}^{n-1}\mathbb{E}\big[\Phi_l\big(Y^0_{l,j},Y^1_{l-1,j},Z^0\big)\big]\Bigg](T-t,x) \\ &= \sum_{l=0}^{n-1}\mathbb{E}\big[\big[\Phi_l\big(Y^0_{l,j},Y^1_{l-1,j},Z^0\big)\big](T-t,x)\big] \\ &= \mathbb{E}[g(x+W^0_t)] + t\,\mathbb{E}\big[f\big(T-t+U^0t,x+W^0_{U^0t},Y^0_{0,j}\big(T-t+U^0t,x+W^0_{U^0t}\big)\big)\big] \\ &\quad + t\sum_{l=1}^{n-1}\mathbb{E}\big[f\big(T-t+U^0t,x+W^0_{U^0t},Y^0_{l,j}\big(T-t+U^0t,x+W^0_{U^0t}\big)\big) \\ &\qquad\qquad - f\big(T-t+U^0t,x+W^0_{U^0t},Y^0_{l-1,j}\big(T-t+U^0t,x+W^0_{U^0t}\big)\big)\big] \\ &= \mathbb{E}[g(x+W^0_t)] + t\,\mathbb{E}\big[f\big(T-t+U^0t,x+W^0_{U^0t},Y^0_{n-1,j}\big(T-t+U^0t,x+W^0_{U^0t}\big)\big)\big].\end{aligned} \tag{3.194}$$

In addition, (3.127), the fact that $\mathbf{W}$ and W^0 are identically distributed, the fact that W^0 and U^0 are independent, and [181, Lemma 2.4] (with $S\leftarrow[0,1]$, $U\leftarrow([0,1]\times\Omega\ni(u,\omega)\mapsto f(t+(T-t)u,x+W^0_{(T-t)u}(\omega),y(t+(T-t)u,x+W^0_{(T-t)u}(\omega)))\in\mathbb{R})$, $Y\leftarrow U^0$

for $x \in \mathbb{R}^d$, $t \in [0,T]$ in the notation of [181, Lemma 2.4]) imply for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}
y(t,x) &= \mathbb{E}[g(x+\mathbf{W}_{T-t})] + \int_t^T \mathbb{E}\big[f\big(s, x+\mathbf{W}_{s-t}, y(s, x+\mathbf{W}_{s-t})\big)\big]\,\mathrm{d}s \\
&= \mathbb{E}[g(x+W^0_{T-t})] + \int_t^T \mathbb{E}\big[f\big(s, x+W^0_{s-t}, y(s, x+W^0_{s-t})\big)\big]\,\mathrm{d}s \\
&= \mathbb{E}[g(x+W^0_{T-t})] \\
&\quad + (T-t)\int_0^1 \mathbb{E}\big[f\big(t+(T-t)u, x+W^0_{(T-t)u}, y\big(t+(T-t)u, x+W^0_{(T-t)u}\big)\big)\big]\,\mathrm{d}u \\
&= \mathbb{E}[g(x+W^0_{T-t})] \\
&\quad + (T-t)\,\mathbb{E}\big[f\big(t+(T-t)U^0, x+W^0_{(T-t)U^0}, y\big(t+(T-t)U^0, x+W^0_{(T-t)U^0}\big)\big)\big].
\end{aligned} \tag{3.195}$$

This and (3.194) demonstrate for all $n,j \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\left|\left[y - \sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l\big(Y^0_{l,j}, Y^1_{l-1,j}, Z^0\big)\big]\right](t,x)\right| \\
&\leq (T-t)\,\mathbb{E}\big[\big|f\big(t+(T-t)U^0, x+W^0_{(T-t)U^0}, y\big(t+(T-t)U^0, x+W^0_{(T-t)U^0}\big)\big) \\
&\qquad\qquad - f\big(t+(T-t)U^0, x+W^0_{(T-t)U^0}, Y^0_{n-1,j}\big(t+(T-t)U^0, x+W^0_{(T-t)U^0}\big)\big)\big|\big] \\
&\leq L(T-t)\,\mathbb{E}\big[\big|y\big(t+(T-t)U^0, x+W^0_{(T-t)U^0}\big) - Y^0_{n-1,j}\big(t+(T-t)U^0, x+W^0_{(T-t)U^0}\big)\big|\big] \\
&= L(T-t)\,\mathbb{E}\big[\big|\big[Y^0_{n-1,j}-y\big]\big(t+(T-t)U^0, x+W^0_{(T-t)U^0}\big)\big|\big].
\end{aligned} \tag{3.196}$$

Jensen's inequality and (3.192) hence ensure for all $n,j \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\left|\left[y - \sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l\big(Y^0_{l,j}, Y^1_{l-1,j}, Z^0\big)\big]\right](t,x)\right|^2 \\
&\leq L^2(T-t)^2\big(\mathbb{E}\big[\big|\big[Y^0_{n-1,j}-y\big]\big(t+(T-t)U^0, x+W^0_{(T-t)U^0}\big)\big|\big]\big)^2 \\
&\leq L^2(T-t)^2\,\mathbb{E}\Big[\big|\big[Y^0_{n-1,j}-y\big]\big(t+(T-t)U^0, x+W^0_{(T-t)U^0}\big)\big|^2\Big] = L^2\Psi_{n-1,j}(t,x).
\end{aligned} \tag{3.197}$$

Furthermore, (3.192), the fact that it holds for every $n,j \in \mathbb{N}$ that $\mathbf{W}$, $Y^0_{n-1,j}$, U^0, and W^0 are independent (cf. Lemma 3.27 and (ii)–(iii) in Proposition 3.8), and [181, Lemma 2.3] (with $S \leftarrow \mathbb{R}^d$, $U \leftarrow (\mathbb{R}^d \times \Omega \ni (w,\omega) \mapsto |(T-t)\cdot[Y^0_{n-1,j}(\omega)-y](t+(T-t)U^0(\omega), \xi + w + W^0_{(T-t)U^0(\omega)}(\omega))|^2 \in [0,\infty))$, $Y \leftarrow \mathbf{W}_t$ for $t \in [0,T]$, $j,n \in \mathbb{N}$ in the notation of [181, Lemma 2.3]) prove for all $n,j \in \mathbb{N}$, $t \in [0,T]$ that

$$\begin{aligned}
&\mathbb{E}[\Psi_{n-1,j}(t,\xi+\mathbf{W}_t)] = \int_{\mathbb{R}^d} \Psi_{n-1,j}(t,\xi+w)\,\big(\mathbf{W}_t(\mathbb{P})_{\mathscr{B}(\mathbb{R}^d)}\big)(\mathrm{d}w) \\
&= \int_{\mathbb{R}^d} \mathbb{E}\Big[\big|(T-t)\cdot\big[Y^0_{n-1,j}-y\big]\big(t+(T-t)U^0, \xi+w+W^0_{(T-t)U^0}\big)\big|^2\Big]\,\big(\mathbf{W}_t(\mathbb{P})_{\mathscr{B}(\mathbb{R}^d)}\big)(\mathrm{d}w) \\
&= \mathbb{E}\Big[\big|(T-t)\cdot\big[Y^0_{n-1,j}-y\big]\big(t+(T-t)U^0, \xi+\mathbf{W}_t+W^0_{(T-t)U^0}\big)\big|^2\Big] \\
&= \mathbb{E}\Big[\big|(T-t)\cdot\big[Y^0_{n-1,j}-y\big]\big(t+(T-t)U^0, \xi+\mathbf{W}_t+W^0_{t+(T-t)U^0}-W^0_t\big)\big|^2\Big].
\end{aligned} \tag{3.198}$$

Combining (3.132) with (3.197), the fact that $\mathbf{U}$ and $\mathbf{W}$ are independent, [181, Lemma 2.3] (with $S \leftarrow [0,1]$, $U \leftarrow ([0,1]\times\Omega \ni (s,\omega) \mapsto {}^{s^{k-1}}/_{(k-1)!}\,\Psi_{n-1,j}(sT, \xi+\mathbf{W}_{sT}(\omega)) \in [0,\infty))$,

$Y \leftarrow \mathbf{U}$ for $j, n, k \in \mathbb{N}$ in the notation of [181, Lemma 2.3]), (3.198), again the fact that it holds for every $n, j \in \mathbb{N}$ that $\mathbf{W}$, $Y^0_{n-1,j}$, U^0, and W^0 are independent, and [181, Lemma 2.10] (with $k \leftarrow k$, $U \leftarrow ([0,T] \times \mathbb{R}^d \times \Omega \ni (t,x,\omega) \mapsto [Y^0_{n-1,j}(\omega)](t,x) - y(t,x) \in \mathbb{R})$, $\mathfrak{r} \leftarrow U^0$, $\mathbb{W} \leftarrow W^0$ for $j, n, k \in \mathbb{N}$ in the notation of [181, Lemma 2.10]) establishes for all $k, n, j \in \mathbb{N}$ that

$$\begin{aligned}
&\mathbb{E}\Bigg[\bigg|\psi_k\bigg(y - \sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l\big(Y^0_{l,j}, Y^1_{l-1,j}, Z^0\big)\big]\bigg)\bigg|^2\Bigg] \\
&= \mathbb{E}\Bigg[\tfrac{\mathbf{U}^{k-1}}{(k-1)!}\bigg|\bigg[y - \sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l\big(Y^0_{l,j}, Y^1_{l-1,j}, Z^0\big)\big]\bigg](\mathbf{U}T, \xi + \mathbf{W}_{\mathbf{U}T})\bigg|^2\Bigg] \\
&\leq L^2\, \mathbb{E}\Big[\tfrac{\mathbf{U}^{k-1}}{(k-1)!}\Psi_{n-1,j}(\mathbf{U}T, \xi + \mathbf{W}_{\mathbf{U}T})\Big] = L^2 \int_0^1 \tfrac{s^{k-1}}{(k-1)!}\, \mathbb{E}[\Psi_{n-1,j}(sT, \xi + \mathbf{W}_{sT})]\, \mathrm{d}s \\
&= \tfrac{L^2}{T^k} \int_0^T \tfrac{t^{k-1}}{(k-1)!}\, \mathbb{E}[\Psi_{n-1,j}(t, \xi + \mathbf{W}_t)]\, \mathrm{d}t \qquad (3.199)\\
&= \tfrac{L^2}{T^k} \int_0^T \tfrac{t^{k-1}}{(k-1)!}\, \mathbb{E}\Big[\big|(T-t)\cdot\big[Y^0_{n-1,j} - y\big]\big(t + (T-t)U^0, \xi + \mathbf{W}_t + W^0_{t+(T-t)U^0} - W^0_t\big)\big|^2\Big]\, \mathrm{d}t \\
&\leq \tfrac{(LT)^2}{T^{k+1}} \int_0^T \tfrac{t^k}{k!}\, \mathbb{E}\Big[\big|\big[Y^0_{n-1,j} - y\big](t, \xi + \mathbf{W}_t)\big|^2\Big]\, \mathrm{d}t.
\end{aligned}$$

This, the fact that it holds for every $n, j \in \mathbb{N}$ that $Y^0_{n-1,j}$, $\mathbf{W}$, and $\mathbf{U}$ are independent, [181, Lemma 2.3] (with $S \leftarrow [0,1]$, $U \leftarrow ([0,1] \times \Omega \ni (s,\omega) \mapsto {}^{s^k}\!/_{k!}|[Y^0_{n-1,j}(\omega) - y](sT, \xi + \mathbf{W}_{sT}(\omega))|^2 \in [0,\infty))$, $Y \leftarrow \mathbf{U}$ for $j, n, k \in \mathbb{N}$ in the notation of [181, Lemma 2.3]), and (3.132) show for all $k, n, j \in \mathbb{N}$ that

$$\begin{aligned}
&\mathbb{E}\Bigg[\bigg|\psi_k\bigg(y - \sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l\big(Y^0_{l,j}, Y^1_{l-1,j}, Z^0\big)\big]\bigg)\bigg|^2\Bigg] \\
&\leq (LT)^2 \int_0^1 \tfrac{s^k}{k!}\, \mathbb{E}\Big[\big|\big[Y^0_{n-1,j} - y\big](sT, \xi + \mathbf{W}_{sT})\big|^2\Big]\, \mathrm{d}s \qquad (3.200)\\
&= (LT)^2\, \mathbb{E}\Big[\tfrac{\mathbf{U}^k}{k!}\big|\big[Y^0_{n-1,j} - y\big](\mathbf{U}T, \xi + \mathbf{W}_{\mathbf{U}T})\big|^2\Big] = (LT)^2\, \mathbb{E}\Big[\big|\psi_{k+1}\big(Y^0_{n-1,j} - y\big)\big|^2\Big].
\end{aligned}$$

Moreover, (3.132), (3.197), and the fact that it holds for all $n, j \in \mathbb{N}$ that $(Y^0_{n-1,j}, U^0, W^0)$ and $(Y^0_{n-1,j}, \mathbf{U}, \mathbf{W})$ are identically distributed demonstrate for all $n, j \in \mathbb{N}$ that

$$\begin{aligned}
&\mathbb{E}\Bigg[\bigg|\psi_0\bigg(y - \sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l\big(Y^0_{l,j}, Y^1_{l-1,j}, Z^0\big)\big]\bigg)\bigg|^2\Bigg] = \bigg|\bigg[y - \sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l\big(Y^0_{l,j}, Y^1_{l-1,j}, Z^0\big)\big]\bigg](0,\xi)\bigg|^2 \\
&\leq (LT)^2\, \mathbb{E}\Big[\big|\big[Y^0_{n-1,j} - y\big]\big(U^0T, \xi + W^0_{U^0T}\big)\big|^2\Big] \qquad (3.201)\\
&= (LT)^2\, \mathbb{E}\Big[\big|\big[Y^0_{n-1,j} - y\big]\big(\mathbf{U}T, \xi + \mathbf{W}_{\mathbf{U}T}\big)\big|^2\Big] = (LT)^2\, \mathbb{E}\Big[\big|\psi_1\big(Y^0_{n-1,j} - y\big)\big|^2\Big].
\end{aligned}$$

The proof of Lemma 3.31 is thus complete. □

3.2.3 Complexity analysis

3.2.3.1 MLP approximations in fixed space dimensions

Proposition 3.32. *Let* $d \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $T \in (0,\infty)$, $L, p, \mathfrak{B}, \kappa, C \in [0,\infty)$, $\Theta = \bigcup_{n=1}^{\infty} \mathbb{Z}^n$, $(M_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ *satisfy* $\liminf_{j\to\infty} M_j = \infty$, $\sup_{j\in\mathbb{N}} {M_{j+1}}/{M_j} \leq \mathfrak{B}$, *and* $\sup_{j\in\mathbb{N}} {M_j}/{j} \leq \kappa$, *let* $f \in C([0,T]\times\mathbb{R}^d\times\mathbb{R},\mathbb{R})$, $g \in C(\mathbb{R}^d,\mathbb{R})$ *satisfy for all* $t \in [0,T]$, $x \in \mathbb{R}^d$, $v, w \in \mathbb{R}$ *that* $\max\{|f(t,x,0)|,|g(x)|\} \leq L\max\{1,\|x\|_{\mathbb{R}^d}^p\}$ *and* $|f(t,x,v)-f(t,x,w)| \leq L|v-w|$, *let* $(\Omega,\mathscr{F},\mathbb{P})$ *be a probability space, let* $U^\theta\colon \Omega \to [0,1]$, $\theta \in \Theta$, *be independent on* $[0,1]$ *uniformly distributed random variables, let* $W^\theta\colon [0,T]\times\Omega \to \mathbb{R}^d$, $\theta\in\Theta$, *be independent standard Brownian motions with continuous sample paths, assume that* $(U^\theta)_{\theta\in\Theta}$ *and* $(W^\theta)_{\theta\in\Theta}$ *are independent, assume that*

$$C = \max\Big\{1, e^{LT}\Big[\big(\mathbb{E}\big[|g(\xi+W_T^0)|^2\big]\big)^{1/2} + \sqrt{T}\big(\textstyle\int_0^T \mathbb{E}\big[|f(t,\xi+W_t^0,0)|^2\big]\,\mathrm{d}t\big)^{1/2}\Big]\Big\}, \tag{3.202}$$

let $Y_{n,j}^\theta\colon [0,T]\times\mathbb{R}^d\times\Omega \to \mathbb{R}$, $\theta\in\Theta$, $j\in\mathbb{N}$, $n \in (\mathbb{N}_0\cup\{-1\})$, *satisfy for all* $n, j \in \mathbb{N}$, $\theta\in\Theta$, $t\in[0,T]$, $x\in\mathbb{R}^d$ *that* $Y_{-1,j}^\theta(t,x) = Y_{0,j}^\theta(t,x) = 0$ *and*

$$\begin{aligned} Y_{n,j}^\theta(T-t,x) &= \tfrac{1}{(M_j)^n}\Bigg[\sum_{i=1}^{(M_j)^n} g\big(x+W_t^{(\theta,0,i)}\big)\Bigg] + \sum_{l=0}^{n-1}\tfrac{t}{(M_j)^{n-l}}\Bigg[\sum_{i=1}^{(M_j)^{n-l}} \\ &\Big[f\Big(T-t+U^{(\theta,l,i)}t, x+W_{U^{(\theta,l,i)}t}^{(\theta,l,i)}, Y_{l,j}^{(\theta,l,i)}\big(T-t+U^{(\theta,l,i)}t, x+W_{U^{(\theta,l,i)}t}^{(\theta,l,i)}\big)\Big) \\ &- \mathbb{1}_{\mathbb{N}}(l) f\Big(T-t+U^{(\theta,l,i)}t, x+W_{U^{(\theta,l,i)}t}^{(\theta,l,i)}, Y_{l-1,j}^{(\theta,-l,i)}\big(T-t+U^{(\theta,l,i)}t, x+W_{U^{(\theta,l,i)}t}^{(\theta,l,i)}\big)\Big)\Big]\Bigg], \end{aligned} \tag{3.203}$$

and let $(\mathrm{Cost}_{n,j})_{(n,j)\in(\mathbb{N}_0\cup\{-1\})\times\mathbb{N}} \subseteq \mathbb{N}_0$ *satisfy for all* $n, j \in \mathbb{N}$ *that* $\mathrm{Cost}_{-1,j} = \mathrm{Cost}_{0,j} = 0$ *and*

$$\mathrm{Cost}_{n,j} \leq (M_j)^n d + \sum_{l=0}^{n-1}\big[(M_j)^{n-l}(\mathrm{Cost}_{l,j} + \mathrm{Cost}_{l-1,j} + d + 1)\big]. \tag{3.204}$$

Then

(i) there exists a unique at most polynomially growing viscosity solution $y \in C([0,T]\times\mathbb{R}^d,\mathbb{R})$ *of*

$$\big(\tfrac{\partial y}{\partial t}\big)(t,x) + \tfrac{1}{2}(\Delta_x y)(t,x) + f(t,x,y(t,x)) = 0 \tag{3.205}$$

with $y(T,x) = g(x)$ *for* $(t,x) \in (0,T)\times\mathbb{R}^d$,

(ii) it holds for all $n \in \mathbb{N}$ *that*

$$\big(\mathbb{E}\big[|Y_{n,n}^0(0,\xi) - y(0,\xi)|^2\big]\big)^{1/2} \leq C\left[\frac{e^\kappa(1+(2LT)^2)}{M_n}\right]^{n/2} < \infty, \tag{3.206}$$

(iii) it holds for all $n\in\mathbb{N}$ *that* $\mathrm{Cost}_{n,n} \leq (5M_n)^n d$, *and*

(iv) there exists $(N_\varepsilon)_{\varepsilon\in(0,1]} \subseteq \mathbb{N}$ *such that it holds for all* $\varepsilon \in (0,1]$, $\delta\in(0,\infty)$ *that* $\sup_{n\in\{N_\varepsilon,N_\varepsilon+1,\ldots\}}\big(\mathbb{E}\big[|Y_{n,n}^0(0,\xi)-y(0,\xi)|^2\big]\big)^{1/2} \leq \varepsilon$ *and*

$$\mathrm{Cost}_{N_\varepsilon,N_\varepsilon} \leq 5de^\kappa C^{2(1+\delta)}\Big(1+\sup_{n\in\mathbb{N}}\Big[\tfrac{[5\mathfrak{B}e^{2\kappa}(1+(2LT)^2)]^{(1+\delta)}}{(M_n)^\delta}\Big]^n\Big)\varepsilon^{-2(1+\delta)} < \infty. \tag{3.207}$$

Proof of Proposition 3.32. Throughout this proof assume w.l.o.g. that $L > 0$, assume w.l.o.g. that there exist an on $[0,1]$ uniformly distributed random variable $\mathbf{U}\colon \Omega \to [0,1]$ and a standard Brownian motion $\mathbf{W}\colon [0,T]\times\Omega \to \mathbb{R}^d$ with continuous sample paths such that $\mathbf{U}$, $\mathbf{W}$, $(U^\theta)_{\theta\in\Theta}$, and $(W^\theta)_{\theta\in\Theta}$ are independent, let $\mathfrak{z},\gamma \in [0,\infty)$, $c\in(0,\infty)$, $\mathfrak{y}_{-1},\mathfrak{y}_0 \in C([0,T]\times\mathbb{R}^d,\mathbb{R})$ be given by $\mathfrak{z} = d$, $\gamma = 2$, $c = (LT)^2$, and $\mathfrak{y}_{-1} = \mathfrak{y}_0 = 0$, let $q\in(p,\infty)$, let $\mathcal{Y}\subseteq C([0,T]\times\mathbb{R}^d,\mathbb{R})$ be the set given by

$$\mathcal{Y} = \left\{ v\in C([0,T]\times\mathbb{R}^d,\mathbb{R})\colon \limsup_{\mathbb{N}\ni n\to\infty} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d,\, \|x\|_{\mathbb{R}^d}\geq n} \frac{|v(t,x)|}{\|x\|_{\mathbb{R}^d}^q} = 0 \right\}, \tag{3.208}$$

let $\|\cdot\|_{\mathcal{Y}}\colon \mathcal{Y}\to[0,\infty)$ satisfy for all $v\in\mathcal{Y}$ that

$$\|v\|_{\mathcal{Y}} = \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \left[\frac{|v(t,x)|}{\max\{1,\|x\|_{\mathbb{R}^d}^q\}} \right], \tag{3.209}$$

let $(\mathcal{Z},\mathscr{Z}) = \big([0,1]\times C([0,T],\mathbb{R}^d), \mathscr{B}([0,1])\otimes\mathscr{B}(C([0,T],\mathbb{R}^d))\big)$, let $Z^\theta\colon\Omega\to\mathcal{Z}$, $\theta\in\Theta$, satisfy for all $\theta\in\Theta$ that $Z^\theta = (U^\theta,W^\theta)$, let $(\mathcal{H},\langle\cdot,\cdot\rangle_{\mathcal{H}},\|\cdot\|_{\mathcal{H}}) = (\mathbb{R},\langle\cdot,\cdot\rangle_{\mathbb{R}},|\cdot|)$, let $\mathscr{S} = \sigma_{L(\mathcal{Y},\mathcal{H})}\big(\{\{\varphi\in L(\mathcal{Y},\mathcal{H})\colon \varphi(v)\in\mathcal{B}\}\subseteq L(\mathcal{Y},\mathcal{H})\colon v\in\mathcal{Y},\,\mathcal{B}\in\mathscr{B}(\mathcal{H})\}\big)$, let $\psi_k\colon\Omega\to L(\mathcal{Y},\mathcal{H})$, $k\in\mathbb{N}_0$, satisfy for all $k\in\mathbb{N}_0$, $\omega\in\Omega$, $v\in\mathcal{Y}$ that

$$[\psi_k(\omega)](v) = \begin{cases} v(0,\xi) & : k = 0 \\ \sqrt{\frac{(\mathbf{U}(\omega))^{k-1}}{(k-1)!}}\, v\big(\mathbf{U}(\omega)T, \xi + \mathbf{W}_{\mathbf{U}(\omega)T}(\omega)\big) & : k\in\mathbb{N} \end{cases}, \tag{3.210}$$

and let $\Phi_l\colon \mathcal{Y}\times\mathcal{Y}\times\mathcal{Z}\to\mathcal{Y}$, $l\in\mathbb{N}_0$, satisfy for all $l\in\mathbb{N}_0$, $v,w\in\mathcal{Y}$, $\mathfrak{z} = (\mathfrak{u},\mathfrak{w})\in\mathcal{Z}$, $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$\begin{aligned} &[\Phi_l(v,w,\mathfrak{z})](T-t,x) \\ &= \begin{cases} g(x+\mathfrak{w}_t) + tf\big(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t}, v(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t})\big) & : l = 0 \\ t\big[f\big(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t}, v(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t})\big) & \\ \quad - f\big(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t}, w(T-t+\mathfrak{u}t, x+\mathfrak{w}_{\mathfrak{u}t})\big)\big] & : l\in\mathbb{N} \end{cases} \end{aligned} \tag{3.211}$$

(cf. Lemma 3.21 and Corollary 3.23). Note that the assumption that $\forall\, t\in[0,T]$, $x\in\mathbb{R}^d$, $v,w\in\mathbb{R}\colon \big(\max\{|f(t,x,0)|,|g(x)|\}\leq L\max\{1,\|x\|_{\mathbb{R}^d}^p\}$ and $|f(t,x,v)-f(t,x,w)|\leq L|v-w|\big)$ ensures that there exists a unique at most polynomially growing viscosity solution $y\in C([0,T]\times\mathbb{R}^d,\mathbb{R})$ of

$$\big(\tfrac{\partial y}{\partial t}\big)(t,x) + \tfrac{1}{2}(\Delta_x y)(t,x) + f(t,x,y(t,x)) = 0 \tag{3.212}$$

with $y(T,x) = g(x)$ for $(t,x)\in(0,T)\times\mathbb{R}^d$ (cf., e.g., Hairer, Hutzenthaler, & Jentzen [157, Section 4], Hutzenthaler et al. [181, Corollary 3.11], and Beck et al. [24, Theorem 1.1]). This shows (i). Moreover, the Feynman–Kac formula proves for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$y(t,x) = \mathbb{E}\left[g(x+\mathbf{W}_{T-t}) + \int_t^T f\big(s, x+\mathbf{W}_{s-t}, y(s,x+\mathbf{W}_{s-t})\big)\,ds \right] \tag{3.213}$$

(cf., e.g., [157, Section 4], [181, Corollary 3.11], and [24, Theorem 1.1]). Combining this with [181, Corollary 3.11] demonstrates that

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \left[\frac{|y(t,x)|}{\max\{1,\|x\|_{\mathbb{R}^d}^p\}} \right] < \infty. \tag{3.214}$$

Next observe that

- it holds that $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ is a separable $\mathbb{R}$-Banach space (cf. (ii) in Proposition 3.20),
- it holds that $\min\{\mathfrak{B}, \kappa, C\} \geq 1$, $y \in \mathcal{Y}$ (cf. (3.214) and Lemma 3.21), and $\mathfrak{y}_{-1}, \mathfrak{y}_0 \in \mathcal{Y}$,
- it holds that $(\mathcal{Z}, \mathscr{Z})$ is a measurable space,
- it holds that $Z^\theta \colon \Omega \to \mathcal{Z}$, $\theta \in \Theta$, are i.i.d. $\mathscr{F}/\mathscr{Z}$-measurable functions,
- it holds that $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}}, \|\cdot\|_{\mathcal{H}})$ is a separable $\mathbb{R}$-Hilbert space,
- it holds that $\psi_k \colon \Omega \to L(\mathcal{Y}, \mathcal{H})$, $k \in \mathbb{N}_0$, are $\mathscr{F}/\mathscr{S}$-measurable functions (cf. Lemma 3.25),
- it holds that $(Z^\theta)_{\theta \in \Theta}$ and $(\psi_k)_{k \in \mathbb{N}_0}$ are independent,
- it holds that $\Phi_l \colon \mathcal{Y} \times \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y}$, $l \in \mathbb{N}_0$, are $(\mathscr{B}(\mathcal{Y}) \otimes \mathscr{B}(\mathcal{Y}) \otimes \mathscr{Z})/\mathscr{B}(\mathcal{Y})$-measurable functions (cf. (ii) in Lemma 3.26),
- it holds for all $n \in (\mathbb{N}_0 \cup \{-1\})$, $j \in \mathbb{N}$, $\theta \in \Theta$ that $Y^\theta_{n,j}(\Omega) \subseteq \mathcal{Y}$ (cf. assumption (3.203) and (i) in Lemma 3.27),
- it holds for all $n, j \in \mathbb{N}$, $\theta \in \Theta$ that $Y^\theta_{-1,j} = \mathfrak{y}_{-1}$, $Y^\theta_{0,j} = \mathfrak{y}_0$, and
$$Y^\theta_{n,j} = \sum_{l=0}^{n-1} \tfrac{1}{(M_j)^{n-l}} \left[\sum_{i=1}^{(M_j)^{n-l}} \Phi_l\big(Y^{(\theta,l,i)}_{l,j}, Y^{(\theta,-l,i)}_{l-1,j}, Z^{(\theta,l,i)}\big) \right] \tag{3.215}$$
(cf. assumption (3.203) and (ii) in Lemma 3.27),
- it holds for all $n, j \in \mathbb{N}$ that $\mathrm{Cost}_{-1,j} = \mathrm{Cost}_{0,j} = 0$ and
$$\begin{aligned} \mathrm{Cost}_{n,j} &\leq (M_j)^n d + \sum_{l=0}^{n-1} \big[(M_j)^{n-l}(\mathrm{Cost}_{l,j} + \mathrm{Cost}_{l-1,j} + d + 1)\big] \\ &\leq (M_j)^n \mathfrak{z} + \sum_{l=0}^{n-1} \big[(M_j)^{n-l}(\mathrm{Cost}_{l,j} + \mathrm{Cost}_{l-1,j} + \gamma\mathfrak{z})\big] \end{aligned} \tag{3.216}$$
(cf. assumption (3.204)),
- it holds for all $k \in \mathbb{N}_0$, $j \in \mathbb{N}$ that $\mathbb{E}\big[\|\Phi_k(Y^0_{k,j}, Y^1_{k-1,j}, Z^0)\|_{\mathcal{Y}}\big] < \infty$ (cf. Lemma 3.28), and
- it holds for all $k \in \mathbb{N}_0$, $n, j \in \mathbb{N}$, $u, v \in \mathcal{Y}$ that
$$\max\big\{\mathbb{E}\big[\|\psi_k(\Phi_0(\mathfrak{y}_0, \mathfrak{y}_{-1}, Z^0))\|^2_{\mathcal{H}}\big], \mathbb{1}_{\mathbb{N}}(k)\, \mathbb{E}\big[\|\psi_k(\mathfrak{y}_0 - y)\|^2_{\mathcal{H}}\big]\big\} \leq \tfrac{C^2}{k!}, \tag{3.217}$$
$$\mathbb{E}\big[\|\psi_k(\Phi_n(u, v, Z^0))\|^2_{\mathcal{H}}\big] \leq c\, \mathbb{E}\big[\|\psi_{k+1}(u - v)\|^2_{\mathcal{H}}\big], \tag{3.218}$$
$$\mathbb{E}\left[\left\|\psi_k\Big(y - \sum_{l=0}^{n-1} \mathbb{E}\big[\Phi_l(Y^0_{l,j}, Y^1_{l-1,j}, Z^0)\big]\Big)\right\|^2_{\mathcal{H}}\right] \leq 2c\, \mathbb{E}\Big[\big\|\psi_{k+1}\big(Y^0_{n-1,j} - y\big)\big\|^2_{\mathcal{H}}\Big] \tag{3.219}$$
(cf. (3.213), assumption (3.202), Lemma 3.29, Lemma 3.30, and Lemma 3.31).

Corollary 3.15 hence establishes (ii)–(iv). The proof of Proposition 3.32 is thus complete. □

3.2.3.2 MLP approximations in variable space dimensions

Theorem 3.33. *Let* $T \in (0,\infty)$, $K, L, p, \mathfrak{B}, \kappa \in [0,\infty)$, $\Theta = \bigcup_{n=1}^{\infty} \mathbb{Z}^n$, $(M_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ *satisfy* $\liminf_{j\to\infty} M_j = \infty$, $\sup_{j\in\mathbb{N}} {}^{M_{j+1}}/_{M_j} \leq \mathfrak{B}$, *and* $\sup_{j\in\mathbb{N}} {}^{M_j}/_j \leq \kappa$, *let* $\xi_d \in \mathbb{R}^d$, $d \in \mathbb{N}$, *satisfy* $\sup_{d\in\mathbb{N}} \|\xi_d\|_{\mathbb{R}^d} \leq K$, *for every* $d \in \mathbb{N}$ *let* $f_d \in C([0,T] \times \mathbb{R}^d \times \mathbb{R}, \mathbb{R})$, $g_d \in C(\mathbb{R}^d, \mathbb{R})$ *satisfy for all* $t \in [0,T]$, $x \in \mathbb{R}^d$, $v, w \in \mathbb{R}$ *that* $\max\{|f_d(t,x,0)|, |g_d(x)|\} \leq L\max\{1, \|x\|_{\mathbb{R}^d}^p\}$ *and* $|f_d(t,x,v) - f_d(t,x,w)| \leq L|v-w|$, *let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space, let* $U^\theta \colon \Omega \to [0,1]$, $\theta \in \Theta$, *be independent on* $[0,1]$ *uniformly distributed random variables, for every* $d \in \mathbb{N}$ *let* $W^{d,\theta} \colon [0,T] \times \Omega \to \mathbb{R}^d$, $\theta \in \Theta$, *be independent standard Brownian motions, assume for every* $d \in \mathbb{N}$ *that* $(U^\theta)_{\theta\in\Theta}$ *and* $(W^{d,\theta})_{\theta\in\Theta}$ *are independent, let* $Y^{d,\theta}_{n,j} \colon [0,T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}$, $\theta \in \Theta$, $d, j \in \mathbb{N}$, $n \in (\mathbb{N}_0 \cup \{-1\})$, *satisfy for all* $n, j, d \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ *that* $Y^{d,\theta}_{-1,j}(t,x) = Y^{d,\theta}_{0,j}(t,x) = 0$ *and*

$$
\begin{aligned}
Y^{d,\theta}_{n,j}(T-t,x) &= \tfrac{1}{(M_j)^n}\Bigg[\sum_{i=1}^{(M_j)^n} g_d\big(x + W^{d,(\theta,0,i)}_t\big)\Bigg] + \sum_{l=0}^{n-1} \tfrac{t}{(M_j)^{n-l}}\Bigg[\sum_{i=1}^{(M_j)^{n-l}} \\
&\Big[f_d\Big(T-t+U^{(\theta,l,i)}t, x + W^{d,(\theta,l,i)}_{U^{(\theta,l,i)}t}, Y^{d,(\theta,l,i)}_{l,j}\big(T-t+U^{(\theta,l,i)}t, x + W^{d,(\theta,l,i)}_{U^{(\theta,l,i)}t}\big)\Big) \\
&- \mathbb{1}_{\mathbb{N}}(l) f_d\Big(T-t+U^{d,(\theta,l,i)}t, x + W^{d,(\theta,l,i)}_{U^{(\theta,l,i)}t}, Y^{d,(\theta,-l,i)}_{l-1,j}\big(T-t+U^{(\theta,l,i)}t, x + W^{d,(\theta,l,i)}_{U^{(\theta,l,i)}t}\big)\Big)\Big]\Bigg],
\end{aligned}
\tag{3.220}
$$

and let $(\mathrm{Cost}_{d,n,j})_{(d,n,j)\in\mathbb{N}\times(\mathbb{N}_0\cup\{-1\})\times\mathbb{N}} \subseteq \mathbb{N}_0$ *satisfy for all* $d, n, j \in \mathbb{N}$ *that* $\mathrm{Cost}_{d,-1,j} = \mathrm{Cost}_{d,0,j} = 0$ *and*

$$
\mathrm{Cost}_{d,n,j} \leq (M_j)^n d + \sum_{l=0}^{n-1} \big[(M_j)^{n-l}(\mathrm{Cost}_{d,l,j} + \mathrm{Cost}_{d,l-1,j} + d + 1)\big]. \tag{3.221}
$$

Then

(i) for every $d \in \mathbb{N}$ *there exists a unique at most polynomially growing viscosity solution* $y_d \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$ *of*

$$
\big(\tfrac{\partial y_d}{\partial t}\big)(t,x) + \tfrac{1}{2}(\Delta_x y_d)(t,x) + f_d(t,x,y_d(t,x)) = 0 \tag{3.222}
$$

with $y_d(T,x) = g_d(x)$ *for* $(t,x) \in (0,T) \times \mathbb{R}^d$ *and*

(ii) there exists $(N_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbb{N}$ *such that it holds for all* $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ *that* $\sup_{n\in\{N_{d,\varepsilon}, N_{d,\varepsilon}+1,\ldots\}} \big(\mathbb{E}\big[|Y^{d,0}_{n,n}(0,\xi_d) - y_d(0,\xi_d)|^2\big]\big)^{1/2} \leq \varepsilon$ *and*

$$
\begin{aligned}
\mathrm{Cost}_{d,N_{d,\varepsilon},N_{d,\varepsilon}} \leq \Big[&[4^{p+2}\max\{L,1\}(1+T)^{p/2+1}e^{LT}(\max\{K,p,1\})^p]^{2(1+\delta)}e^\kappa \\
&\cdot \Big(1 + \sup_{n\in\mathbb{N}}\Big[\tfrac{[5\mathfrak{B}e^{2\kappa}(1+(2LT)^2)]^{(1+\delta)}}{(M_n)^\delta}\Big]^n\Big)\Big] d^{1+p(1+\delta)}\varepsilon^{-2(1+\delta)} < \infty.
\end{aligned}
\tag{3.223}
$$

Proof of Theorem 3.33. Throughout this proof assume w.l.o.g. for every $d \in \mathbb{N}$ that $W^{d,\theta} \colon [0,T] \times \Omega \to \mathbb{R}^d$, $\theta \in \Theta$, are independent standard Brownian motions with continuous sample paths (cf., e.g., Klenke [202, Definition 21.8]) and throughout this proof let $C_d \in [1,\infty)$, $d \in \mathbb{N}$, be the real numbers which satisfy for all $d \in \mathbb{N}$ that

$$
C_d = \max\Big\{1, e^{LT}\Big[\big(\mathbb{E}\big[|g_d(\xi_d + W^{d,0}_T)|^2\big]\big)^{1/2} + \sqrt{T}\big(\textstyle\int_0^T \mathbb{E}\big[|f_d(t,\xi_d + W^{d,0}_t, 0)|^2\big]\,\mathrm{d}t\big)^{1/2}\Big]\Big\}. \tag{3.224}
$$

First of all, observe that the Burkholder–Davis–Gundy-type inequality in Da Prato & Zabczyk [94, Lemma 7.7] establishes for all $r \in [2, \infty)$, $d \in \mathbb{N}$, $t \in [0, T]$ that

$$\big(\mathbb{E}\big[\|W_t^{d,0}\|_{\mathbb{R}^d}^r\big]\big)^{1/r} \leq \sqrt{\tfrac{1}{2}r(r-1)td} \leq r\sqrt{\tfrac{Td}{2}}. \tag{3.225}$$

Jensen's inequality and the fact that $\forall\, a, b, r \in [0, \infty)\colon (a+b)^r \leq 2^{\max\{r-1,0\}}(a^r + b^r)$ hence prove for all $d \in \mathbb{N}$ that

$$\begin{aligned}
&\big(\mathbb{E}\big[|g_d(\xi_d + W_T^{d,0})|^2\big]\big)^{1/2} + \sqrt{T}\big(\textstyle\int_0^T \mathbb{E}\big[|f_d(t, \xi_d + W_t^{d,0}, 0)|^2\big]\,\mathrm{d}t\big)^{1/2} \\
&\leq L\big(\mathbb{E}\big[\max\{1, \|\xi_d + W_T^{d,0}\|_{\mathbb{R}^d}^{2p}\}\big]\big)^{1/2} + L\sqrt{T}\big(\textstyle\int_0^T \mathbb{E}\big[\max\{1, \|\xi_d + W_t^{d,0}\|_{\mathbb{R}^d}^{2p}\}\big]\,\mathrm{d}t\big)^{1/2} \\
&\leq L\Big(1 + \big(\mathbb{E}\big[\|\xi_d + W_T^{d,0}\|_{\mathbb{R}^d}^{2p}\big]\big)^{1/2}\Big) + LT\big(1 + \tfrac{1}{T}\textstyle\int_0^T \mathbb{E}\big[\|\xi_d + W_t^{d,0}\|_{\mathbb{R}^d}^{2p}\big]\,\mathrm{d}t\big)^{1/2} \\
&\leq L(1+T) + L\Big(\big(\mathbb{E}\big[\|\xi_d + W_T^{d,0}\|_{\mathbb{R}^d}^{2\max\{p,1\}}\big]\big)^{\frac{1}{2\max\{p,1\}}}\Big)^p \\
&\quad + LT\Big(\big(\tfrac{1}{T}\textstyle\int_0^T \mathbb{E}\big[\|\xi_d + W_t^{d,0}\|_{\mathbb{R}^d}^{2\max\{p,1\}}\big]\,\mathrm{d}t\big)^{\frac{1}{2\max\{p,1\}}}\Big)^p \\
&\leq L(1+T) + L\Big(\|\xi_d\|_{\mathbb{R}^d} + \big(\mathbb{E}\big[\|W_T^{d,0}\|_{\mathbb{R}^d}^{2\max\{p,1\}}\big]\big)^{\frac{1}{2\max\{p,1\}}}\Big)^p \\
&\quad + LT\Big(\|\xi_d\|_{\mathbb{R}^d} + \big(\tfrac{1}{T}\textstyle\int_0^T \mathbb{E}\big[\|W_t^{d,0}\|_{\mathbb{R}^d}^{2\max\{p,1\}}\big]\,\mathrm{d}t\big)^{\frac{1}{2\max\{p,1\}}}\Big)^p \\
&\leq L(1+T) + L\big(\|\xi_d\|_{\mathbb{R}^d} + \max\{p,1\}\sqrt{2Td}\big)^p + LT\big(\|\xi_d\|_{\mathbb{R}^d} + \max\{p,1\}\sqrt{2Td}\big)^p \\
&\leq L(1+T) + L(1+T)2^{\max\{p-1,0\}}\big(\|\xi_d\|_{\mathbb{R}^d}^p + \max\{p^p,1\}(2Td)^{p/2}\big) \\
&\leq d^{p/2}2^{\max\{p,1\}+p/2}L(1+T)^{p/2+1}\big(K^p + \max\{p^p,1\}\big) \\
&\leq d^{p/2}4^{p+1}L(1+T)^{p/2+1}(\max\{K,p,1\})^p.
\end{aligned} \tag{3.226}$$

This and (3.224) show for all $d \in \mathbb{N}$, $\delta \in (0, \infty)$ that

$$\begin{aligned}
5(C_d)^{2(1+\delta)} &\leq 5[d^{p/2}4^{p+1}\max\{L,1\}(1+T)^{p/2+1}e^{LT}(\max\{K,p,1\})^p]^{2(1+\delta)} \\
&\leq [4^{p+2}\max\{L,1\}(1+T)^{p/2+1}e^{LT}(\max\{K,p,1\})^p]^{2(1+\delta)}d^{p(1+\delta)}.
\end{aligned} \tag{3.227}$$

Combining this with Proposition 3.32 completes the proof of Theorem 3.33. □

Corollary 3.34. *Let $T \in (0, \infty)$, $p \in [0, \infty)$, $\Theta = \bigcup_{n=1}^\infty \mathbb{Z}^n$, $(M_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$, $(\xi_d)_{d\in\mathbb{N}} \subseteq \mathbb{R}^d$ satisfy $\sup_{j\in\mathbb{N}}(M_{j+1}/M_j + M_j/j + \|\xi_j\|_{\mathbb{R}^j}) < \infty = \liminf_{j\to\infty} M_j$, let $f\colon \mathbb{R} \to \mathbb{R}$ be a Lipschitz continuous function, let $g_d \in C(\mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, satisfy $\sup_{d\in\mathbb{N}, x\in\mathbb{R}^d} {|g_d(x)|}/{\max\{1, \|x\|_{\mathbb{R}^d}^p\}} < \infty$, let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $U^\theta\colon \Omega \to [0,1]$, $\theta \in \Theta$, be independent on $[0,1]$ uniformly distributed random variables, let $W^{d,\theta}\colon [0,T] \times \Omega \to \mathbb{R}^d$, $\theta \in \Theta$, $d \in \mathbb{N}$, be independent standard Brownian motions, assume that $(U^\theta)_{\theta\in\Theta}$ and $(W^{d,\theta})_{(d,\theta)\in\mathbb{N}\times\Theta}$ are independent, let $Y_{n,j}^{d,\theta}\colon [0,T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}$, $\theta \in \Theta$, $d, j \in \mathbb{N}$, $n \in (\mathbb{N}_0 \cup \{-1\})$, satisfy for all $n, j, d \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that $Y_{-1,j}^{d,\theta}(t,x) = Y_{0,j}^{d,\theta}(t,x) = 0$ and*

$$\begin{aligned}
Y_{n,j}^{d,\theta}(T-t,x) &= \sum_{l=0}^{n-1} \tfrac{t}{(M_j)^{n-l}} \Bigg[\sum_{i=1}^{(M_j)^{n-l}} \Big[f\Big(Y_{l,j}^{d,(\theta,l,i)}\big(T-t+U^{(\theta,l,i)}t, x + W_{U^{(\theta,l,i)}t}^{d,(\theta,l,i)}\big)\Big) \\
&\quad - \mathbb{1}_{\mathbb{N}}(l) f\Big(Y_{l-1,j}^{d,(\theta,-l,i)}\big(T-t+U^{(\theta,l,i)}t, x + W_{U^{(\theta,l,i)}t}^{d,(\theta,l,i)}\big)\Big)\Big]\Bigg] + \tfrac{1}{(M_j)^n}\Bigg[\sum_{i=1}^{(M_j)^n} g_d\big(x + W_t^{d,(\theta,0,i)}\big)\Bigg],
\end{aligned} \tag{3.228}$$

and let $(\mathrm{Cost}_{d,n,j})_{(d,n,j)\in\mathbb{N}\times(\mathbb{N}_0\cup\{-1\})\times\mathbb{N}} \subseteq \mathbb{N}_0$ *satisfy for all* $d,n,j \in \mathbb{N}$ *that* $\mathrm{Cost}_{d,-1,j} = \mathrm{Cost}_{d,0,j} = 0$ *and*

$$\mathrm{Cost}_{d,n,j} \leq (M_j)^n d + \sum_{l=0}^{n-1} \big[(M_j)^{n-l}(\mathrm{Cost}_{d,l,j} + \mathrm{Cost}_{d,l-1,j} + d + 1)\big]. \tag{3.229}$$

Then

(i) *for every* $d \in \mathbb{N}$ *there exists a unique at most polynomially growing viscosity solution* $y_d \in C([0,T]\times\mathbb{R}^d,\mathbb{R})$ *of*

$$\big(\tfrac{\partial y_d}{\partial t}\big)(t,x) + \tfrac{1}{2}(\Delta_x y_d)(t,x) + f(y_d(t,x)) = 0 \tag{3.230}$$

with $y_d(T,x) = g_d(x)$ *for* $(t,x) \in (0,T)\times\mathbb{R}^d$ *and*

(ii) *there exist* $(N_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbb{N}$ *and* $(C_\delta)_{\delta\in(0,\infty)} \subseteq (0,\infty)$ *such that it holds for all* $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ *that* $\mathrm{Cost}_{d,N_{d,\varepsilon},N_{d,\varepsilon}} \leq C_\delta d^{1+p(1+\delta)}\varepsilon^{-2(1+\delta)}$ *and*

$$\sup_{n\in\{N_{d,\varepsilon},N_{d,\varepsilon}+1,\ldots\}} \big(\mathbb{E}\big[|Y^{d,0}_{n,n}(0,\xi_d) - y_d(0,\xi_d)|^2\big]\big)^{1/2} \leq \varepsilon. \tag{3.231}$$

Chapter 4

Solving high-dimensional optimal stopping problems using deep learning

The content of this chapter is a slightly modified extract of the preprint Becker et al. [30].

In this chapter we propose a deep learning based algorithm for solving general possibly high-dimensional optimal stopping problems (cf. Section 1.3 in Chapter 1). Step by step we present the derivation and implementation of the algorithm, that is described in Framework 4.2 in Subsection 4.2.2, and provide more details of our approach summarised in (I)–(III) in Section 1.3. In addition, we report a large number of numerical experiments, which demonstrate that the algorithm is highly effective for solving high-dimensional optimal stopping problems, in terms of both accuracy and speed.

This chapter is organised in the following way. In Section 4.1 we present the main ideas from which the proposed algorithm is derived. More specifically, in Subsection 4.1.1 we illustrate how an optimal stopping problem in the context of American option pricing is typically formulated. Thereafter, a replacement of this continuous time problem by a corresponding discrete time optimal stopping problem is discussed by means of an example in Subsection 4.1.2. Subsection 4.1.3 is devoted to the statement and proof of an elementary, but crucial result about factorising general discrete stopping times in terms of compositions of measurable functions (cf. Lemma 4.2), which lies at the heart of the neural network architecture we propose in Subsection 4.1.4 to approximate general discrete stopping times. This construction, in turn, is exploited in Subsection 4.1.5 to transform the discrete time optimal stopping problem from Subsection 4.1.2 into the search of a maximum of a suitable objective function (cf. (I) in Section 1.3). In Subsection 4.1.6 we suggest to employ stochastic gradient ascent-type optimisation algorithms to find approximate maximum points of the objective function (cf. (II) in Section 1.3). As a last step, we explain in Subsection 4.1.7 how we calculate final approximations for both the American option price as well as an optimal exercise strategy (cf. (III) in Section 1.3). In Section 4.2 we introduce the proposed algorithm in a concise way, first for a special case for the sake of clarity (cf. Subsection 4.2.1) and second in more generality so that, in particular, a rigorous description of our implementations is fully covered (cf. Subsections 4.2.2 and 4.2.3). Following this, in Section 4.3 first a few theoretical results are presented (cf. Subsection 4.3.1), which are used to design numerical example problems and to provide reference values. Thereafter, we describe in detail a large number of example problems, on which our proposed algorithm has been tested, and present numerical results for each of these problems. In particular, the examples include the optimal stopping of Brownian motions (cf. Subsection 4.3.3.1), the pricing of certain exotic American geometric average put and call-type options (cf. Subsection 4.3.3.2), the pricing of Bermudan max-call options in up to 5000 dimensions (cf. Subsection 4.3.4.1), the pricing of an American strangle spread basket option in five dimensions (cf. Subsection 4.3.4.2), the pricing of an American put basket option in Dupire's local volatility model in five dimensions (cf. Subsection 4.3.4.3), and the pricing of an exotic path-dependent financial derivative of a single underlying, which is modelled as a 100-dimensional optimal stopping problem (cf. Subsection 4.3.4.4). The numerical results for the examples in Subsections 4.3.3.1.2, 4.3.3.2.1, 4.3.3.2.2, 4.3.3.2.3, and 4.3.4.1.3 are compared to calculated reference values that can be easily obtained due to the specific design of the considered optimal stopping problem. Moreover, the examples in Subsections 4.3.3.2.2, 4.3.4.1.1, 4.3.4.1.3, 4.3.4.2, 4.3.4.3, and 4.3.4.4 are taken from the literature and our corresponding numerical results are compared to reference values from the literature (where available).

4.1 Main ideas of the proposed algorithm

In this section we outline the main ideas that lead to the formulation of the proposed algorithm in Subsections 4.2.1 and 4.2.2 by considering the example of pricing an American option. The proposed algorithm in Framework 4.2 in Subsection 4.2.2 is, however, general enough to also be applied to optimal stopping problems where there are no specific assumptions on the dynamics of the underlying stochastic process, as long as it can be cheaply simulated (cf. Subsection 4.2.3). Furthermore, often in practice and, in particular, in the case of Bermudan option pricing (cf. many of the examples in Section 4.3) the optimal stopping problem of interest is not a continuous time problem but is already formulated in discrete time. In such a situation there is no need for a time discretisation, as described in Subsection 4.1.2 below, and the proposed algorithm in Framework 4.2 can be applied directly.

4.1.1 The American option pricing problem

Let $T \in (0,\infty)$, $d \in \mathbb{N} = \{1,2,3,\ldots\}$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $\mathscr{F} = (\mathscr{F}_t)_{t\in[0,T]}$ that satisfies the usual conditions (cf., e.g., [199, Definition 2.25 in Section 1.2]), let $\xi\colon \Omega \to \mathbb{R}^d$ be an $\mathscr{F}_0/\mathcal{B}(\mathbb{R}^d)$-measurable function which satisfies for all $p \in (0,\infty)$ that $\mathbb{E}\big[\|\xi\|_{\mathbb{R}^d}^p\big] < \infty$, let $W\colon [0,T]\times\Omega \to \mathbb{R}^d$ be a standard $(\Omega,\mathcal{F},\mathbb{P},\mathscr{F})$-Brownian motion with continuous sample paths, let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be Lipschitz continuous functions, let $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ be an $\mathscr{F}$-adapted continuous solution process of the stochastic differential equation

$$dX_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \qquad X_0 = \xi, \qquad t \in [0,T], \tag{4.1}$$

let $\mathbb{F} = (\mathbb{F}_t)_{t\in[0,T]}$ be the filtration generated by X, and let $g\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ be a continuous and at most polynomially growing function. We think of X as a model for the price processes of d underlyings (say, d stock prices) under the risk-neutral pricing measure $\mathbb{P}$ (cf., e.g., Kallsen [198]) and we are then interested in approximatively pricing the

American option on the process $(X_t)_{t\in[0,T]}$ with the discounted pay-off function $g\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$, that is, we intend to compute the real number

$$\sup\Big\{\mathbb{E}\big[g(\tau, X_\tau)\big] : {\scriptstyle \tau\colon \Omega\to[0,T] \text{ is an} \atop \mathbb{F}\text{-stopping time}}\Big\}. \tag{4.2}$$

In addition to the *price* of the American option in the model (4.1) there is also a high demand from the financial engineering industry to compute an approximately *optimal exercise strategy*, that is, to compute a stopping time which approximately reaches the supremum in (4.2).

In a very simple example of (4.1)–(4.2), we can think of an *American put option* in the one-dimensional Black–Scholes model, in which there are an interest rate $r\in\mathbb{R}$, a dividend rate $\delta\in[0,\infty)$, a volatility $\beta\in(0,\infty)$, and a strike price $K\in(0,\infty)$ such that it holds for all $x\in\mathbb{R}$, $t\in[0,T]$ that $d=1$, $\mu(x)=(r-\delta)\,x$, $\sigma(x)=\beta\,x$, and $g(t,x)=e^{-rt}\max\{K-x,0\}$.

4.1.2 Temporal discretisation

To derive the proposed approximation algorithm we first apply the Euler–Maruyama scheme to the stochastic differential equation (4.1) (cf. (4.5)–(4.6) below) and we employ a suitable time discretisation for the optimal stopping problem (4.2). For this let $N\in\mathbb{N}$ be a natural number and let $t_0,t_1,\dots,t_N\in[0,T]$ be real numbers with

$$0=t_0<t_1<\dots<t_N=T \tag{4.3}$$

(such that the maximal mesh size $\max_{n\in\{0,1,\dots,N-1\}}(t_{n+1}-t_n)$ is sufficiently small). Observe that (4.1) ensures that for all $n\in\{0,1,\dots,N-1\}$ it holds $\mathbb{P}$-a.s. that

$$X_{t_{n+1}} = X_{t_n} + \int_{t_n}^{t_{n+1}} \mu(X_s)\,ds + \int_{t_n}^{t_{n+1}} \sigma(X_s)\,dW_s. \tag{4.4}$$

Note that (4.4) suggests for every $n\in\{0,1,\dots,N-1\}$ that

$$X_{t_{n+1}} \approx X_{t_n} + \mu(X_{t_n})\,(t_{n+1}-t_n) + \sigma(X_{t_n})\,\big(W_{t_{n+1}}-W_{t_n}\big). \tag{4.5}$$

The approximation scheme associated to (4.5) is referred to as the Euler–Maruyama scheme in the literature (cf., e.g., Maruyama [239] and Kloeden & Platen [203]). More formally, let $\mathcal{X}\colon\{0,1,\dots,N\}\times\Omega\to\mathbb{R}^d$ be the stochastic process which satisfies for all $n\in\{0,1,\dots,N-1\}$ that $\mathcal{X}_0=\xi$ and

$$\mathcal{X}_{n+1} = \mathcal{X}_n + \mu(\mathcal{X}_n)\,(t_{n+1}-t_n) + \sigma(\mathcal{X}_n)\,\big(W_{t_{n+1}}-W_{t_n}\big) \tag{4.6}$$

and let $\mathfrak{F}=(\mathfrak{F}_n)_{n\in\{0,1,\dots,N\}}$ be the filtration generated by $\mathcal{X}$. Combining this with (4.5) suggests the approximation

$$\sup\Big\{\mathbb{E}\big[g(t_\tau,\mathcal{X}_\tau)\big] : {\scriptstyle \tau\colon\Omega\to\{0,1,\dots,N\} \text{ is an} \atop \mathfrak{F}\text{-stopping time}}\Big\} \approx \sup\Big\{\mathbb{E}\big[g(\tau,X_\tau)\big] : {\scriptstyle \tau\colon\Omega\to[0,T] \text{ is an} \atop \mathbb{F}\text{-stopping time}}\Big\} \tag{4.7}$$

for the price (4.2) of the American option in Subsection 4.1.1. Below we employ, in particular, (4.7) to derive the proposed approximation algorithm.

4.1.3 Factorisation lemma for stopping times

The derivation of the proposed approximation algorithm is in parts based on an elementary reformulation of time-discrete stopping times (cf. the left hand side of (4.7) above) in terms of measurable functions that appropriately characterise the behaviour of the stopping time; cf. (4.10) and (4.9) in Lemma 4.2 below. The proof of Lemma 4.2 employs the following well-known factorisation result, Lemma 4.1. Lemma 4.1 follows, e.g., from Klenke [202, Corollary 1.97].

Lemma 4.1 (Factorisation lemma)**.** *Let $(S, \mathcal{S})$ be a measurable space, let Ω be a set, let $B \in \mathcal{B}(\mathbb{R} \cup \{-\infty, \infty\})$, and let $X : \Omega \to S$ and $Y : \Omega \to B$ be functions. Then it holds that Y is $\{X^{-1}(A) : A \in \mathcal{S}\}/\mathcal{B}(B)$-measurable if and only if there exists an $\mathcal{S}/\mathcal{B}(B)$-measurable function $f : S \to B$ such that*

$$Y = f \circ X. \tag{4.8}$$

We are now ready to present the above mentioned Lemma 4.2. This elementary lemma is a consequence of Lemma 4.1 above.

Lemma 4.2 (Factorisation lemma for stopping times)**.** *Let $d, N \in \mathbb{N}$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{X} : \{0, 1, \dots, N\} \times \Omega \to \mathbb{R}^d$ be a stochastic process, and let $\mathbb{F} = (\mathbb{F}_n)_{n \in \{0,1,\dots,N\}}$ be the filtration generated by $\mathcal{X}$. Then*

(i) for all Borel measurable functions $\mathbb{U}_n : (\mathbb{R}^d)^{n+1} \to \{0, 1\}$, $n \in \{0, 1, \dots, N\}$, with $\forall\, x_0, x_1, \dots, x_N \in \mathbb{R}^d : \sum_{n=0}^{N} \mathbb{U}_n(x_0, x_1, \dots, x_n) = 1$ it holds that the function

$$\Omega \ni \omega \mapsto \sum_{n=0}^{N} n\, \mathbb{U}_n\big(\mathcal{X}_0(\omega), \mathcal{X}_1(\omega), \dots, \mathcal{X}_n(\omega)\big) \in \{0, 1, \dots, N\} \tag{4.9}$$

is an $\mathbb{F}$-stopping time and

(ii) for every $\mathbb{F}$-stopping time $\tau : \Omega \to \{0, 1, \dots, N\}$ there exist Borel measurable functions $\mathbb{U}_n : (\mathbb{R}^d)^{n+1} \to \{0, 1\}$, $n \in \{0, 1, \dots, N\}$, which satisfy $\forall\, x_0, x_1, \dots, x_N \in \mathbb{R}^d : \sum_{n=0}^{N} \mathbb{U}_n(x_0, x_1, \dots, x_n) = 1$ and

$$\tau = \sum_{n=0}^{N} n\, \mathbb{U}_n(\mathcal{X}_0, \mathcal{X}_1, \dots, \mathcal{X}_n). \tag{4.10}$$

Proof of Lemma 4.2. Note that for all Borel measurable functions $\mathbb{U}_n : (\mathbb{R}^d)^{n+1} \to \{0, 1\}$, $n \in \{0, 1, \dots, N\}$, with $\forall\, x_0, x_1, \dots, x_N \in \mathbb{R}^d : \sum_{n=0}^{N} \mathbb{U}_n(x_0, x_1, \dots, x_n) = 1$ and all $k \in \{0, 1, \dots, N\}$ it holds that

$$\begin{aligned}
&\left\{\omega \in \Omega : \sum_{n=0}^{N} n\, \mathbb{U}_n\big(\mathcal{X}_0(\omega), \mathcal{X}_1(\omega), \dots, \mathcal{X}_n(\omega)\big) = k\right\} \\
&= \big\{\omega \in \Omega : \mathbb{U}_k\big(\mathcal{X}_0(\omega), \mathcal{X}_1(\omega), \dots, \mathcal{X}_k(\omega)\big) = 1\big\} \\
&= \left\{\omega \in \Omega : (\mathcal{X}_0(\omega), \mathcal{X}_1(\omega), \dots, \mathcal{X}_k(\omega)) \in \underbrace{(\mathbb{U}_k)^{-1}(\{1\})}_{\in \mathcal{B}((\mathbb{R}^d)^{k+1})}\right\} \in \mathbb{F}_k.
\end{aligned} \tag{4.11}$$

This establishes (i). It thus remains to prove (ii). For this let $\tau\colon \Omega \to \{0,1,\dots,N\}$ be an $\mathbb{F}$-stopping time. Observe that for every function $\varrho\colon \Omega \to \{0,1,\dots,N\}$ and every $\omega \in \Omega$ it holds that

$$\varrho(\omega) = \sum_{n=0}^{N} n \, \mathbb{1}_{\{\varrho=n\}}(\omega). \tag{4.12}$$

Next note that for every $n \in \{0,1,\dots,N\}$ it holds that the function

$$\Omega \ni \omega \mapsto \mathbb{1}_{\{\tau=n\}}(\omega) \in \{0,1\} \tag{4.13}$$

is $\mathbb{F}_n/\mathcal{B}(\{0,1\})$-measurable. This and the fact that

$$\forall\, n \in \{0,1,\dots,N\}\colon \sigma_\Omega\big((\mathcal{X}_0,\mathcal{X}_1,\dots,\mathcal{X}_n)\big) = \mathbb{F}_n \tag{4.14}$$

ensures that for every $n \in \{0,1,\dots,N\}$ it holds that the function

$$\Omega \ni \omega \mapsto \mathbb{1}_{\{\tau=n\}}(\omega) \in \{0,1\} \tag{4.15}$$

is $\sigma_\Omega((\mathcal{X}_0,\mathcal{X}_1,\dots,\mathcal{X}_n))/\mathcal{B}(\{0,1\})$-measurable. Lemma 4.1 hence demonstrates that there exist Borel measurable functions $\mathbb{V}_n\colon (\mathbb{R}^d)^{n+1} \to \{0,1\}$, $n \in \{0,1,\dots,N\}$, which satisfy for all $n \in \{0,1,\dots,N\}$, $\omega \in \Omega$ that

$$\mathbb{1}_{\{\tau=n\}}(\omega) = \mathbb{V}_n\big(\mathcal{X}_0(\omega),\mathcal{X}_1(\omega),\dots,\mathcal{X}_n(\omega)\big). \tag{4.16}$$

Next let $\mathbb{U}_n\colon (\mathbb{R}^d)^{n+1} \to \mathbb{R}$, $n \in \{0,1,\dots,N\}$, be the functions which satisfy for all $n \in \{0,1,\dots,N\}$, $x_0,x_1,\dots,x_n \in \mathbb{R}^d$ that

$$\begin{aligned}&\mathbb{U}_n(x_0,x_1,\dots,x_n)\\ &= \max\{\mathbb{V}_n(x_0,x_1,\dots,x_n), n+1-N\}\left[1 - \sum_{k=0}^{n-1} \mathbb{U}_k(x_0,x_1,\dots,x_k)\right].\end{aligned} \tag{4.17}$$

Observe that (4.17), in particular, ensures that for all $x_0,x_1,\dots,x_N \in \mathbb{R}^d$ it holds that

$$\mathbb{U}_N(x_0,x_1,\dots,x_N) = \left[1 - \sum_{k=0}^{N-1} \mathbb{U}_k(x_0,x_1,\dots,x_k)\right]. \tag{4.18}$$

Hence, we obtain that for all $x_0,x_1,\dots,x_N \in \mathbb{R}^d$ it holds that

$$\sum_{k=0}^{N} \mathbb{U}_k(x_0,x_1,\dots,x_k) = 1. \tag{4.19}$$

In addition, note that (4.17) assures that for all $x_0 \in \mathbb{R}^d$ it holds that

$$\mathbb{U}_0(x_0) = \mathbb{V}_0(x_0). \tag{4.20}$$

Induction, the fact that

$$\forall\, n \in \{0,1,\dots,N\},\ x_0,x_1,\dots,x_n \in \mathbb{R}^d\colon \mathbb{V}_n(x_0,x_1,\dots,x_n) \in \{0,1\}, \tag{4.21}$$

and (4.17) hence demonstrate that for all $n \in \{0, 1, \dots, N\}$, $x_0, x_1, \dots, x_n \in \mathbb{R}^d$ it holds that

$$\left\{ \mathbb{U}_0(x_0), \mathbb{U}_1(x_0, x_1), \dots, \mathbb{U}_n(x_0, x_1, \dots, x_n), \sum_{k=0}^{n} \mathbb{U}_k(x_0, x_1, \dots, x_k) \right\} \subseteq \{0, 1\}. \tag{4.22}$$

Moreover, note that (4.17), induction, and the fact that the functions $\mathbb{V}_n \colon (\mathbb{R}^d)^{n+1} \to \{0, 1\}$, $n \in \{0, 1, \dots, N\}$, are Borel measurable ensure that for every $n \in \{0, 1, \dots, N\}$ it holds that the function

$$(\mathbb{R}^d)^{n+1} \ni (x_0, x_1, \dots, x_n) \mapsto \mathbb{U}_n(x_0, x_1, \dots, x_n) \in \{0, 1\} \tag{4.23}$$

is also Borel measurable. In the next step we observe that (4.20), (4.17), (4.21), and induction assure that for all $n \in \{0, 1, \dots, N\}$, $x_0, x_1, \dots, x_n \in \mathbb{R}^d$ with $n + 1 - N \leq \sum_{k=0}^{n} \mathbb{V}_k(x_0, x_1, \dots, x_k) \leq 1$ it holds that

$$\forall\, k \in \{0, 1, \dots, n\} \colon \mathbb{U}_k(x_0, x_1, \dots, x_k) = \mathbb{V}_k(x_0, x_1, \dots, x_k). \tag{4.24}$$

In addition, note that (4.16) shows that for all $\omega \in \Omega$ it holds that

$$\sum_{k=0}^{N} \mathbb{V}_k\big(\mathcal{X}_0(\omega), \mathcal{X}_1(\omega), \dots, \mathcal{X}_k(\omega)\big) = \sum_{k=0}^{N} \mathbb{1}_{\{\tau=k\}}(\omega) = 1. \tag{4.25}$$

This, (4.24), and again (4.16) imply that for all $k \in \{0, 1, \dots, N\}$, $\omega \in \Omega$ it holds that

$$\mathbb{U}_k\big(\mathcal{X}_0(\omega), \mathcal{X}_1(\omega), \dots, \mathcal{X}_k(\omega)\big) = \mathbb{V}_k\big(\mathcal{X}_0(\omega), \mathcal{X}_1(\omega), \dots, \mathcal{X}_k(\omega)\big) = \mathbb{1}_{\{\tau=k\}}(\omega). \tag{4.26}$$

Equation (4.12) hence proves that for all $\omega \in \Omega$ it holds that

$$\tau(\omega) = \sum_{n=0}^{N} n\, \mathbb{U}_n\big(\mathcal{X}_0(\omega), \mathcal{X}_1(\omega), \dots, \mathcal{X}_n(\omega)\big). \tag{4.27}$$

Combining this with (4.19) and (4.23) establishes (ii). The proof of Lemma 4.2 is thus complete. ☐

4.1.4 Neural network architectures for stopping times

In the next step we employ multilayer neural network approximations for the functions $\mathbb{U}_n \colon (\mathbb{R}^d)^{n+1} \to \{0, 1\}$, $n \in \{0, 1, \dots, N\}$, in the factorisation lemma, Lemma 4.2 above. In the following we refer to these functions as 'stopping time factors'. Consider again the setting in Subsections 4.1.1 and 4.1.2, for every $\mathfrak{F}$-stopping time $\tau \colon \Omega \to \{0, 1, \dots, N\}$ let $\mathbb{U}_{n,\tau} \colon (\mathbb{R}^d)^{n+1} \to \{0, 1\}$, $n \in \{0, 1, \dots, N\}$, be Borel measurable functions which satisfy $\forall\, x_0, x_1, \dots, x_N \in \mathbb{R}^d \colon \sum_{n=0}^{N} \mathbb{U}_{n,\tau}(x_0, x_1, \dots, x_n) = 1$ and

$$\tau = \sum_{n=0}^{N} n\, \mathbb{U}_{n,\tau}(\mathcal{X}_0, \mathcal{X}_1, \dots, \mathcal{X}_n) \tag{4.28}$$

(cf. (ii) in Lemma 4.2), let $\nu \in \mathbb{N}$ be a sufficiently large natural number, and for every $n \in \{0, 1, \dots, N\}$, $\theta \in \mathbb{R}^\nu$ let $u_{n,\theta} \colon \mathbb{R}^d \to (0, 1)$ and $U_{n,\theta} \colon (\mathbb{R}^d)^{n+1} \to (0, 1)$ be Borel measurable functions which satisfy for all $x_0, x_1, \dots, x_n \in \mathbb{R}^d$ that

$$U_{n,\theta}(x_0, x_1, \dots, x_n) = \max\{u_{n,\theta}(x_n), n + 1 - N\} \left[1 - \sum_{k=0}^{n-1} U_{k,\theta}(x_0, x_1, \dots, x_k) \right] \tag{4.29}$$

(cf. (4.17) above). Observe that for all $\theta \in \mathbb{R}^\nu$, $x_0, x_1, \dots, x_N \in \mathbb{R}^d$ it holds that

$$\sum_{n=0}^{N} U_{n,\theta}(x_0, x_1, \dots, x_n) = 1. \tag{4.30}$$

We think of $\nu \in \mathbb{N}$ as the number of parameters in the employed artificial neural networks and for every appropriate $\mathfrak{F}$-stopping time $\tau\colon \Omega \to \{0, 1, \dots, N\}$ we think of the functions $U_{n,\theta}\colon (\mathbb{R}^d)^{n+1} \to (0,1)$ for $n \in \{0, 1, \dots, N\}$ and suitable $\theta \in \mathbb{R}^\nu$ as appropriate approximations for the stopping time factors $\mathbb{U}_{n,\tau}\colon (\mathbb{R}^d)^{n+1} \to \{0,1\}$, $n \in \{0, 1, \dots, N\}$. Due to (4.30) for every $\theta \in \mathbb{R}^\nu$ the stochastic process

$$\{0, 1, \dots, N\} \times \Omega \ni (n, \omega) \mapsto U_{n,\theta}\big(\mathcal{X}_0(\omega), \mathcal{X}_1(\omega), \dots, \mathcal{X}_n(\omega)\big) \in (0,1) \tag{4.31}$$

can also be viewed as an in an appropriate sense 'randomised' stopping time (cf., e.g., [291, Definition 1 in Subsection 3.1] and, e.g., [125, Section 1.1]). Furthermore, since $\mathcal{X}\colon \{0, 1, \dots, N\} \times \Omega \to \mathbb{R}^d$ is a Markov process, for every $n \in \{0, 1, \dots, N\}$ we only consider functions $u_{n,\theta}\colon \mathbb{R}^d \to (0,1)$, $\theta \in \mathbb{R}^\nu$, which are defined on $\mathbb{R}^d$ instead of $(\mathbb{R}^d)^{n+1}$ and which in (4.29) only depend on $x_n \in \mathbb{R}^d$ instead of $(x_0, x_1, \dots, x_n) \in (\mathbb{R}^d)^{n+1}$ (cf. (4.29) and (4.17) above and [28, Theorem 1 and Remark 2 in Subsection 2.1]). We suggest to choose the functions $u_{n,\theta}\colon \mathbb{R}^d \to (0,1)$, $\theta \in \mathbb{R}^\nu$, $n \in \{0, 1, \dots, N-1\}$, as multilayer feedforward neural networks (cf. [28, Corollary 5 in Subsection 2.2] and, e.g., [16, 92, 175]). For example, for every $k \in \mathbb{N}$ let $\mathcal{L}_k\colon \mathbb{R}^k \to \mathbb{R}^k$ be the function which satisfies for all $x = (x_1, \dots, x_k) \in \mathbb{R}^k$ that

$$\mathcal{L}_k(x) = \left(\frac{\exp(x_1)}{\exp(x_1)+1}, \frac{\exp(x_2)}{\exp(x_2)+1}, \dots, \frac{\exp(x_k)}{\exp(x_k)+1}\right), \tag{4.32}$$

for every $\theta = (\theta_1, \dots, \theta_\nu) \in \mathbb{R}^\nu$, $v \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$, $k, l \in \mathbb{N}$ with $v + k(l+1) \leq \nu$ let $A^{\theta,v}_{k,l}\colon \mathbb{R}^l \to \mathbb{R}^k$ be the affine linear function which satisfies for all $x = (x_1, \dots, x_l) \in \mathbb{R}^l$ that

$$A^{\theta,v}_{k,l}(x) = \begin{pmatrix} \theta_{v+1} & \theta_{v+2} & \dots & \theta_{v+l} \\ \theta_{v+l+1} & \theta_{v+l+2} & \dots & \theta_{v+2l} \\ \theta_{v+2l+1} & \theta_{v+2l+2} & \dots & \theta_{v+3l} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{v+(k-1)l+1} & \theta_{v+(k-1)l+2} & \dots & \theta_{v+kl} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_l \end{pmatrix} + \begin{pmatrix} \theta_{v+kl+1} \\ \theta_{v+kl+2} \\ \theta_{v+kl+3} \\ \vdots \\ \theta_{v+kl+k} \end{pmatrix}, \tag{4.33}$$

and assume for all $n \in \{0, 1, \dots, N-1\}$, $\theta \in \mathbb{R}^\nu$ that $\nu \geq N(2d+1)(d+1)$ and

$$u_{n,\theta} = \mathcal{L}_1 \circ A^{\theta,(2nd+n)(d+1)}_{1,d} \circ \mathcal{L}_d \circ A^{\theta,(2nd+n+1)(d+1)}_{d,d} \circ \mathcal{L}_d \circ A^{\theta,((2n+1)d+n+1)(d+1)}_{d,d}. \tag{4.34}$$

The functions in (4.34) provide artificial neural networks with 4 layers (1 input layer with d neurons, 2 hidden layers each with d neurons, and 1 output layer with 1 neuron) and the multidimensional version of the standard logistic function $\mathbb{R} \ni x \mapsto \exp(x)/(\exp(x)+1) \in (0,1)$ (cf. (4.32) above) as activation functions. In our numerical simulations in Section 4.3 we use this type of activation function only just in front of the output layer and we employ instead the multidimensional version of the rectifier function $\mathbb{R} \ni x \mapsto \max\{x, 0\} \in [0, \infty)$ as activation functions just in front of the hidden layers. But in order to keep the illustration here as short as possible we only employ the multidimensional version of the standard logistic function as activation functions in (4.32)–(4.34) above. Furthermore, note that in contrast to the choice of the functions $u_{n,\theta}\colon \mathbb{R}^d \to (0,1)$, $\theta \in \mathbb{R}^\nu$, $n \in \{0,1,\ldots,N-1\}$, the choice of the functions $u_{N,\theta}\colon \mathbb{R}^d \to (0,1)$, $\theta \in \mathbb{R}^\nu$, has no influence on the approximate stopping time factors $U_{n,\theta}\colon (\mathbb{R}^d)^{n+1} \to (0,1)$, $\theta \in \mathbb{R}^\nu$, $n \in \{0,1,\ldots,N\}$ (cf. (4.29) above).

4.1.5 Formulation of the objective function

Recall that we intend to compute the real number

$$\sup\left\{\mathbb{E}\big[g(t_\tau,\mathcal{X}_\tau)\big] : \substack{\tau\colon \Omega\to\{0,1,\ldots,N\} \text{ is an} \\ \mathfrak{F}\text{-stopping time}}\right\} \tag{4.35}$$

as an approximation of the American option price (4.2) (cf. (4.7) in Subsection 4.1.2). By employing neural network architectures for stopping times (cf. Subsection 4.1.4 above), we next propose to replace the search over all $\mathfrak{F}$-stopping times for finding the supremum in (4.35) by a search over the artificial neural network parameters $\theta \in \mathbb{R}^\nu$ (cf. (4.38) below). For this, observe that (4.28) implies for all $\mathfrak{F}$-stopping times $\tau\colon \Omega \to \{0,1,\ldots,N\}$ and all $n \in \{0,1,\ldots,N\}$ that

$$\mathbb{1}_{\{\tau=n\}} = \mathbb{U}_{n,\tau}(\mathcal{X}_0,\mathcal{X}_1,\ldots,\mathcal{X}_n). \tag{4.36}$$

Therefore, for all $\mathfrak{F}$-stopping times $\tau\colon \Omega \to \{0,1,\ldots,N\}$ it holds that

$$\begin{aligned} g(t_\tau,\mathcal{X}_\tau) &= \left[\sum_{n=0}^{N} \mathbb{1}_{\{\tau=n\}}\right] g(t_\tau,\mathcal{X}_\tau) = \sum_{n=0}^{N} \mathbb{1}_{\{\tau=n\}}\, g(t_n,\mathcal{X}_n) \\ &= \sum_{n=0}^{N} \mathbb{U}_{n,\tau}(\mathcal{X}_0,\mathcal{X}_1,\ldots,\mathcal{X}_n)\, g(t_n,\mathcal{X}_n). \end{aligned} \tag{4.37}$$

Combining this with (i) in Lemma 4.2 and (4.30) inspires the approximation

$$\begin{aligned} &\sup\left\{\mathbb{E}\big[g(t_\tau,\mathcal{X}_\tau)\big] : \substack{\tau\colon \Omega\to\{0,1,\ldots,N\} \text{ is an} \\ \mathfrak{F}\text{-stopping time}}\right\} \\ &= \sup\left\{\mathbb{E}\left[\sum_{n=0}^{N} \mathbb{U}_{n,\tau}(\mathcal{X}_0,\mathcal{X}_1,\ldots,\mathcal{X}_n)\, g(t_n,\mathcal{X}_n)\right] : \substack{\tau\colon \Omega\to\{0,1,\ldots,N\} \text{ is an} \\ \mathfrak{F}\text{-stopping time}}\right\} \\ &= \sup\left\{\mathbb{E}\left[\sum_{n=0}^{N} \mathbb{V}_n(\mathcal{X}_0,\mathcal{X}_1,\ldots,\mathcal{X}_n)\, g(t_n,\mathcal{X}_n)\right] : \substack{\mathbb{V}_n\colon (\mathbb{R}^d)^{n+1}\to\{0,1\},\, n\in\{0,1,\ldots,N\}, \\ \text{are Borel measurable functions with} \\ \forall\, x_0,x_1,\ldots,x_N\in\mathbb{R}^d\colon\, \sum_{n=0}^{N} \mathbb{V}_n(x_0,x_1,\ldots,x_n)=1}\right\} \\ &= \sup\left\{\mathbb{E}\left[\sum_{n=0}^{N} \mathfrak{V}_n(\mathcal{X}_0,\mathcal{X}_1,\ldots,\mathcal{X}_n)\, g(t_n,\mathcal{X}_n)\right] : \substack{\mathfrak{V}_n\colon (\mathbb{R}^d)^{n+1}\to[0,1],\, n\in\{0,1,\ldots,N\}, \\ \text{are Borel measurable functions with} \\ \forall\, x_0,x_1,\ldots,x_N\in\mathbb{R}^d\colon\, \sum_{n=0}^{N} \mathfrak{V}_n(x_0,x_1,\ldots,x_n)=1}\right\} \\ &= \sup\left\{\mathbb{E}\left[\sum_{n=0}^{N} \mathfrak{U}_n(\mathcal{X}_0,\mathcal{X}_1,\ldots,\mathcal{X}_n)\, g(t_n,\mathcal{X}_n)\right] : \substack{\mathfrak{U}_n\colon (\mathbb{R}^d)^{n+1}\to(0,1),\, n\in\{0,1,\ldots,N\}, \\ \text{are Borel measurable functions with} \\ \forall\, x_0,x_1,\ldots,x_N\in\mathbb{R}^d\colon\, \sum_{n=0}^{N} \mathfrak{U}_n(x_0,x_1,\ldots,x_n)=1}\right\} \\ &\approx \sup\left\{\mathbb{E}\left[\sum_{n=0}^{N} U_{n,\theta}(\mathcal{X}_0,\mathcal{X}_1,\ldots,\mathcal{X}_n)\, g(t_n,\mathcal{X}_n)\right] : \theta \in \mathbb{R}^\nu\right\}. \end{aligned} \tag{4.38}$$

In view of this, our numerical solution for approximatively computing (4.35) consists of trying to find an approximate maximiser of the objective function

$$\mathbb{R}^\nu \ni \theta \mapsto \mathbb{E}\left[\sum_{n=0}^{N} U_{n,\theta}(\mathcal{X}_0, \mathcal{X}_1, \ldots, \mathcal{X}_n)\, g(t_n, \mathcal{X}_n)\right] \in \mathbb{R}. \tag{4.39}$$

4.1.6 Stochastic gradient ascent optimisation algorithms

Local/global maxima of the objective function (4.39) can be approximately reached by maximising the expectation of the random objective function

$$\mathbb{R}^\nu \times \Omega \ni (\theta, \omega) \mapsto \sum_{n=0}^{N} U_{n,\theta}(\mathcal{X}_0(\omega), \mathcal{X}_1(\omega), \ldots, \mathcal{X}_n(\omega))\, g(t_n, \mathcal{X}_n(\omega)) \in \mathbb{R} \tag{4.40}$$

by means of a stochastic gradient ascent-type optimisation algorithm. This yields a sequence of random parameter vectors along which we expect the objective function (4.39) to increase. More formally, applying under suitable hyptheses stochastic gradient ascent-type optimisation algorithms to (4.39) results in random approximations

$$\Theta_m = (\Theta_m^{(1)}, \ldots, \Theta_m^{(\nu)}) \colon \Omega \to \mathbb{R}^\nu \tag{4.41}$$

for $m \in \{0, 1, 2, \ldots\}$ of the local/global maximum points of the objective function (4.39), where $m \in \{0, 1, 2, \ldots\}$ is the number of steps of the employed stochastic gradient ascent-type optimisation algorithm.

4.1.7 Price and optimal exercise time for American-style options

The approximation algorithm sketched in Subsection 4.1.6 above allows us to approximatively compute both the *price* and an *optimal exercise strategy* for the American option (cf. Subsection 4.1.1). Let $M \in \mathbb{N}$ and consider a realisation $\widehat{\Theta}_M \in \mathbb{R}^\nu$ of the random variable $\Theta_M \colon \Omega \to \mathbb{R}^\nu$. Then for sufficiently large $N, \nu, M \in \mathbb{N}$ a candidate for a suitable approximation of the American option price is the real number

$$\mathbb{E}\left[\sum_{n=0}^{N} U_{n,\widehat{\Theta}_M}(\mathcal{X}_0, \mathcal{X}_1, \ldots, \mathcal{X}_n)\, g(t_n, \mathcal{X}_n)\right] \tag{4.42}$$

and a candidate for a suitable approximation of an optimal exercise strategy for the American option is the function

$$\Omega \ni \omega \mapsto \sum_{n=0}^{N} n\, U_{n,\widehat{\Theta}_M}(\mathcal{X}_0(\omega), \mathcal{X}_1(\omega), \ldots, \mathcal{X}_n(\omega)) \in [0, N]. \tag{4.43}$$

Note, however, that in general the function (4.43) does not take values in $\{0, 1, \ldots, N\}$ and hence is not a proper stopping time. Similarly, note that in general it is not clear whether there exists an exercise strategy such that the number (4.42) is equal to the expected discounted pay-off under this exercise strategy. For these reasons we suggest other candidates for suitable approximations of the price and an optimal exercise strategy for the American option. More specifically, for every $\theta \in \mathbb{R}^\nu$ let $\tau_\theta \colon \Omega \to \{0, 1, \ldots, N\}$ be the $\mathfrak{F}$-stopping time given by

$$\tau_\theta = \min\left\{ n \in \{0, 1, \ldots, N\} \colon \sum_{k=0}^{n} U_{k,\theta}(\mathcal{X}_0, \ldots, \mathcal{X}_k) \geq 1 - U_{n,\theta}(\mathcal{X}_0, \ldots, \mathcal{X}_n) \right\}. \tag{4.44}$$

Then for sufficiently large $N, \nu, M \in \mathbb{N}$ we use a suitable Monte Carlo approximation of the real number

$$\mathbb{E}\Big[g\big(t_{\tau_{\widehat{\Theta}_M}}, \mathcal{X}_{\tau_{\widehat{\Theta}_M}}\big)\Big] \tag{4.45}$$

as a suitable implementable approximation of the price of the American option (cf. (4.2) in Subsection 4.1.1 above and (4.58) in Subsection 4.2.1 below) and we use the random variable

$$\tau_{\widehat{\Theta}_M} \colon \Omega \to \{0, 1, \ldots, N\} \tag{4.46}$$

as a suitable implementable approximation of an optimal exercise strategy for the American option. Note that one has

$$\begin{aligned} \tau_{\widehat{\Theta}_M} &= \min\left\{ n \in \{0, 1, \ldots, N\} \colon U_{n,\widehat{\Theta}_M}(\mathcal{X}_0, \ldots, \mathcal{X}_n) \geq 1 - \sum_{k=0}^{n} U_{k,\widehat{\Theta}_M}(\mathcal{X}_0, \ldots, \mathcal{X}_k) \right\} \\ &= \min\left\{ n \in \{0, 1, \ldots, N\} \colon U_{n,\widehat{\Theta}_M}(\mathcal{X}_0, \ldots, \mathcal{X}_n) \geq \sum_{k=n+1}^{N} U_{k,\widehat{\Theta}_M}(\mathcal{X}_0, \ldots, \mathcal{X}_k) \right\}. \end{aligned} \tag{4.47}$$

This shows that the exercise strategy $\tau_{\widehat{\Theta}_M} \colon \Omega \to \{0, 1, \ldots, N\}$ exercises at the first time index $n \in \{0, 1, \ldots, N\}$ for which the approximate stopping time factor associated to the mesh point t_n is at least as large as the combined approximate stopping time factors associated to all later mesh points. Furthermore, observe that it holds that

$$\mathbb{E}\Big[g\big(t_{\tau_{\widehat{\Theta}_M}}, \mathcal{X}_{\tau_{\widehat{\Theta}_M}}\big)\Big] \leq \sup\left\{ \mathbb{E}\big[g(t_\tau, \mathcal{X}_\tau)\big] \colon \substack{\tau \colon \Omega \to \{0,1,\ldots,N\} \text{ is an} \\ \mathfrak{F}\text{-stopping time}} \right\}. \tag{4.48}$$

Roughly speaking, this illustrates that Monte Carlo approximations of the number (4.45) are typically low-biased approximations for the American option price (4.2). Finally, we point out that, in comparison with the deep learning based approximation method for solving optimal stopping problems in Becker, Cheridito, & Jentzen [28], the parameters $\widehat{\Theta}_M \in \mathbb{R}^\nu$ determining an approximate optimal exercise strategy (cf. (4.46) above) are obtained using a single training procedure to approximately maximise a single objective function (cf. (4.39) above) and not found recursively through a sequence of training procedures along with different random objective functions (cf. [28, Subsections 2.2 and 2.3]).

4.2 Details of the proposed algorithm

4.2.1 Formulation of the proposed algorithm in a special case

In this subsection we describe the proposed algorithm in the specific situation where the objective is to solve the American option pricing problem described in Subsection 4.1.1, where *batch normalisation* (cf. Ioffe & Szegedy [184]) is not employed in the proposed algorithm, and where the plain vanilla stochastic gradient ascent approximation method with a constant learning rate $\gamma \in (0,\infty)$ and without mini-batches is the employed stochastic approximation algorithm. The general framework, which includes the setting in this subsection as a special case, can be found in Subsection 4.2.2 below.

Framework 4.1 (Specific case)**.** *Let $T, \gamma \in (0,\infty)$, $d, N \in \mathbb{N}$, $\nu = N(2d+1)(d+1)$, let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$, $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$, and $g\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ be Borel measurable functions, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $\xi^m\colon \Omega \to \mathbb{R}^d$, $m \in \mathbb{N}$, be independent random variables, let $W^m\colon [0,T]\times\Omega \to \mathbb{R}^d$, $m\in\mathbb{N}$, be independent $\mathbb{P}$-standard Brownian motions with continuous sample paths, assume that $(\xi^m)_{m\in\mathbb{N}}$ and $(W^m)_{m\in\mathbb{N}}$ are independent, let $t_0, t_1, \ldots, t_N \in [0,T]$ be real numbers with $0 = t_0 < t_1 < \ldots < t_N = T$, let $\mathcal{X}^m\colon \{t_0,t_1,\ldots,t_N\}\times\Omega \to \mathbb{R}^d$, $m\in\mathbb{N}$, be the stochastic processes which satisfy for all $m\in\mathbb{N}$, $n\in\{0,1,\ldots,N-1\}$ that $\mathcal{X}^m_{t_0} = \xi^m$ and*

$$\mathcal{X}^m_{t_{n+1}} = \mathcal{X}^m_{t_n} + \mu\big(\mathcal{X}^m_{t_n}\big)(t_{n+1}-t_n) + \sigma\big(\mathcal{X}^m_{t_n}\big)\big(W^m_{t_{n+1}} - W^m_{t_n}\big), \tag{4.49}$$

for every $k\in\mathbb{N}$ let $\mathcal{L}_k\colon \mathbb{R}^k \to \mathbb{R}^k$ be the function which satisfies for all $x = (x_1,\ldots,x_k) \in \mathbb{R}^k$ that

$$\mathcal{L}_k(x) = \left(\frac{\exp(x_1)}{\exp(x_1)+1}, \frac{\exp(x_2)}{\exp(x_2)+1}, \ldots, \frac{\exp(x_k)}{\exp(x_k)+1}\right), \tag{4.50}$$

for every $\theta = (\theta_1,\ldots,\theta_\nu)\in\mathbb{R}^\nu$, $v\in\mathbb{N}_0$, $k,l\in\mathbb{N}$ with $v + k(l+1) \le \nu$ let $A^{\theta,v}_{k,l}\colon \mathbb{R}^l \to \mathbb{R}^k$ be the function which satisfies for all $x = (x_1,\ldots,x_l)\in\mathbb{R}^l$ that

$$A^{\theta,v}_{k,l}(x) = \left(\theta_{v+kl+1} + \left[\sum_{i=1}^{l} x_i\,\theta_{v+i}\right], \ldots, \theta_{v+kl+k} + \left[\sum_{i=1}^{l} x_i\,\theta_{v+(k-1)l+i}\right]\right), \tag{4.51}$$

for every $\theta\in\mathbb{R}^\nu$ let $u_{n,\theta}\colon \mathbb{R}^d \to (0,1)$, $n\in\{0,1,\ldots,N\}$, be functions which satisfy for all $n\in\{0,1,\ldots,N-1\}$ that

$$u_{n,\theta} = \mathcal{L}_1 \circ A^{\theta,(2nd+n)(d+1)}_{1,d} \circ \mathcal{L}_d \circ A^{\theta,(2nd+n+1)(d+1)}_{d,d} \circ \mathcal{L}_d \circ A^{\theta,((2n+1)d+n+1)(d+1)}_{d,d}, \tag{4.52}$$

for every $n\in\{0,1,\ldots,N\}$, $\theta\in\mathbb{R}^\nu$ let $U_{n,\theta}\colon (\mathbb{R}^d)^{n+1} \to (0,1)$ be the function which satisfies for all $x_0, x_1, \ldots, x_n \in \mathbb{R}^d$ that

$$U_{n,\theta}(x_0,x_1,\ldots,x_n) = \max\{u_{n,\theta}(x_n), n+1-N\}\left[1 - \sum_{k=0}^{n-1} U_{k,\theta}(x_0,x_1,\ldots,x_k)\right], \tag{4.53}$$

for every $m\in\mathbb{N}$ let $\phi^m\colon \mathbb{R}^\nu\times\Omega \to \mathbb{R}$ be the function which satisfies for all $\theta\in\mathbb{R}^\nu$, $\omega\in\Omega$ that

$$\phi^m(\theta,\omega) = \sum_{n=0}^{N}\Big[U_{n,\theta}\big(\mathcal{X}^m_{t_0}(\omega), \mathcal{X}^m_{t_1}(\omega), \ldots, \mathcal{X}^m_{t_n}(\omega)\big)\, g\big(t_n, \mathcal{X}^m_{t_n}(\omega)\big)\Big], \tag{4.54}$$

for every $m\in\mathbb{N}$ let $\Phi^m\colon \mathbb{R}^\nu\times\Omega \to \mathbb{R}^\nu$ be the function which satisfies for all $\theta\in\mathbb{R}^\nu$, $\omega\in\Omega$ that

$$\Phi^m(\theta,\omega) = (\nabla_\theta \phi^m)(\theta,\omega), \tag{4.55}$$

let $\Theta\colon \mathbb{N}_0\times\Omega \to \mathbb{R}^\nu$ be a stochastic process which satisfies for all $m\in\mathbb{N}$ that

$$\Theta_m = \Theta_{m-1} + \gamma\cdot\Phi^m(\Theta_{m-1}), \tag{4.56}$$

and for every $j \in \mathbb{N}$, $\theta \in \mathbb{R}^\nu$ *let* $\tau_{j,\theta}\colon \Omega \to \{t_0, t_1, \ldots, t_N\}$ *be the random variable given by*

$$\tau_{j,\theta} = \min\Big\{ s \in [0,T] \colon \Big(\exists\, n \in \{0,1,\ldots,N\} \colon \\ \big[s = t_n \text{ and } \textstyle\sum_{k=0}^{n} U_{k,\theta}\big(\mathcal{X}^j_{t_0},\ldots,\mathcal{X}^j_{t_k}\big) \geq 1 - U_{n,\theta}\big(\mathcal{X}^j_{t_0},\ldots,\mathcal{X}^j_{t_n}\big) \big] \Big) \Big\}. \tag{4.57}$$

Consider the setting in Framework 4.1, assume that μ and σ are globally Lipschitz continuous, and assume that g is continuous and at most polynomially growing. In the case of sufficiently large $N, M, J \in \mathbb{N}$ and sufficiently small $\gamma \in (0,\infty)$ we then think of the random real number

$$\frac{1}{J}\left[\sum_{j=1}^{J} g\big(\tau_{M+j,\Theta_M}, \mathcal{X}^{M+j}_{\tau_{M+j,\Theta_M}}\big) \right] \tag{4.58}$$

as an approximation of the price of the American option with the discounted pay-off function g and for every $j \in \mathbb{N}$ we think of the random variable

$$\tau_{M+j,\Theta_M}\colon \Omega \to \{t_0, t_1, \ldots, t_N\} \tag{4.59}$$

as an approximation of an *optimal exercise strategy* associated to the underlying time-discrete path $(\mathcal{X}^{M+j}_t)_{t\in\{t_0,t_1,\ldots,t_N\}}$ (cf. Subsection 4.1.1 above and Section 4.3 below).

4.2.2 Formulation of the proposed algorithm in the general case

In this subsection we extend the framework in Subsection 4.2.1 above and describe the proposed algorithm in the general case.

Framework 4.2. *Let* $T \in (0,\infty)$, $d, N, M, \nu, \varsigma, \varrho \in \mathbb{N}$, *let* $g\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ *be a Borel measurable function, let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, let* $t_0, t_1, \ldots, t_N \in [0,T]$ *be real numbers with* $0 = t_0 < t_1 < \ldots < t_N = T$, *let* $\mathcal{X}^{m,j} = (\mathcal{X}^{m,j,(1)}, \ldots, \mathcal{X}^{m,j,(d)})\colon \{t_0, t_1, \ldots, t_N\} \times \Omega \to \mathbb{R}^d$, $j \in \mathbb{N}$, $m \in \mathbb{N}_0$, *be i.i.d. stochastic processes, for every* $n \in \{0,1,\ldots,N\}$, $\theta \in \mathbb{R}^\nu$, $\mathbf{s} \in \mathbb{R}^\varsigma$ *let* $u^{\theta,\mathbf{s}}_n\colon \mathbb{R}^d \to (0,1)$ *be a function, for every* $n \in \{0,1,\ldots,N\}$, $\theta \in \mathbb{R}^\nu$, $\mathbf{s} \in \mathbb{R}^\varsigma$ *let* $U^{\theta,\mathbf{s}}_n\colon (\mathbb{R}^d)^{n+1} \to (0,1)$ *be the function which satisfies for all* $x_0, x_1, \ldots, x_n \in \mathbb{R}^d$ *that*

$$U^{\theta,\mathbf{s}}_n(x_0, x_1, \ldots, x_n) = \max\{u^{\theta,\mathbf{s}}_n(x_n), n+1-N\} \left[1 - \sum_{k=0}^{n-1} U^{\theta,\mathbf{s}}_k(x_0, x_1, \ldots, x_k) \right], \tag{4.60}$$

let $(J_m)_{m\in\mathbb{N}_0} \subseteq \mathbb{N}$ *be a sequence, for every* $m \in \mathbb{N}$, $\mathbf{s} \in \mathbb{R}^\varsigma$ *let* $\phi^{m,\mathbf{s}}\colon \mathbb{R}^\nu \times \Omega \to \mathbb{R}$ *be the function which satisfies for all* $\theta \in \mathbb{R}^\nu$, $\omega \in \Omega$ *that*

$$\phi^{m,\mathbf{s}}(\theta,\omega) = \frac{1}{J_m} \sum_{j=1}^{J_m} \sum_{n=0}^{N} \Big[U^{\theta,\mathbf{s}}_n\big(\mathcal{X}^{m,j}_{t_0}(\omega), \mathcal{X}^{m,j}_{t_1}(\omega), \ldots, \mathcal{X}^{m,j}_{t_n}(\omega)\big)\, g\big(t_n, \mathcal{X}^{m,j}_{t_n}(\omega)\big) \Big], \tag{4.61}$$

for every $m \in \mathbb{N}$, $\mathbf{s} \in \mathbb{R}^\varsigma$ *let* $\Phi^{m,\mathbf{s}}\colon \mathbb{R}^\nu \times \Omega \to \mathbb{R}^\nu$ *be a function which satisfies for all* $\omega \in \Omega$, $\theta \in \{\eta \in \mathbb{R}^\nu \colon \phi^{m,\mathbf{s}}(\cdot,\omega)\colon \mathbb{R}^\nu \to \mathbb{R} \text{ is differentiable at } \eta\}$ *that*

$$\Phi^{m,\mathbf{s}}(\theta,\omega) = (\nabla_\theta \phi^{m,\mathbf{s}})(\theta,\omega), \tag{4.62}$$

let $\mathcal{S}\colon \mathbb{R}^{\varsigma} \times \mathbb{R}^{\nu} \times (\mathbb{R}^d)^{\{0,1,\ldots,N-1\}\times\mathbb{N}} \to \mathbb{R}^{\varsigma}$ *be a function, for every* $m \in \mathbb{N}$ *let* $\Psi_m\colon \mathbb{R}^{\varrho} \times \mathbb{R}^{\nu} \to \mathbb{R}^{\varrho}$ *and* $\psi_m\colon \mathbb{R}^{\varrho} \to \mathbb{R}^{\nu}$ *be functions, let* $\mathbb{S}\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^{\varsigma}$, $\Xi\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^{\varrho}$, *and* $\Theta\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^{\nu}$ *be stochastic processes which satisfy for all* $m \in \mathbb{N}$ *that*

$$\mathbb{S}_m = \mathcal{S}\big(\mathbb{S}_{m-1}, \Theta_{m-1}, (\mathcal{X}^{m,j}_{t_n})_{(n,j)\in\{0,1,\ldots,N-1\}\times\mathbb{N}}\big), \tag{4.63}$$

$$\Xi_m = \Psi_m(\Xi_{m-1}, \Phi^{m,\mathbb{S}_m}(\Theta_{m-1})), \qquad \text{and} \qquad \Theta_m = \Theta_{m-1} + \psi_m(\Xi_m), \tag{4.64}$$

for every $j \in \mathbb{N}$, $\theta \in \mathbb{R}^{\nu}$, $\mathbf{s} \in \mathbb{R}^{\varsigma}$ *let* $\tau^{j,\theta,\mathbf{s}}\colon \Omega \to \{t_0, t_1, \ldots, t_N\}$ *be the random variable given by*

$$\begin{aligned}\tau^{j,\theta,\mathbf{s}} = \min\Bigg\{ s \in [0,T]\colon \Big(\exists\, n \in \{0,1,\ldots,N\}\colon \\ \big[s = t_n \text{ and } \textstyle\sum_{k=0}^{n} U^{\theta,\mathbf{s}}_k\big(\mathcal{X}^{0,j}_{t_0},\ldots,\mathcal{X}^{0,j}_{t_k}\big) \geq 1 - U^{\theta,\mathbf{s}}_n\big(\mathcal{X}^{0,j}_{t_0},\ldots,\mathcal{X}^{0,j}_{t_n}\big)\big]\Big)\Bigg\},\end{aligned} \tag{4.65}$$

and let $\mathcal{P}\colon \Omega \to \mathbb{R}$ *be the random variable which satisfies for all* $\omega \in \Omega$ *that*

$$\mathcal{P}(\omega) = \frac{1}{J_0}\left[\sum_{j=1}^{J_0} g\big(\tau^{j,\Theta_M(\omega),\mathbb{S}_M(\omega)}(\omega), \mathcal{X}^{0,j}_{\tau^{j,\Theta_M(\omega),\mathbb{S}_M(\omega)}(\omega)}(\omega)\big)\right]. \tag{4.66}$$

Consider the setting in Framework 4.2. Under suitable further assumptions, in the case of sufficiently large $N, M, \nu, J_0 \in \mathbb{N}$ we think of the random real number

$$\mathcal{P} = \frac{1}{J_0}\left[\sum_{j=1}^{J_0} g\big(\tau^{j,\Theta_M,\mathbb{S}_M}, \mathcal{X}^{0,j}_{\tau^{j,\Theta_M,\mathbb{S}_M}}\big)\right] \tag{4.67}$$

as an approximation of the price of the American option with the discounted pay-off function g and for every $j \in \mathbb{N}$ we think of the random variable

$$\tau^{j,\Theta_M,\mathbb{S}_M}\colon \Omega \to \{t_0, t_1, \ldots, t_N\} \tag{4.68}$$

as an approximation of an *optimal exercise strategy* associated to the underlying time-discrete path $(\mathcal{X}^{0,j}_t)_{t\in\{t_0,t_1,\ldots,t_N\}}$ (cf. Subsection 4.1.1 above and Section 4.3 below).

4.2.3 Comments on the proposed algorithm

Note that the lack in Framework 4.2 of any assumptions on the dynamics of the stochastic process $(\mathcal{X}^{0,1}_t)_{t\in\{t_0,t_1,\ldots,t_N\}}$ allows us to approximatively compute the optimal pay-off as well as an optimal exercise strategy for very general optimal stopping problems where, in particular, the stochastic process under consideration is not necessarily a solution of any stochastic differential equation. We only require that the stochastic process $(\mathcal{X}^{0,1}_t)_{t\in\{t_0,t_1,\ldots,t_N\}}$ can be simulated efficiently and formally we still rely on the Markov assumption (cf. Subsection 4.1.4 above). In addition, observe that any particular choice of the functions $u^{\theta,\mathbf{s}}_N\colon \mathbb{R}^d \to (0,1)$, $\mathbf{s} \in \mathbb{R}^{\varsigma}$, $\theta \in \mathbb{R}^{\nu}$, has no influence on the proposed algorithm (cf. (4.60)). Furthermore, the dynamics in (4.64) associated with the stochastic processes $(\Xi_m)_{m\in\mathbb{N}_0}$ and $(\Theta_m)_{m\in\mathbb{N}_0}$ allow us to incorporate different stochastic approximation algorithms such as

- plain vanilla stochastic gradient ascent with or without mini-batches (cf. (4.56) above) as well as
- adaptive moment estimation (Adam) with mini-batches (cf. Kingma & Ba [201] and (4.92)–(4.93) in Subsection 4.3.2 below)

into the algorithm in Subsection 4.2.2 (cf. E, Han, & Jentzen [110, Subsection 3.3]). The dynamics in (4.63) associated with the stochastic process $(\mathbb{S}_m)_{m\in\mathbb{N}_0}$ in turn, allow us to incorporate batch normalisation (cf. Ioffe & Szegedy [184] and the beginning of Section 4.3 below) into the algorithm in Subsection 4.2.2. In that case we think of $(\mathbb{S}_m)_{m\in\mathbb{N}_0}$ as a bookkeeping process keeping track of approximatively calculated means and standard deviations as well as of the number of steps $m \in \mathbb{N}_0$ of the employed stochastic approximation algorithm.

4.3 Numerical examples of pricing American-style derivatives

In this section we test the algorithm in Framework 4.2 on several different examples of pricing American-style financial derivatives.

In each of the examples below we employ the general approximation algorithm in Framework 4.2 above in conjunction with the Adam optimiser (cf. Kingma & Ba [201]) with varying learning rates and with mini-batches (cf. Subsection 4.3.2 below for a precise description).

Furthermore, in the context of Framework 4.2 we employ $N-1$ fully connected feedforward neural networks in each of our implementations for the examples below where the initial value $\mathcal{X}^{0,1}_{t_0}$ is deterministic. In that case the data entering the functions $u^{\theta,\mathbf{s}}_0\colon \mathbb{R}^d \to (0,1)$, $\mathbf{s} \in \mathbb{R}^\varsigma$, $\theta \in \mathbb{R}^\nu$, is deterministic (cf. (4.60)–(4.61)). Therefore, a training procedure is not necessary for the approximative calculations of these functions but is only carried out for the functions $u^{\theta,\mathbf{s}}_1, \dots, u^{\theta,\mathbf{s}}_{N-1}\colon \mathbb{R}^d \to (0,1)$, $\mathbf{s} \in \mathbb{R}^\varsigma$, $\theta \in \mathbb{R}^\nu$. If, however, the initial value $\mathcal{X}^{0,1}_{t_0}$ is not deterministic (cf. the example in Subsection 4.3.4.4 below), a training procedure is carried out for all the functions $u^{\theta,\mathbf{s}}_0, u^{\theta,\mathbf{s}}_1, \dots, u^{\theta,\mathbf{s}}_{N-1}\colon \mathbb{R}^d \to (0,1)$, $\mathbf{s} \in \mathbb{R}^\varsigma$, $\theta \in \mathbb{R}^\nu$, and in that case we hence employ N fully connected feedforward neural networks (cf. Becker, Cheridito, & Jentzen [28, Remark 6 in Subsection 2.3]).

All neural networks employed have one input layer, two hidden layers, and one output layer. As non-linear activation functions just in front of the two hidden layers we employ the multidimensional version of the rectifier function $\mathbb{R} \ni x \mapsto \max\{x, 0\} \in [0,\infty)$, whereas just in front of the output layer we employ the standard logistic function $\mathbb{R} \ni x \mapsto {}^{\exp(x)}/_{(\exp(x)+1)} \in (0,1)$ as non-linear activation function. In addition, batch normalisation (cf. Ioffe & Szegedy [184]) is applied just before the first linear transformation, just before each of the two non-linear activation functions in front of the hidden layers as well as just before the non-linear activation function in front of the output layer. We use Xavier initialisation (cf. Glorot & Bengio [136]) to initialise all weights in the neural networks.

All the examples presented below were implemented in Python. The corresponding Python codes were run, unless stated otherwise (cf. Subsection 4.3.4.1.2 as well as the last sentence in Subsection 4.3.4.1.3 below), in single precision (float32) on a NVIDIA GeForce GTX 1080 GPU with 1974 MHz core clock and 8 GB GDDR5X memory with 1809.5 MHz clock rate, where the underlying system consisted of an Intel Core i7-6800K 3.4 GHz CPU with 64 GB DDR4-2133 memory running Tensorflow 1.5 on Ubuntu 16.04. We would like to point out that no special emphasis has been put on optimising computation speed. In many cases some of the algorithm parameters could be adjusted in order to obtain similarly accurate results in shorter runtime.

4.3.1 Theoretical considerations

Before we present the optimal stopping problem examples on which we have tested the algorithm in Framework 4.2 (cf. Subsections 4.3.3 and 4.3.4 below), we recall a few theoretical results, which are used to design some of these examples and to provide reference values. The elementary and well-known result in Lemma 4.3 below specifies the distributions of linear combinations of independent and identically distributed centred Gaussian random variables which take values in a separable normed $\mathbb{R}$-vector space.

Lemma 4.3. *Let $n \in \mathbb{N}$, $\gamma = (\gamma_1, \dots, \gamma_n) \in \mathbb{R}^n$, let $(V, \|\cdot\|_V)$ be a separable normed $\mathbb{R}$-vector space, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X_i \colon \Omega \to V$, $i \in \{1, \dots, n\}$, be i.i.d. centred Gaussian random variables. Then it holds that*

$$\left(\sum_{i=1}^{n} \gamma_i \, X_i\right)(\mathbb{P})_{\mathcal{B}(V)} = (\|\gamma\|_{\mathbb{R}^n} \, X_1)(\mathbb{P})_{\mathcal{B}(V)}. \tag{4.69}$$

Proof of Lemma 4.3. Throughout this proof let $Y_1, Y_2 \colon \Omega \to V$ be the random variables given by $Y_1 = \sum_{i=1}^{n} \gamma_i \, X_i$ and $Y_2 = \|\gamma\|_{\mathbb{R}^n} \, X_1$. Note that for every $\varphi \in V'$ it holds that $\varphi \circ X_i \colon \Omega \to \mathbb{R}$, $i \in \{1, \dots, n\}$, are independent and identically distributed centred Gaussian random variables. This implies that for all $\varphi \in V'$ it holds that

$$\begin{aligned}
\mathbb{E}\big[e^{\mathbf{i}\,\varphi(Y_1)}\big] &= \mathbb{E}\big[e^{\mathbf{i}\sum_{i=1}^{n} \gamma_i \, \varphi(X_i)}\big] = \mathbb{E}\left[\prod_{i=1}^{n} e^{\mathbf{i}\,\gamma_i \, \varphi(X_i)}\right] = \prod_{i=1}^{n} \mathbb{E}\big[e^{\mathbf{i}\,(\gamma_i \, \varphi)(X_i)}\big] \\
&= \prod_{i=1}^{n} \exp\!\big(-\tfrac{1}{2}\,\mathbb{E}\big[|(\gamma_i \, \varphi)(X_i)|^2\big]\big) = \prod_{i=1}^{n} \exp\!\big(-\tfrac{1}{2}\,\mathbb{E}\big[|(\gamma_i \, \varphi)(X_1)|^2\big]\big) \\
&= \exp\!\left(-\tfrac{1}{2}\,\mathbb{E}\left[\sum_{i=1}^{n} |\gamma_i \, \varphi(X_1)|^2\right]\right) = \exp\!\big(-\tfrac{1}{2}\,\mathbb{E}\big[|(\|\gamma\|_{\mathbb{R}^n} \, \varphi)(X_1)|^2\big]\big) \\
&= \mathbb{E}\big[e^{\mathbf{i}\,\|\gamma\|_{\mathbb{R}^n} \, \varphi(X_1)}\big] = \mathbb{E}\big[e^{\mathbf{i}\,\varphi(Y_2)}\big].
\end{aligned} \tag{4.70}$$

This and, e.g., Jentzen, Salimova, & Welti [193, Lemma 4.10] establish that

$$Y_1(\mathbb{P})_{\mathcal{B}(V)} = Y_2(\mathbb{P})_{\mathcal{B}(V)}. \tag{4.71}$$

The proof of Lemma 4.3 is thus complete. ☐

The next elementary and well-known corollary follows directly from Lemma 4.3.

Corollary 4.4. *Let $d \in \mathbb{N}$, $\gamma = (\gamma_1, \dots, \gamma_d) \in \mathbb{R}^d$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $W = (W^{(1)}, \dots, W^{(d)}) \colon [0, T] \times \Omega \to \mathbb{R}^d$ be a $\mathbb{P}$-standard Brownian motion with continuous sample paths. Then it holds that*

$$\left(\sum_{i=1}^{d} \gamma_i \, W^{(i)}\right)(\mathbb{P})_{\mathcal{B}(C([0,T],\mathbb{R}))} = \big(\|\gamma\|_{\mathbb{R}^d} \, W^{(1)}\big)(\mathbb{P})_{\mathcal{B}(C([0,T],\mathbb{R}))}. \tag{4.72}$$

The next elementary and well-known result, Proposition 4.5, states that the distribution of a product of multiple correlated geometric Brownian motions is equal to the distribution of a single particular geometric Brownian motion.

Proposition 4.5. *Let* $T, \epsilon \in (0,\infty)$, $d \in \mathbb{N}$, $\mathfrak{S} = (\varsigma_1, \dots, \varsigma_d) \in \mathbb{R}^{d\times d}$, $\xi = (\xi_1, \dots, \xi_d), \alpha = (\alpha_1, \dots, \alpha_d), \beta = (\beta_1, \dots, \beta_d) \in \mathbb{R}^d$, *let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, let* $\mathscr{F}^{(i)} = (\mathscr{F}^{(i)}_t)_{t\in[0,T]}$, $i \in \{1,2\}$, *be filtrations on* $(\Omega, \mathcal{F}, \mathbb{P})$ *that satisfy the usual conditions, let* $W = (W^{(1)}, \dots, W^{(d)})\colon [0,T] \times \Omega \to \mathbb{R}^d$ *be a standard* $(\Omega, \mathcal{F}, \mathbb{P}, \mathscr{F}^{(1)})$*-Brownian motion with continuous sample paths, let* $\mathrm{W}\colon [0,T]\times\Omega \to \mathbb{R}$ *be a standard* $(\Omega, \mathcal{F}, \mathbb{P}, \mathscr{F}^{(2)})$*-Brownian motion with continuous sample paths, let* $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$, $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$, $\mathbf{P}\colon C([0,T], \mathbb{R}^d) \to C([0,T],\mathbb{R})$, *and* $\mathbf{G}\colon C([0,T],\mathbb{R}) \to C([0,T],\mathbb{R})$ *be the functions which satisfy for all* $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, $u^{(1)} = (u^{(1)}_s)_{s\in[0,T]}, \dots, u^{(d)} = (u^{(d)}_s)_{s\in[0,T]} \in C([0,T],\mathbb{R})$, $t \in [0,T]$ *that* $\mu(x) = (\alpha_1 x_1, \dots, \alpha_d x_d)$, $\sigma(x) = \operatorname{diag}(\beta_1 x_1, \dots, \beta_d x_d)\,\mathfrak{S}^*$, $(\mathbf{G}[u^{(1)}])_t = \exp\!\big(\epsilon \big[\sum_{i=1}^d \alpha_i - \|\beta_i\varsigma_i\|^2_{\mathbb{R}^d}/2\big] t + \epsilon\, \|\mathfrak{S}\,\beta\|_{\mathbb{R}^d}\, u^{(1)}_t\big) \prod_{i=1}^d |\xi_i|^\epsilon$, *and* $(\mathbf{P}[(u^{(1)}, \dots, u^{(d)})])_t = \prod_{i=1}^d \big|u^{(i)}_t\big|^\epsilon$, *let* $X = (X^{(1)}, \dots, X^{(d)})\colon [0,T]\times\Omega \to \mathbb{R}^d$ *be an* $\mathscr{F}^{(1)}$*-adapted stochastic process with continuous sample paths, let* $Y\colon [0,T]\times\Omega\to\mathbb{R}$ *be an* $\mathscr{F}^{(2)}$*-adapted stochastic process with continuous sample paths, and assume that for all* $t\in[0,T]$ *it holds* $\mathbb{P}$*-a.s. that*

$$X_t = \xi + \int_0^t \mu(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}W_s, \tag{4.73}$$

$$Y_t = \prod_{i=1}^d |\xi_i|^\epsilon + \left(\epsilon\left[\sum_{i=1}^d \alpha_i - \frac{\|\beta_i\varsigma_i\|^2_{\mathbb{R}^d}}{2}\right] + \frac{\|\epsilon\,\mathfrak{S}\,\beta\|^2_{\mathbb{R}^d}}{2}\right)\int_0^t Y_s\,\mathrm{d}s + \epsilon\,\|\mathfrak{S}\,\beta\|_{\mathbb{R}^d}\int_0^t Y_s\,\mathrm{d}\mathrm{W}_s. \tag{4.74}$$

Then

(i) for all $i\in\{1,\dots,d\}$, $t\in[0,T]$ *it holds* $\mathbb{P}$*-a.s. that*

$$X^{(i)}_t = \exp\!\left(\left[\alpha_i - \frac{\|\beta_i\varsigma_i\|^2_{\mathbb{R}^d}}{2}\right] t + \beta_i \langle \varsigma_i, W_t\rangle_{\mathbb{R}^d}\right)\xi_i, \tag{4.75}$$

(ii) it holds that $\mathbf{P}$ *and* $\mathbf{G}$ *are continuous functions, and*

(iii) it holds that

$$(\mathbf{P}\circ X)(\mathbb{P})_{\mathcal{B}(C([0,T],\mathbb{R}))} = (\mathbf{G}\circ \mathrm{W})(\mathbb{P})_{\mathcal{B}(C([0,T],\mathbb{R}))} = Y(\mathbb{P})_{\mathcal{B}(C([0,T],\mathbb{R}))}. \tag{4.76}$$

Proof of Proposition 4.5. Throughout this proof let $\gamma = (\gamma_1, \dots, \gamma_d) \in \mathbb{R}^d$ be the vector given by $\gamma = \mathfrak{S}\,\beta$, let $Z^{(i)}\colon [0,T]\times\Omega\to\mathbb{R}$, $i\in\{1,\dots,d\}$, be the stochastic processes which satisfy for all $i\in\{1,\dots,d\}$, $t\in[0,T]$ that

$$Z^{(i)}_t = \left[\alpha_i - \frac{\|\beta_i\varsigma_i\|^2_{\mathbb{R}^d}}{2}\right] t + \beta_i\langle\varsigma_i, W_t\rangle_{\mathbb{R}^d}, \tag{4.77}$$

and let $\tilde{\mathbf{G}}\colon C([0,T],\mathbb{R}) \to C([0,T],\mathbb{R})$ be the function which satisfies for all $u = (u_s)_{s\in[0,T]} \in C([0,T],\mathbb{R})$, $t\in[0,T]$ that

$$(\tilde{\mathbf{G}}[u])_t = \exp\!\left(\epsilon\left[\sum_{i=1}^d \alpha_i - \frac{\|\beta_i\varsigma_i\|^2_{\mathbb{R}^d}}{2}\right] t + \epsilon\, u_t\right)\prod_{i=1}^d |\xi_i|^\epsilon. \tag{4.78}$$

Observe that for all $i \in \{1, \ldots, d\}$, $t \in [0, T]$ it holds $\mathbb{P}$-a.s. that

$$X_t^{(i)} = \xi_i + \alpha_i \int_0^t X_s^{(i)} \, \mathrm{d}s + \beta_i \int_0^t X_s^{(i)} \langle \varsigma_i, \mathrm{d}W_s \rangle_{\mathbb{R}^d}. \tag{4.79}$$

In addition, note that (4.77) implies that for all $i \in \{1, \ldots, d\}$, $t \in [0, T]$ it holds $\mathbb{P}$-a.s. that

$$Z_t^{(i)} = \int_0^t \alpha_i - \frac{\|\beta_i \varsigma_i\|_{\mathbb{R}^d}^2}{2} \, \mathrm{d}s + \int_0^t \beta_i \langle \varsigma_i, \mathrm{d}W_s \rangle_{\mathbb{R}^d}. \tag{4.80}$$

Itô's formula hence shows that for all $i \in \{1, \ldots, d\}$, $t \in [0, T]$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} e^{Z_t^{(i)}} \xi_i &= \xi_i + \left[\alpha_i - \frac{\|\beta_i \varsigma_i\|_{\mathbb{R}^d}^2}{2}\right] \int_0^t e^{Z_s^{(i)}} \xi_i \, \mathrm{d}s + \beta_i \int_0^t e^{Z_s^{(i)}} \xi_i \langle \varsigma_i, \mathrm{d}W_s \rangle_{\mathbb{R}^d} \\ &\quad + \frac{\|\beta_i \varsigma_i\|_{\mathbb{R}^d}^2}{2} \int_0^t e^{Z_s^{(i)}} \xi_i \, \mathrm{d}s \\ &= \xi_i + \alpha_i \int_0^t e^{Z_s^{(i)}} \xi_i \, \mathrm{d}s + \beta_i \int_0^t e^{Z_s^{(i)}} \xi_i \langle \varsigma_i, \mathrm{d}W_s \rangle_{\mathbb{R}^d}. \end{aligned} \tag{4.81}$$

Combining this and (4.79) with, e.g., Da Prato & Zabczyk [94, (i) in Theorem 7.4] proves that for all $i \in \{1, \ldots, d\}$, $t \in [0, T]$ it holds $\mathbb{P}$-a.s. that

$$X_t^{(i)} = e^{Z_t^{(i)}} \xi_i = \exp\!\left(\left[\alpha_i - \frac{\|\beta_i \varsigma_i\|_{\mathbb{R}^d}^2}{2}\right] t + \beta_i \langle \varsigma_i, W_t \rangle_{\mathbb{R}^d} \right) \xi_i. \tag{4.82}$$

This establishes (i). In the next step note that (ii) is clear. It thus remains to prove (iii). For this observe that (i) establishes that for all $t \in [0, T]$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} (\mathbf{P}[X])_t &= \prod_{i=1}^d |X_t^{(i)}|^\epsilon = \prod_{i=1}^d \left[\exp\!\left(\epsilon \left[\alpha_i - \frac{\|\beta_i \varsigma_i\|_{\mathbb{R}^d}^2}{2}\right] t + \epsilon \, \beta_i \langle \varsigma_i, W_t \rangle_{\mathbb{R}^d} \right) |\xi_i|^\epsilon \right] \\ &= \exp\!\left(\epsilon \left[\sum_{i=1}^d \alpha_i - \frac{\|\beta_i \varsigma_i\|_{\mathbb{R}^d}^2}{2} \right] t + \epsilon \left\langle \sum_{i=1}^d \varsigma_i \beta_i, W_t \right\rangle_{\mathbb{R}^d} \right) \prod_{i=1}^d |\xi_i|^\epsilon \\ &= \exp\!\left(\epsilon \left[\sum_{i=1}^d \alpha_i - \frac{\|\beta_i \varsigma_i\|_{\mathbb{R}^d}^2}{2} \right] t + \epsilon \, \langle \gamma, W_t \rangle_{\mathbb{R}^d} \right) \prod_{i=1}^d |\xi_i|^\epsilon \\ &= \left(\tilde{\mathbf{G}}\left[\sum_{i=1}^d \gamma_i \, W^{(i)} \right] \right)_t. \end{aligned} \tag{4.83}$$

Continuity hence implies that it holds $\mathbb{P}$-a.s. that

$$\mathbf{P}[X] = \tilde{\mathbf{G}}\left[\sum_{i=1}^d \gamma_i \, W^{(i)} \right]. \tag{4.84}$$

Moreover, note that (i) shows that for all $t \in [0, T]$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} Y_t &= \exp\!\left(\left\{ \epsilon \left[\sum_{i=1}^d \alpha_i - \frac{\|\beta_i \varsigma_i\|_{\mathbb{R}^d}^2}{2} \right] + \frac{\|\epsilon \, \mathfrak{S} \, \beta\|_{\mathbb{R}^d}^2}{2} - \frac{\|\epsilon \, \mathfrak{S} \, \beta\|_{\mathbb{R}^d}^2}{2} \right\} t + \epsilon \, \|\mathfrak{S} \, \beta\|_{\mathbb{R}^d} \, \mathrm{W}_t \right) \prod_{i=1}^d |\xi_i|^\epsilon \\ &= \exp\!\left(\epsilon \left[\sum_{i=1}^d \alpha_i - \frac{\|\beta_i \varsigma_i\|_{\mathbb{R}^d}^2}{2} \right] t + \epsilon \, \|\mathfrak{S} \, \beta\|_{\mathbb{R}^d} \, \mathrm{W}_t \right) \prod_{i=1}^d |\xi_i|^\epsilon \\ &= (\mathbf{G}[\mathrm{W}])_t. \end{aligned} \tag{4.85}$$

This and continuity establish that it holds $\mathbb{P}$-a.s. that

$$Y = \mathbf{G}[\mathrm{w}]. \tag{4.86}$$

Furthermore, observe that Corollary 4.4 ensures that

$$\left(\sum_{i=1}^{d} \gamma_i \, W^{(i)}\right)(\mathbb{P})_{\mathcal{B}(C([0,T],\mathbb{R}))} = \big(\|\gamma\|_{\mathbb{R}^d} \, W^{(1)}\big)(\mathbb{P})_{\mathcal{B}(C([0,T],\mathbb{R}))}. \tag{4.87}$$

The fact that $\tilde{\mathbf{G}} \colon C([0,T],\mathbb{R}) \to C([0,T],\mathbb{R})$ is a Borel measurable function, (ii), the fact that $\forall \, u \in C([0,T],\mathbb{R}) \colon \tilde{\mathbf{G}}[\|\gamma\|_{\mathbb{R}^d} \, u] = \mathbf{G}[u]$, (4.84), and (4.86) hence demonstrate that

$$\begin{aligned}
&(\mathbf{P} \circ X)(\mathbb{P})_{\mathcal{B}(C([0,T],\mathbb{R}))} = \left(\tilde{\mathbf{G}} \circ \left(\sum_{i=1}^{d} \gamma_i \, W^{(i)}\right)\right)(\mathbb{P})_{\mathcal{B}(C([0,T],\mathbb{R}))} \\
&= \big(\tilde{\mathbf{G}} \circ \big(\|\gamma\|_{\mathbb{R}^d} \, W^{(1)}\big)\big)(\mathbb{P})_{\mathcal{B}(C([0,T],\mathbb{R}))} = \big(\mathbf{G} \circ W^{(1)}\big)(\mathbb{P})_{\mathcal{B}(C([0,T],\mathbb{R}))} \\
&= (\mathbf{G} \circ \mathrm{w})(\mathbb{P})_{\mathcal{B}(C([0,T],\mathbb{R}))} = Y(\mathbb{P})_{\mathcal{B}(C([0,T],\mathbb{R}))}.
\end{aligned} \tag{4.88}$$

The proof of Proposition 4.5 is thus complete. □

In the next result, Lemma 4.6, we recall the well-known formula for the price of a European call option on a single stock in the Black–Scholes model.

Lemma 4.6. *Let $T, \xi, \sigma \in (0,\infty)$, $r, c \in \mathbb{R}$, let $\Phi \colon \mathbb{R} \to \mathbb{R}$ be the function which satisfies for all $x \in \mathbb{R}$ that $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}y^2} \, \mathrm{d}y$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $\mathscr{F} = (\mathscr{F}_t)_{t \in [0,T]}$ that satisfies the usual conditions, let $W \colon [0,T] \times \Omega \to \mathbb{R}$ be a standard $(\Omega, \mathcal{F}, \mathbb{P}, \mathscr{F})$-Brownian motion with continuous sample paths, and let $X \colon [0,T] \times \Omega \to \mathbb{R}$ be an $\mathscr{F}$-adapted stochastic process with continuous sample paths which satisfies that for all $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that*

$$X_t = \xi + (r - c) \int_0^t X_s \, \mathrm{d}s + \sigma \int_0^t X_s \, \mathrm{d}W_s. \tag{4.89}$$

Then it holds for all $K \in \mathbb{R}$ that

$$\begin{aligned}
&\mathbb{E}\big[e^{-rT} \max\{X_T - K, 0\}\big] \\
&= \begin{cases} e^{-cT} \xi \, \Phi\Big(\frac{\big(r - c + \frac{\sigma^2}{2}\big)T + \ln(\xi/K)}{\sigma\sqrt{T}}\Big) - K e^{-rT} \, \Phi\Big(\frac{\big(r - c - \frac{\sigma^2}{2}\big)T + \ln(\xi/K)}{\sigma\sqrt{T}}\Big) & : K > 0 \\ e^{-cT}\xi - K e^{-rT} & : K \leq 0 \end{cases}.
\end{aligned} \tag{4.90}$$

4.3.2 Setting

Framework 4.3. *Assume Framework 4.2, let $\zeta_1 = 0.9$, $\zeta_2 = 0.999$, $\varepsilon \in (0,\infty)$, $(\gamma_m)_{m \in \mathbb{N}} \subseteq (0,\infty)$, $\xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d$, let $\mathscr{F} = (\mathscr{F}_t)_{t \in [0,T]}$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ that satisfies the usual conditions, let $W^{m,j} = (W^{m,j,(1)}, \ldots, W^{m,j,(d)}) \colon [0,T] \times \Omega \to \mathbb{R}^d$, $j \in \mathbb{N}$, $m \in \mathbb{N}_0$, be independent standard $(\Omega, \mathcal{F}, \mathbb{P}, \mathscr{F})$-Brownian motions with continuous sample paths, let $\mu \colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma \colon \mathbb{R}^d \to \mathbb{R}^{d \times d}$ be Lipschitz continuous functions, let*

$X = (X^{(1)}, \dots, X^{(d)})\colon [0,T] \times \Omega \to \mathbb{R}^d$ be an $\mathscr{F}$-adapted stochastic process with continuous sample paths which satisfies that for all $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that

$$X_t = \xi + \int_0^t \mu(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}W_s^{0,1}, \tag{4.91}$$

assume for all $n \in \{0, 1, \dots, N\}$ that $\varrho = 2\nu$, $\Xi_0 = 0$, and $t_n = \frac{nT}{N}$, and assume for all $m \in \mathbb{N}$, $x = (x_1, \dots, x_\nu)$, $y = (y_1, \dots, y_\nu)$, $\eta = (\eta_1, \dots, \eta_\nu) \in \mathbb{R}^\nu$ that

$$\Psi_m(x, y, \eta) = \big(\zeta_1 x + (1 - \zeta_1)\eta, \zeta_2 y + (1 - \zeta_2)((\eta_1)^2, \dots, (\eta_\nu)^2)\big) \tag{4.92}$$

and

$$\psi_m(x, y) = \left(\left[\sqrt{\tfrac{|y_1|}{1-(\zeta_2)^m}} + \varepsilon \right]^{-1} \frac{\gamma_m x_1}{1 - (\zeta_1)^m}, \dots, \left[\sqrt{\tfrac{|y_\nu|}{1-(\zeta_2)^m}} + \varepsilon \right]^{-1} \frac{\gamma_m x_\nu}{1 - (\zeta_1)^m} \right). \tag{4.93}$$

Equations (4.92)–(4.93) in Framework 4.3 describe the Adam optimiser with possibly varying learning rates (cf. Kingma & Ba [201] and, e.g., E, Han, & Jentzen [110, (4.3)–(4.4) in Subsection 4.1 and (5.4)–(5.5) in Subsection 5.2]). Furthermore, in the context of pricing American-style financial derivatives, we think

- of T as the maturity,
- of d as the dimension of the associated optimal stopping problem,
- of N as the time discretisation parameter employed,
- of M as the total number of training steps employed in the Adam optimiser,
- of g as the discounted pay-off function,
- of $\{t_0, t_1, \dots, t_N\}$ as the discrete time grid employed,
- of J_0 as the number of Monte Carlo samples employed in the final integration for the price approximation,
- of $(J_m)_{m\in\mathbb{N}}$ as the sequence of batch sizes employed in the Adam optimiser,
- of ζ_1 as the momentum decay factor, of ζ_2 as the second momentum decay factor, and of ε as the regularising factor employed in the Adam optimiser,
- of $(\gamma_m)_{m\in\mathbb{N}}$ as the sequence of learning rates employed in the Adam optimiser,
- and, where applicable, of X as a continuous-time model for d underlying stock prices with initial prices ξ, drift coefficient function μ, and diffusion coefficient function σ.

Moreover, note that for every $m \in \mathbb{N}_0$, $j \in \mathbb{N}$ the stochastic processes $W^{m,j,(1)} = (W_t^{m,j,(1)})_{t\in[0,T]}$, …, $W^{m,j,(d)} = (W_t^{m,j,(d)})_{t\in[0,T]}$ are the components of the d-dimensional standard Brownian motion $W^{m,j} = (W_t^{m,j})_{t\in[0,T]}$ and hence each a one-dimensional standard Brownian motion.

4.3.3 Examples with known one-dimensional representation

In this subsection we test the algorithm in Framework 4.2 in the case of several very simple optimal stopping problem examples in which the d-dimensional optimal stopping problem under consideration has been designed in such a way that it can be represented as a one-dimensional optimal stopping problem. This representation allows us to employ a numerical method for the one-dimensional optimal stopping problem to compute reference values for the original d-dimensional optimal stopping problem. We refer to Subsection 4.3.4 below for more challenging examples where a one-dimensional representation is not known.

4.3.3.1 Optimal stopping of a Brownian motion

4.3.3.1.1 A Bermudan two-exercise put-type example

In this subsection we test the algorithm in Framework 4.2 on the example of optimally stopping a correlated Brownian motion under a put option inspired pay-off function with two possible exercise dates.

Assume Framework 4.3, let $r = 0.02 = 2\%$, $\beta = 0.3 = 30\%$, $\chi = 95$, $K = 90$, $Q = (Q_{i,j})_{(i,j)\in\{1,\dots,d\}^2}, \mathfrak{S} \in \mathbb{R}^{d\times d}$ satisfy for all $i \in \{1,\dots,d\}$ that $Q_{i,i} = 1$, $\forall\, j \in \{1,\dots,d\}\setminus\{i\}\colon Q_{i,j} = 0.1$, and $\mathfrak{S}\,\mathfrak{S}^* = Q$, let $\mathbb{F} = (\mathbb{F}_t)_{t\in[0,T]}$ be the filtration generated by $W^{0,1}$, and assume for all $m, j \in \mathbb{N}$, $n \in \{0,1,2\}$, $s \in [0,T]$, $x = (x_1,\dots,x_d) \in \mathbb{R}^d$ that $T = 1$, $N = 2$, $M = 500$, $\mathcal{X}^{m-1,j}_{t_n} = \mathfrak{S}\, W^{m-1,j}_{t_n}$, $J_0 = 4\,096\,000$, $J_m = 8192$, $\varepsilon = 10^{-8}$, $\gamma_m = 5\,[10^{-2}\,\mathbb{1}_{[1,100]}(m) + 10^{-3}\,\mathbb{1}_{(100,300]}(m) + 10^{-4}\,\mathbb{1}_{(300,\infty)}(m)]$, and

$$g(s,x) = e^{-rs}\max\!\left\{K - \exp\!\left(\left[r - \tfrac{1}{2}\beta^2\right]s + \tfrac{\beta\sqrt{10}}{\sqrt{d(d+9)}}[x_1 + \ldots + x_d]\right)\chi, 0\right\}. \tag{4.94}$$

Note that the distribution of the random variable $\Omega \ni \omega \mapsto ([0,T] \ni t \mapsto g(t, \mathfrak{S}\, W^{0,1}_t(\omega)) \in \mathbb{R}) \in C([0,T],\mathbb{R})$ does not depend on the dimension d (cf. Corollary 4.4). The random variable $\mathcal{P}$ provides approximations for the real number

$$\sup\!\left\{\mathbb{E}\!\left[g(\tau, \mathfrak{S}\, W^{0,1}_\tau)\right] : \substack{\tau\colon \Omega\to\{t_0,t_1,t_2\}\text{ is an}\\ (\mathbb{F}_t)_{t\in\{t_0,t_1,t_2\}}\text{-stopping time}}\right\}. \tag{4.95}$$

In Table 4.1 we show approximations for the mean and for the standard deviation of $\mathcal{P}$ and the average runtime in seconds needed for calculating one realisation of $\mathcal{P}$ for $d \in \{1, 5, 10, 50, 100, 500, 1000\}$. For each case the calculations of the results in Table 4.1 are based on 10 independent realisations of $\mathcal{P}$, which have been obtained from an implementation in Python.

4.3.3.1.2 An American put-type example

In this subsection we test the algorithm in Framework 4.2 on the example of optimally stopping a standard Brownian motion under a put option inspired pay-off function.

Assume Framework 4.3, let $r = 0.06 = 6\%$, $\beta = 0.4 = 40\%$, $\chi = K = 40$, let $\mathbb{F} = (\mathbb{F}_t)_{t\in[0,T]}$ be the filtration generated by $W^{0,1}$, let $\mathfrak{F} = (\mathfrak{F}_t)_{t\in[0,T]}$ be the filtration generated by $W^{0,1,(1)}$, and assume for all $m, j \in \mathbb{N}$, $n \in \{0,1,\dots,N\}$, $s \in [0,T]$, $x = (x_1,\dots,x_d) \in \mathbb{R}^d$ that $T = 1$, $N = 50$, $M = 1500\,\mathbb{1}_{[1,50]}(d) + 1800\,\mathbb{1}_{(50,100]}(d) + 3000\,\mathbb{1}_{(100,\infty)}(d)$, $\mathcal{X}^{m-1,j}_{t_n} =$

Dimension d	Mean of $\mathcal{P}$	Standard deviation of $\mathcal{P}$	Runtime in sec. for one realisation of $\mathcal{P}$
1	7.890	0.004	4.2
5	7.892	0.007	4.2
10	7.892	0.005	4.4
50	7.890	0.005	5.5
100	7.891	0.004	7.3
500	7.891	0.005	24.3
1000	7.892	0.007	54.4

Table 4.1: Numerical simulations of the algorithm in Framework 4.2 for optimally stopping a correlated Brownian motion in the case of the Bermudan two-exercise put-type example in Subsection 4.3.3.1.1.

$W^{m-1,j}_{t_n}$, $J_0 = 4\,096\,000$, $J_m = 8192\,\mathbb{1}_{[1,50]}(d)+4096\,\mathbb{1}_{(50,100]}(d)+2048\,\mathbb{1}_{(100,\infty)}(d)$, $\varepsilon = 0.001$, $\gamma_m = 5\,[10^{-2}\,\mathbb{1}_{[1,M/3]}(m) + 10^{-3}\,\mathbb{1}_{(M/3,2M/3]}(m) + 10^{-4}\,\mathbb{1}_{(2M/3,\infty)}(m)]$, and

$$g(s,x) = e^{-rs}\max\Big\{K - \exp\Big(\big[r - \tfrac{1}{2}\beta^2\big]s + \tfrac{\beta}{\sqrt{d}}[x_1 + \ldots + x_d]\Big)\chi, 0\Big\}. \tag{4.96}$$

The random variable $\mathcal{P}$ provides approximations for the real number

$$\sup\Big\{\mathbb{E}\big[g(\tau, W^{0,1}_\tau)\big] : \substack{\tau\colon\Omega\to[0,T]\text{ is an}\\ \mathbb{F}\text{-stopping time}}\Big\}. \tag{4.97}$$

We show approximations for the mean of $\mathcal{P}$, for the standard deviation of $\mathcal{P}$, and for the relative L^1-approximation error associated to $\mathcal{P}$, the uncorrected sample standard deviation of the relative approximation error associated to $\mathcal{P}$, and the average runtime in seconds needed for calculating one realisation of $\mathcal{P}$ for $d \in \{1, 5, 10, 50, 100, 500, 1000\}$ in Table 4.2. For each case the calculations of the results in Table 4.2 are based on 10 independent realisations of $\mathcal{P}$, which have been obtained from an implementation in Python. Furthermore, in the approximative calculations of the relative approximation error associated to $\mathcal{P}$ the exact number (4.97) has been replaced, independently of the dimension d, by the real number

$$\sup\Big\{\mathbb{E}\Big[e^{-r\tau}\max\Big\{K - \exp\Big(\big[r - \tfrac{1}{2}\beta^2\big]\tau + \beta\, W^{0,1,(1)}_\tau\Big)\chi, 0\Big\}\Big] : \substack{\tau\colon\Omega\to[0,T]\text{ is an}\\ \mathfrak{F}\text{-stopping time}}\Big\} \tag{4.98}$$

(cf. Corollary 4.4), which, in turn, has been replaced by the value 5.318 (cf. Longstaff & Schwartz [232, Table 1 in Section 3]). This value has been calculated using the binomial tree method on M. Smirnov's website [290] with 20 000 nodes. Note that (4.98) corresponds to the price of an American put option on a single stock in the Black–Scholes model with initial stock price χ, interest rate r, volatility β, strike price K, and maturity T.

4.3.3.2 Geometric average-type options

4.3.3.2.1 An American geometric average put-type example

In this subsection we test the algorithm in Framework 4.2 on the example of pricing an American geometric average put-type option on up to 200 distinguishable stocks in the Black–Scholes model.

Dimension d	Mean of $\mathcal{P}$	Standard deviation of $\mathcal{P}$	Rel. L^1-approx. err.	Standard deviation of the rel. approx. err.	Runtime in sec. for one realisation of $\mathcal{P}$
1	5.311	0.002	0.0013	0.0004	78.6
5	5.310	0.003	0.0015	0.0005	91.3
10	5.309	0.003	0.0017	0.0005	104.6
50	5.306	0.003	0.0022	0.0006	215.7
100	5.305	0.004	0.0025	0.0006	245.1
500	5.298	0.003	0.0037	0.0005	1006.3
1000	5.294	0.003	0.0046	0.0006	2266.0

Table 4.2: Numerical simulations of the algorithm in Framework 4.2 for optimally stopping a standard Brownian motion in the case of the American put-type example in Subsection 4.3.3.1.2. In the approximative calculations of the relative approximation errors the exact number (4.97) has been replaced by the value 5.318, which has been obtained using the binomial tree method on M. Smirnov's website [290].

Assume Framework 4.3, assume that $d \in \{40, 80, 120, \ldots\}$, let $\beta = (\beta_1, \ldots, \beta_d) \in \mathbb{R}^d$, $\rho, \tilde{\delta}, \tilde{\beta}, \delta_1, \delta_2, \ldots, \delta_d \in \mathbb{R}$, $r = 0.6$, $K = 95$, $\tilde{\xi} = 100$ satisfy for all $i \in \{1, \ldots, d\}$ that $\beta_i = \min\{0.04\,[(i-1) \bmod 40], 1.6 - 0.04\,[(i-1) \bmod 40]\}$, $\rho = \frac{1}{d}\|\beta\|^2_{\mathbb{R}^d} = \frac{1}{40}\sum_{i=1}^{40}(\beta_i)^2 = 0.2136$, $\delta_i = r - \frac{\rho}{d}\big(i - \frac{1}{2}\big) - \frac{1}{5\sqrt{d}}$, $\tilde{\delta} = r - \frac{1}{\sqrt{d}}\sum_{i=1}^{d}(r - \delta_i) + \frac{\sqrt{d}-1}{2d}\|\beta\|^2_{\mathbb{R}^d} = r - \frac{\rho}{2} - \frac{1}{5} = 0.2932$, and $\tilde{\beta} = \frac{1}{\sqrt{d}}\|\beta\|_{\mathbb{R}^d} = \sqrt{\rho}$, let $Y\colon [0,T] \times \Omega \to \mathbb{R}$ be an $\mathscr{F}$-adapted stochastic process with continuous sample paths which satisfies that for all $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that

$$Y_t = \tilde{\xi} + (r - \tilde{\delta}) \int_0^t Y_s \,\mathrm{d}s + \tilde{\beta} \int_0^t Y_s \,\mathrm{d}W_s^{0,1,(1)}, \tag{4.99}$$

let $\mathbb{F} = (\mathbb{F}_t)_{t\in[0,T]}$ be the filtration generated by X, let $\mathfrak{F} = (\mathfrak{F}_t)_{t\in[0,T]}$ be the filtration generated by Y, and assume for all $m, j \in \mathbb{N}$, $n \in \{0, 1, \ldots, N\}$, $i \in \{1, \ldots, d\}$, $s \in [0,T]$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ that $T = 1$, $N = 100$, $M = 1800\,\mathbb{1}_{[1,120]}(d) + 3000\,\mathbb{1}_{(120,\infty)}(d)$, $J_0 = 4\,096\,000$, $J_m = 8192\,\mathbb{1}_{[1,120]}(d) + 4096\,\mathbb{1}_{(120,\infty)}(d)$, $\varepsilon = 10^{-8}$, $\gamma_m = 5\,[10^{-2}\,\mathbb{1}_{[1,M/3]}(m) + 10^{-3}\,\mathbb{1}_{(M/3,2M/3]}(m) + 10^{-4}\,\mathbb{1}_{(2M/3,\infty)}(m)]$, $\xi_i = (100)^{1/\sqrt{d}}$, $\mu(x) = ((r-\delta_1)\,x_1, \ldots, (r-\delta_d)\,x_d)$, $\sigma(x) = \operatorname{diag}(\beta_1 x_1, \ldots, \beta_d x_d)$, that

$$\mathcal{X}_{t_n}^{m-1,j,(i)} = \exp\!\big(\big[r - \delta_i - \tfrac{1}{2}(\beta_i)^2\big]t_n + \beta_i\, W_{t_n}^{m-1,j,(i)}\big)\,\xi_i, \tag{4.100}$$

and that

$$g(s,x) = e^{-rs} \max\left\{K - \left[\prod_{k=1}^{d} |x_k|^{1/\sqrt{d}}\right], 0\right\}. \tag{4.101}$$

The random variable $\mathcal{P}$ provides approximations for the price

$$\sup\left\{\mathbb{E}\big[g(\tau, X_\tau)\big] : \substack{\tau\colon \Omega\to[0,T] \text{ is an} \\ \mathbb{F}\text{-stopping time}}\right\}. \tag{4.102}$$

In Table 4.3 we show approximations for the mean of $\mathcal{P}$, for the standard deviation of $\mathcal{P}$, and for the relative L^1-approximation error associated to $\mathcal{P}$, the uncorrected sample standard deviation of the relative approximation error associated to $\mathcal{P}$, and the average runtime in seconds needed for calculating one realisation of $\mathcal{P}$ for $d \in \{40, 80, 120, 160, 200\}$.

Dimension d	Mean of $\mathcal{P}$	Standard deviation of $\mathcal{P}$	Rel. L^1-approx. err.	Standard deviation of the rel. approx. err.	Runtime in sec. for one realisation of $\mathcal{P}$
40	6.510	0.004	0.0053	0.0006	477.7
80	6.508	0.003	0.0056	0.0005	793.5
120	6.505	0.003	0.0061	0.0005	934.7
160	6.504	0.003	0.0062	0.0005	1203.9
200	6.504	0.004	0.0063	0.0005	1475.0

Table 4.3: Numerical simulations of the algorithm in Framework 4.2 for pricing the American geometric average put-type option from the example in Subsection 4.3.3.2.1. In the approximative calculations of the relative approximation errors the exact value of the price (4.102) has been replaced by the value 6.545, which has been obtained using the binomial tree method on M. Smirnov's website [290].

For each case the calculations of the results in Table 4.3 are based on 10 independent realisations of $\mathcal{P}$, which have been obtained from an implementation in PYTHON. Furthermore, in the approximative calculations of the relative approximation error associated to $\mathcal{P}$ the exact value of the price (4.102) has been replaced, independently of the dimension d, by the real number

$$\sup\Big\{\mathbb{E}\big[e^{-r\tau}\max\{K-Y_\tau,0\}\big] : \substack{\tau\colon \Omega\to[0,T] \text{ is an} \\ \mathfrak{F}\text{-stopping time}}\Big\}, \tag{4.103}$$

(cf. Proposition 4.5), which, in turn, has been replaced by the value 6.545. The latter has been calculated using the binomial tree method on M. Smirnov's website [290] with 20 000 nodes. Note that (4.103) corresponds to the price of an American put option on a single stock in the Black–Scholes model with initial stock price $\tilde{\xi}$, interest rate r, dividend rate $\tilde{\delta}$, volatility $\tilde{\beta}$, strike price K, and maturity T.

4.3.3.2.2 An American geometric average call-type example

In this subsection we test the algorithm in Framework 4.2 on the example of pricing an American geometric average call-type option on up to 100 correlated stocks in the Black–Scholes model. This example is taken from Sirignano & Spiliopoulos [288, Subsection 4.3], from where we consider the cases with 3, 20, and 100 dimensions.

Assume Framework 4.3, let $r = 0\%$, $\delta = 0.02 = 2\%$, $\beta = 0.25 = 25\%$, $K = \tilde{\xi} = 1$, $Q = (Q_{i,j})_{(i,j)\in\{1,\dots,d\}^2}, \mathfrak{S} = (\varsigma_1,\dots,\varsigma_d) \in \mathbb{R}^{d\times d}$, $\tilde{\delta}, \tilde{\beta} \in \mathbb{R}$ satisfy for all $i \in \{1,\dots,d\}$ that $Q_{i,i} = 1$, $\forall j \in \{1,\dots,d\} \setminus \{i\}\colon Q_{i,j} = 0.75$, $\mathfrak{S}^*\mathfrak{S} = Q$, $\tilde{\delta} = \delta + \frac{1}{2}(\beta^2 - (\tilde{\beta})^2)$, and $\tilde{\beta} = \frac{\beta}{2d}\sqrt{d\,(3d+1)}$, let $Y\colon [0,T]\times\Omega \to \mathbb{R}$ be an $\mathscr{F}$-adapted stochastic process with continuous sample paths which satisfies that for all $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that

$$Y_t = \tilde{\xi} + (r-\tilde{\delta})\int_0^t Y_s\,\mathrm{d}s + \tilde{\beta}\int_0^t Y_s\,\mathrm{d}W_s^{0,1,(1)}, \tag{4.104}$$

let $\mathbb{F} = (\mathbb{F}_t)_{t\in[0,T]}$ be the filtration generated by X, let $\mathfrak{F} = (\mathfrak{F}_t)_{t\in[0,T]}$ be the filtration generated by Y, and assume for all $m, j \in \mathbb{N}$, $n \in \{0,1,\dots,N\}$, $i \in \{1,\dots,d\}$, $s \in [0,T]$,

$x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ that $T = 2$, $N = 50$, $M = 1600$, $J_0 = 4\,096\,000$, $J_m = 8192$, $\varepsilon = 10^{-8}$, $\gamma_m = 5\,[10^{-2}\,\mathbb{1}_{[1,400]}(m) + 10^{-3}\,\mathbb{1}_{(400,800]}(m) + 10^{-4}\,\mathbb{1}_{(800,\infty)}(m)]$, $\xi_i = 1$, $\mu(x) = (r - \delta)\,x$, $\sigma(x) = \beta\,\mathrm{diag}(x_1, \ldots, x_d)\,\mathfrak{S}^*$, that

$$\mathcal{X}^{m-1,j,(i)}_{t_n} = \exp\!\Big(\big[r - \delta - \tfrac{1}{2}\beta^2\big]\,t_n + \beta\,\big\langle \varsigma_i, W^{m-1,j}_{t_n}\big\rangle_{\mathbb{R}^d}\Big)\,\xi_i, \tag{4.105}$$

and that

$$g(s, x) = e^{-rs} \max\!\left\{\left[\prod_{k=1}^{d} |x_k|^{1/d}\right] - K, 0\right\}. \tag{4.106}$$

The random variable $\mathcal{P}$ provides approximations for the price

$$\sup\!\left\{\mathbb{E}\big[g(\tau, X_\tau)\big] : \substack{\tau\colon \Omega\to[0,T] \text{ is an} \\ \mathbb{F}\text{-stopping time}}\right\}. \tag{4.107}$$

Table 4.4 shows approximations for the mean of $\mathcal{P}$, for the standard deviation of $\mathcal{P}$, for the real number

$$\sup\!\left\{\mathbb{E}\big[e^{-r\tau}\max\{Y_\tau - K, 0\}\big] : \substack{\tau\colon \Omega\to[0,T] \text{ is an} \\ \mathfrak{F}\text{-stopping time}}\right\}, \tag{4.108}$$

and for the relative L^1-approximation error associated to $\mathcal{P}$, the uncorrected sample standard deviation of the relative approximation error associated to $\mathcal{P}$, and the average runtime in seconds needed for calculating one realisation of $\mathcal{P}$ for $d \in \{3, 20, 100\}$. The approximative calculations of the mean of $\mathcal{P}$, of the standard deviation of $\mathcal{P}$, and of the relative L^1-approximation error associated to $\mathcal{P}$, the computations of the uncorrected sample standard deviation of the relative approximation error associated to $\mathcal{P}$ as well as the computations of the average runtime for calculating one realisation of $\mathcal{P}$ in Table 4.4 each are based on 10 independent realisations of $\mathcal{P}$, which have been obtained from an implementation in PYTHON. Furthermore, in the approximative calculations of the relative approximation error associated to $\mathcal{P}$ the exact value of the price (4.107) has been replaced by the number (4.108) (cf. Proposition 4.5), which has been approximatively calculated using the binomial tree method on M. Smirnov's website [290] with 20 000 nodes. Note that (4.108) corresponds to the price of an American call option on a single stock in the Black–Scholes model with initial stock price $\tilde{\xi}$, interest rate r, dividend rate $\tilde{\delta}$, volatility $\tilde{\beta}$, strike price K, and maturity T.

Dimension d	Mean of $\mathcal{P}$	Standard deviation of $\mathcal{P}$	Price (4.108)	Rel. L^1-approx. err.	Standard deviation of the rel. approx. err.	Runtime in sec. for one realisation of $\mathcal{P}$
3	0.10699	0.00007	0.10719	0.0019	0.0006	92.1
20	0.10007	0.00006	0.10033	0.0026	0.0006	146.9
100	0.09903	0.00006	0.09935	0.0032	0.0006	409.0

Table 4.4: Numerical simulations of the algorithm in Framework 4.2 for pricing the American geometric average call-type option from the example in Subsection 4.3.3.2.2. In the approximative calculations of the relative approximation errors the exact value of the price (4.107) has been replaced by the number (4.108), which has been approximatively calculated using the binomial tree method on M. Smirnov's website [290].

4.3.3.2.3 Another American geometric average call-type example

In this subsection we test the algorithm in Framework 4.2 on the example of pricing an American geometric average call-type option on up to 400 distinguishable stocks in the Black–Scholes model.

Assume Framework 4.3, assume that $d \in \{40, 80, 120, \dots\}$, let $\beta = (\beta_1, \dots, \beta_d) \in \mathbb{R}^d$, $\alpha_1, \dots, \alpha_d \in \mathbb{R}$, $r, \tilde{\beta} \in (0, \infty)$, $K = 95$, $\tilde{\xi} = 100$ satisfy for all $i \in \{1, \dots, d\}$ that $\beta_i = \frac{0.4\, i}{d}$, $\alpha_i = \min\{0.01\, [(i-1) \bmod 40], 0.4 - 0.01\, [(i-1) \bmod 40]\}$, $r = \frac{1}{d}\sum_{i=1}^{d} \alpha_i - \frac{d-1}{2d^2} \|\beta\|_{\mathbb{R}^d}^2 = 0.1 - \frac{0.08}{d^2}(d-1)\big(\frac{d}{3} + \frac{1}{2} + \frac{1}{6d}\big)$, and $\tilde{\beta} = \frac{1}{d}\|\beta\|_{\mathbb{R}^d} = \frac{0.4}{d}\big(\frac{d}{3} + \frac{1}{2} + \frac{1}{6d}\big)^{1/2}$, let $Y \colon [0,T] \times \Omega \to \mathbb{R}$ be an $\mathscr{F}$-adapted stochastic process with continuous sample paths which satisfies that for all $t \in [0, T]$ it holds $\mathbb{P}$-a.s. that

$$Y_t = \tilde{\xi} + r \int_0^t Y_s \, \mathrm{d}s + \tilde{\beta} \int_0^t Y_s \, \mathrm{d}W_s^{0,1,(1)}, \tag{4.109}$$

let $\mathbb{F} = (\mathbb{F}_t)_{t \in [0,T]}$ be the filtration generated by X, let $\mathfrak{F} = (\mathfrak{F}_t)_{t \in [0,T]}$ be the filtration generated by Y, and assume for all $m, j \in \mathbb{N}$, $n \in \{0, 1, \dots, N\}$, $i \in \{1, \dots, d\}$, $s \in [0, T]$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ that $T = 3$, $N = 50$, $M = 1500$, $J_0 = 4\,096\,000$, $J_m = 8192$, $\varepsilon = 10^{-8}$, $\gamma_m = 5\,[10^{-2}\, \mathbb{1}_{[1, M/3]}(m) + 10^{-3}\, \mathbb{1}_{(M/3, 2M/3]}(m) + 10^{-4}\, \mathbb{1}_{(2M/3, \infty)}(m)]$, $\xi_i = 100$, $\mu(x) = (\alpha_1 x_1, \dots, \alpha_d x_d)$, $\sigma(x) = \operatorname{diag}(\beta_1 x_1, \dots, \beta_d x_d)$, that

$$\mathcal{X}_{t_n}^{m-1,j,(i)} = \exp\!\big(\big[\alpha_i - \tfrac{1}{2}(\beta_i)^2\big] t_n + \beta_i\, W_{t_n}^{m-1,j,(i)}\big)\, \xi_i, \tag{4.110}$$

and that

$$g(s, x) = e^{-rs} \max\!\left\{ \left[\prod_{k=1}^{d} |x_k|^{1/d} \right] - K, 0 \right\}. \tag{4.111}$$

The random variable $\mathcal{P}$ provides approximations for the price

$$\sup\!\left\{ \mathbb{E}\big[g(\tau, X_\tau)\big] : \substack{\tau \colon \Omega \to [0,T] \text{ is an} \\ \mathbb{F}\text{-stopping time}} \right\}. \tag{4.112}$$

In Table 4.5 we show approximations for the mean of $\mathcal{P}$, for the standard deviation of $\mathcal{P}$, for the real number

$$\mathbb{E}\big[e^{-rT} \max\{Y_T - K, 0\}\big], \tag{4.113}$$

and for the relative L^1-approximation error associated to $\mathcal{P}$, the uncorrected sample standard deviation of the relative approximation error associated to $\mathcal{P}$, and the average runtime in seconds needed for calculating one realisation of $\mathcal{P}$ for $d \in \{40, 80, 120, 160, 200, 400\}$. The approximative calculations of the mean of $\mathcal{P}$, of the standard deviation of $\mathcal{P}$, and of the relative L^1-approximation error associated to $\mathcal{P}$, the computations of the uncorrected sample standard deviation of the relative approximation error associated to $\mathcal{P}$ as well as the computations of the average runtime for calculating one realisation of $\mathcal{P}$ in Table 4.5 each are based on 10 independent realisations of $\mathcal{P}$, which have been obtained from an implementation in Python. Moreover, in the approximative calculations of the relative approximation error associated to $\mathcal{P}$ the exact value of the price (4.112) has been replaced by the real number

$$\sup\!\left\{ \mathbb{E}\big[e^{-r\tau} \max\{Y_\tau - K, 0\}\big] : \substack{\tau \colon \Omega \to [0,T] \text{ is an} \\ \mathfrak{F}\text{-stopping time}} \right\} \tag{4.114}$$

Dimension d	Mean of $\mathcal{P}$	Standard deviation of $\mathcal{P}$	Price (4.113)	Rel. L^1-approx. err.	Standard deviation of the rel. approx. err.	Runtime in sec. for one realisation of $\mathcal{P}$
40	23.6877	0.0030	23.6883	0.00012	0.00004	196.3
80	23.7235	0.0020	23.7235	0.00006	0.00005	323.3
120	23.7360	0.0019	23.7357	0.00006	0.00005	442.7
160	23.7415	0.0007	23.7419	0.00003	0.00002	569.9
200	23.7451	0.0014	23.7456	0.00005	0.00004	692.9
400	23.7528	0.0009	23.7531	0.00004	0.00002	1434.7

Table 4.5: Numerical simulations of the algorithm in Framework 4.2 for pricing the American geometric average call-type option from the example in Subsection 4.3.3.2.3. In the approximative calculations of the relative approximation errors the exact value of the price (4.112) has been replaced by the number (4.113), which has been approximatively computed in MATLAB.

(cf. Proposition 4.5). It is well-known (cf., e.g., Shreve [286, Corollary 8.5.3]) that the number (4.114) is equal to the number (4.113), which has been approximatively computed in MATLAB R2017b using Lemma 4.6 above. Note that (4.114) corresponds to the price of an American call option on a single stock in the Black–Scholes model with initial stock price $\tilde{\xi}$, interest rate r, volatility $\tilde{\beta}$, strike price K, and maturity T, while (4.113) corresponds to the price of a European call option on a single stock in the Black–Scholes model with initial stock price $\tilde{\xi}$, interest rate r, volatility $\tilde{\beta}$, strike price K, and maturity T.

4.3.4 Examples without known one-dimensional representation

In Subsection 4.3.3 above numerical results for examples with a one-dimensional representation can be found. We test in this subsection several examples where such a representation is not known.

4.3.4.1 Max-call options

4.3.4.1.1 A Bermudan max-call standard benchmark example

In this subsection we test the algorithm in Framework 4.2 on the example of pricing a Bermudan max-call option on up to 500 stocks in the Black–Scholes model (cf. Becker, Cheridito, & Jentzen [28, Subsection 4.1]). In the case of up to five underlying stocks this example is a standard benchmark example in the literature (cf., e.g., [60, Subsection 5.4], [232, Subsection 8.1], [6, Section 4], [164, Subsection 5.1], [270, Subsection 4.3], [126, Subsection 3.9], [41, Subsection 4.2], [58, Subsection 5.3], [36, Subsection 6.1], [188, Subsection 4.1], [276, Subsection 7.2], [38, Subsection 6.1], [223, Subsection 5.2.1]).

Assume Framework 4.3, let $r = 0.05 = 5\%$, $\delta = 0.1 = 10\%$, $\beta = 0.2 = 20\%$, $K = 100$, let $\mathbb{F} = (\mathbb{F}_t)_{t\in[0,T]}$ be the filtration generated by X, and assume for all $m, j \in \mathbb{N}$, $n \in \{0, 1, \ldots, N\}$, $i \in \{1, \ldots, d\}$, $s \in [0, T]$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ that $T = 3$, $N = 9$, $M = 3000 + d$, $J_0 = 4\,096\,000$, $J_m = 8192$, $\varepsilon = 0.1$, $\gamma_m = 5\,[10^{-2}\,\mathbb{1}_{[1,500+d/5]}(m) +$

$10^{-3}\,\mathbb{1}_{(500+d/5,1500+3d/5]}(m) + 10^{-4}\,\mathbb{1}_{(1500+3d/5,\infty)}(m)]$, $\xi_i = \xi_1$, $\mu(x) = (r-\delta)\,x$, $\sigma(x) = \beta\,\mathrm{diag}(x_1,\ldots,x_d)$, that

$$\mathcal{X}_{t_n}^{m-1,j,(i)} = \exp\!\big(\big[r-\delta-\tfrac{1}{2}\beta^2\big]t_n + \beta\, W_{t_n}^{m-1,j,(i)}\big)\,\xi_1, \tag{4.115}$$

and that

$$g(s,x) = e^{-rs}\max\!\big\{\max\{x_1,\ldots,x_d\}-K,0\big\}. \tag{4.116}$$

The random variable $\mathcal{P}$ provides approximations for the price

$$\sup\!\Big\{\mathbb{E}\big[g(\tau,X_\tau)\big] : \substack{\tau\colon \Omega\to\{t_0,t_1,\ldots,t_N\}\text{ is an}\\ (\mathbb{F}_t)_{t\in\{t_0,t_1,\ldots,t_N\}}\text{-stopping time}}\Big\}. \tag{4.117}$$

In Table 4.6 we show approximations for the mean and for the standard deviation of $\mathcal{P}$, binomial approximations as well as 95% confidence intervals for the price (4.117) according to Andersen & Broadie [6, Table 2 in Section 4] (where available), 95% confidence intervals for the price (4.117) according to Broadie & Cao [58, Table 3 in Subsection 5.3] (where available), and the average runtime in seconds needed for calculating one realisation of $\mathcal{P}$ for $(d,\xi_1)\in\{2,3,5\}\times\{90,100,110\}$. In addition, we list approximations for the mean and for the standard deviation of $\mathcal{P}$ and the average runtime in seconds needed for calculating one realisation of $\mathcal{P}$ for $(d,\xi_1)\in\{10,20,30,50,100,200,500\}\times\{90,100,110\}$ in Table 4.7. The approximative calculations of the mean and of the standard deviation of $\mathcal{P}$ as well as the computations of the average runtime for calculating one realisation of $\mathcal{P}$ in Tables 4.6 and 4.7 each are based on 10 independent realisations of $\mathcal{P}$, which have been obtained from an implementation in PYTHON.

4.3.4.1.2 A high-dimensional Bermudan max-call benchmark example

In this subsection we test the algorithm in Framework 4.2 on the example of pricing the Bermudan max-call option from the example in Subsection 4.3.4.1.1 in a case with 5000 underlying stocks. All PYTHON codes corresponding to this example were run in single precision (float32) on a NVIDIA Tesla P100 GPU with 1328 MHz core clock and 16 GB HBM2 memory with 1408 MHz clock rate.

Assume Framework 4.3, let $r = 0.05 = 5\%$, $\delta = 0.1 = 10\%$, $\beta = 0.2 = 20\%$, $K = 100$, let $\mathbb{F} = (\mathbb{F}_t)_{t\in[0,T]}$ be the filtration generated by X, and assume for all $m,j\in\mathbb{N}$, $n\in\{0,1,\ldots,N\}$, $i\in\{1,\ldots,d\}$, $s\in[0,T]$, $x=(x_1,\ldots,x_d)\in\mathbb{R}^d$ that $T=3$, $d=5000$, $N=9$, $J_0=2^{20}$, $J_m=1024$, $\varepsilon=10^{-8}$, $\gamma_m = 10^{-2}\,\mathbb{1}_{[1,2000]}(m) + 10^{-3}\,\mathbb{1}_{(2000,4000]}(m) + 10^{-4}\,\mathbb{1}_{(4000,\infty)}(m)$, $\xi_i = 100$, $\mu(x) = (r-\delta)\,x$, $\sigma(x) = \beta\,\mathrm{diag}(x_1,\ldots,x_d)$, that

$$\mathcal{X}_{t_n}^{m-1,j,(i)} = \exp\!\big(\big[r-\delta-\tfrac{1}{2}\beta^2\big]t_n + \beta\, W_{t_n}^{m-1,j,(i)}\big)\,\xi_i, \tag{4.118}$$

and that

$$g(s,x) = e^{-rs}\max\!\big\{\max\{x_1,\ldots,x_d\}-K,0\big\}. \tag{4.119}$$

For sufficiently large $M\in\mathbb{N}$ the random variable $\mathcal{P}$ provides approximations for the price

$$\sup\!\Big\{\mathbb{E}\big[g(\tau,X_\tau)\big] : \substack{\tau\colon \Omega\to\{t_0,t_1,\ldots,t_N\}\text{ is an}\\ (\mathbb{F}_t)_{t\in\{t_0,t_1,\ldots,t_N\}}\text{-stopping time}}\Big\}. \tag{4.120}$$

In Table 4.8 we show a realisation of $\mathcal{P}$, a 95% confidence interval for the corresponding realisation of the random variable

$$\Omega\ni \mathbf{w}\mapsto \mathbb{E}\Big[g\Big(\tau^{1,\Theta_M(\mathbf{w}),\mathbb{S}_M(\mathbf{w})},\mathcal{X}^{0,1}_{\tau^{1,\Theta_M(\mathbf{w}),\mathbb{S}_M(\mathbf{w})}}\Big)\Big]\in\mathbb{R}, \tag{4.121}$$

Dimension d	Initial value ξ_1	Mean of $\mathcal{P}$	Standard deviation of $\mathcal{P}$	Binomial value in [6]	95% confidence interval in [6]	95% confidence interval in [58]	Runtime in sec. for one realisation of $\mathcal{P}$
2	90	8.072	0.005	8.075	[8.053, 8.082]	–	31.3
2	100	13.899	0.008	13.902	[13.892, 13.934]	–	31.7
2	110	21.344	0.006	21.345	[21.316, 21.359]	–	31.7
3	90	11.275	0.005	11.29	[11.265, 11.308]	–	32.5
3	100	18.687	0.006	18.69	[18.661, 18.728]	–	32.6
3	110	27.560	0.009	27.58	[27.512, 27.663]	–	32.5
5	90	16.628	0.010	–	[16.602, 16.655]	[16.620, 16.653]	33.3
5	100	26.144	0.008	–	[26.109, 26.292]	[26.115, 26.164]	32.9
5	110	36.763	0.011	–	[36.704, 36.832]	[36.710, 36.798]	33.3

Table 4.6: Numerical simulations of the algorithm in Framework 4.2 for pricing the Bermudan max-call option from the example in Subsection 4.3.4.1.1 for $d \in \{2, 3, 5\}$.

Dimension d	Initial value ξ_1	Mean of $\mathcal{P}$	Standard deviation of $\mathcal{P}$	Runtime in sec. for one realisation of $\mathcal{P}$
10	90	26.200	0.010	35.5
10	100	38.278	0.010	35.5
10	110	50.817	0.011	35.5
20	90	37.697	0.011	42.6
20	100	51.569	0.008	42.6
20	110	65.514	0.010	42.6
30	90	44.822	0.008	49.2
30	100	59.521	0.010	49.2
30	110	74.231	0.010	49.2
50	90	53.897	0.008	63.4
50	100	69.574	0.012	63.3
50	110	85.256	0.013	63.3
100	90	66.361	0.016	101.1
100	100	83.386	0.009	101.2
100	110	100.429	0.014	101.2
200	90	78.996	0.009	179.1
200	100	97.411	0.011	179.1
200	110	115.827	0.009	179.0
500	90	95.976	0.007	507.7
500	100	116.249	0.013	507.4
500	110	136.541	0.005	507.8

Table 4.7: Numerical simulations of the algorithm in Framework 4.2 for pricing the Bermudan max-call option from the example in Subsection 4.3.4.1.1 for $d \in \{10, 20, 30, 50, 100, 200, 500\}$.

the corresponding realisation of the relative approximation error associated to $\mathcal{P}$, and the runtime in seconds need or calculating the realisation of $\mathcal{P}$ for $M \in \{0, 250, 500, \ldots, 2000\} \cup \{6000\}$. In addition, Figure 4.1 depicts a realisation of the relative approximation error associated to $\mathcal{P}$ against $M \in \{0, 10, 20, \ldots, 2000\}$. For each case the 95% confidence interval for the realisation of the random variable (4.121) in Table 4.8 has been computed based on the corresponding realisation of $\mathcal{P}$, the corresponding sample standard deviation, and the 0.975 quantile of the standard normal distribution (cf., e.g., [28, Subsection 3.3]). Moreover, in the approximative calculations of the realisation of the relative approximation error associated to $\mathcal{P}$ in Table 4.8 and Figure 4.1 the exact value of the price (4.120) has been replaced by the value 165.430, which corresponds to a realisation of $\mathcal{P}$ with $M = 6000$ (cf. Table 4.8).

Number of steps M	Realisation of $\mathcal{P}$	95% confidence interval	Rel. approx. error	Runtime in sec.
0	106.711	[106.681, 106.741]	0.35495	157.3
250	132.261	[132.170, 132.353]	0.20050	271.7
500	156.038	[155.975, 156.101]	0.05677	386.0
750	103.764	[103.648, 103.879]	0.37276	500.4
1000	161.128	[161.065, 161.191]	0.02601	614.3
1250	162.756	[162.696, 162.816]	0.01616	728.8
1500	164.498	[164.444, 164.552]	0.00563	842.8
1750	163.858	[163.803, 163.913]	0.00950	957.3
2000	165.452	[165.400, 165.505]	0.00014	1071.9
6000	165.430	[165.378, 165.483]	0.00000	2899.5

Table 4.8: Numerical simulations of the algorithm in Framework 4.2 for pricing the Bermudan max-call option on 5000 stocks from the example in Subsection 4.3.4.1.2. In the approximative calculations of the relative approximation error the exact value of the price (4.120) has been replaced by the value 165.430, which corresponds to a realisation of $\mathcal{P}$ with $M = 6000$.

4.3.4.1.3 Another Bermudan max-call example

In this subsection we test the algorithm in Framework 4.2 on the example of pricing a Bermudan max-call option for different maturities and strike prices on up to 400 *correlated* stocks, that do not pay dividends, in the Black–Scholes model. This example is taken from Barraquand & Martineau [15, Section VII].

Assume Framework 4.3, let $\tau = {}^{30}\!/_{365}$, $r = 0.05\,\tau = 5\%\cdot\tau$, $\beta = 0.4\sqrt{\tau} = 40\%\cdot\sqrt{\tau}$, $K \in \{35, 40, 45\}$, $Q = (Q_{i,j})_{(i,j)\in\{1,\ldots,d\}^2}$, $\mathfrak{S} = (\varsigma_1, \ldots, \varsigma_d) \in \mathbb{R}^{d\times d}$ satisfy for all $i \in \{1, \ldots, d\}$ that $Q_{i,i} = 1$, $\forall\, j \in \{1, \ldots, d\} \setminus \{i\}\colon Q_{i,j} = 0.5$, and $\mathfrak{S}^*\,\mathfrak{S} = Q$, let $\mathbb{F} = (\mathbb{F}_t)_{t\in[0,T]}$ be the filtration generated by X, and assume for all $m, j \in \mathbb{N}$, $n \in \{0, 1, \ldots, N\}$, $i \in \{1, \ldots, d\}$, $s \in [0, T]$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ that $N = 10$, $M = 1600$, $J_0 = 4\,096\,000$, $J_m = 8192$, $\varepsilon = 0.001$, $\gamma_m = 5\,[10^{-2}\,\mathbb{1}_{[1,400]}(m) + 10^{-3}\,\mathbb{1}_{(400,800]}(m) + 10^{-4}\,\mathbb{1}_{(800,\infty)}(m)]$, $\xi_i = 40$, $\mu(x) = r\,x$, $\sigma(x) = \beta\,\mathrm{diag}(x_1, \ldots, x_d)\,\mathfrak{S}^*$, that

$$\mathcal{X}^{m-1,j,(i)}_{t_n} = \exp\!\Big(\big[r - \tfrac{1}{2}\beta^2\big]t_n + \beta\,\big\langle \varsigma_i, W^{m-1,j}_{t_n}\big\rangle_{\mathbb{R}^d}\Big)\,\xi_i, \tag{4.122}$$

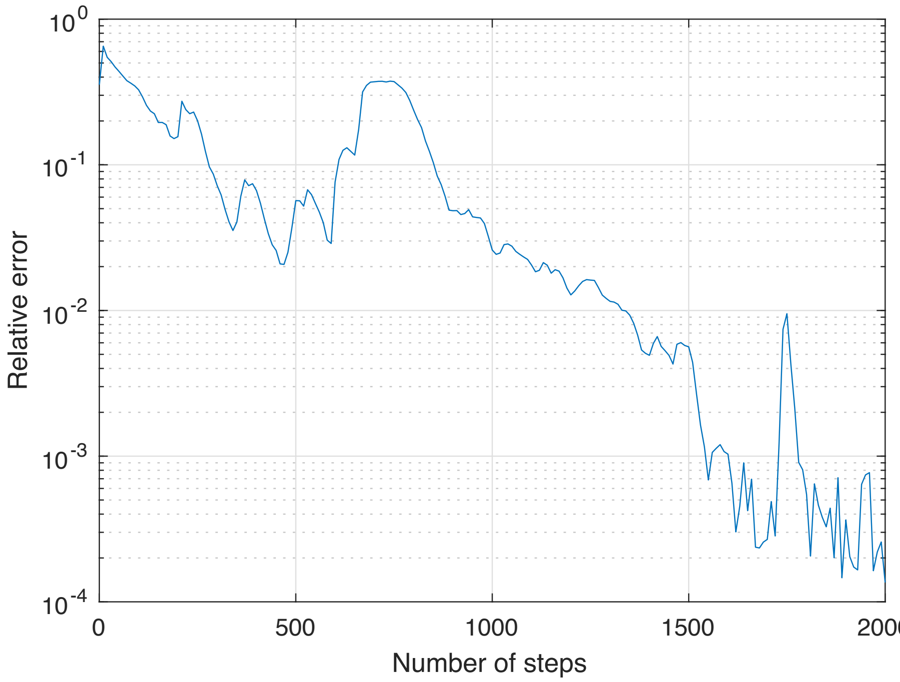

Figure 4.1: Plot of a realisation of the relative approximation error $\frac{|\mathcal{P}-165.430|}{165.430}$ against $M \in \{0, 10, 20, \dots, 2000\}$ in the case of the Bermudan max-call option on 5000 stocks from the example in Subsection 4.3.4.1.2.

and that

$$g(s, x) = e^{-rs} \max\bigl\{\max\{x_1, \dots, x_d\} - K, 0\bigr\}. \tag{4.123}$$

The random variable $\mathcal{P}$ provides approximations for the price

$$\sup\left\{\mathbb{E}\bigl[g(\tau, X_\tau)\bigr] : \substack{\tau\colon \Omega \to \{t_0, t_1, \dots, t_N\} \text{ is an} \\ (\mathbb{F}_t)_{t \in \{t_0, t_1, \dots, t_N\}}\text{-stopping time}}\right\}. \tag{4.124}$$

In Table 4.9 we show approximations for the mean and for the standard deviation of $\mathcal{P}$, Monte Carlo approximations for the European max-call option price

$$\mathbb{E}\bigl[g(T, X_T)\bigr] \tag{4.125}$$

corresponding to (4.124), approximations for the price (4.124) according to [15, Table 4 in Section VII] (where available), and the average runtime in seconds needed for calculating one realisation of $\mathcal{P}$ for

$$(d, T, K) \in \left\{\begin{array}{ccc} (10, 1, 35), & (10, 1, 40), & (10, 1, 45), \\ (10, 4, 35), & (10, 4, 40), & (10, 4, 45), \\ (10, 7, 35), & (10, 7, 40), & (10, 7, 45), \\ (400, 12, 35), & (400, 12, 40), & (400, 12, 45) \end{array}\right\}. \tag{4.126}$$

The approximative calculations of the mean and of the standard deviation of $\mathcal{P}$ as well as the computations of the average runtime for calculating one realisation of $\mathcal{P}$ in Table 4.9 each are based on 10 independent realisations of $\mathcal{P}$, which have been obtained from an implementation in PYTHON. Furthermore, the Monte Carlo approximations for the European price (4.125) in Table 4.9 each are calculated in double precision (float64) and are based on $2 \cdot 10^{10}$ independent realisations of the random variable $\Omega \ni \omega \mapsto g(T, X_T(\omega)) \in \mathbb{R}$.

Dimension d	Maturity T	Strike price K	Mean of $\mathcal{P}$	Standard deviation of $\mathcal{P}$	European price (4.125)	Price in [15]	Runtime in sec. for one realisation of $\mathcal{P}$
10	1	35	10.365	0.002	10.365	10.36	24.4
10	1	40	5.540	0.002	5.540	5.54	24.3
10	1	45	1.897	0.001	1.896	1.90	24.3
10	4	35	16.519	0.003	16.520	16.53	24.3
10	4	40	11.869	0.007	11.870	11.87	24.3
10	4	45	7.801	0.003	7.804	7.81	24.3
10	7	35	20.913	0.007	20.916	20.92	24.3
10	7	40	16.374	0.008	16.374	16.38	24.3
10	7	45	12.271	0.006	12.277	12.28	24.3
400	12	35	55.712	0.009	55.714	–	247.7
400	12	40	50.969	0.020	50.964	–	247.8
400	12	45	46.233	0.010	46.234	–	247.6

Table 4.9: Numerical simulations of the algorithm in Framework 4.2 for pricing the Bermudan max-call option from the example in Subsection 4.3.4.1.3.

4.3.4.2 A strangle spread basket option

In this subsection we test the algorithm in Framework 4.2 on the example of pricing an American strangle spread basket option on five correlated stocks in the Black–Scholes model. This example is taken from Kohler, Krzyżak, & Todorovic [207, Section 4] (cf. also Kohler [204, Section 3] and Kohler, Krzyżak, & Walk [208, Section 4]).

Assume Framework 4.3, let $r = 0.05 = 5\%$, $K_1 = 75$, $K_2 = 90$, $K_3 = 110$, $K_4 = 125$, let $\mathfrak{S} = (\varsigma_1, \ldots, \varsigma_5) \in \mathbb{R}^{5\times 5}$ be given by

$$\mathfrak{S} = \begin{pmatrix} 0.3024 & 0.1354 & 0.0722 & 0.1367 & 0.1641 \\ 0.1354 & 0.2270 & 0.0613 & 0.1264 & 0.1610 \\ 0.0722 & 0.0613 & 0.0717 & 0.0884 & 0.0699 \\ 0.1367 & 0.1264 & 0.0884 & 0.2937 & 0.1394 \\ 0.1641 & 0.1610 & 0.0699 & 0.1394 & 0.2535 \end{pmatrix}, \tag{4.127}$$

let $\mathbb{F} = (\mathbb{F}_t)_{t\in[0,T]}$ be the filtration generated by X, and assume for all $m, j \in \mathbb{N}$, $n \in \{0, 1, \ldots, N\}$, $i \in \{1, \ldots, d\}$, $s \in [0, T]$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ that $T = 1$, $d = 5$, $N = 48$, $M = 750$, $J_0 = 4\,096\,000$, $J_m = 8192$, $\varepsilon = 10^{-8}$, $\gamma_m = 5\,[10^{-2}\,\mathbb{1}_{[1,250]}(m) + 10^{-3}\,\mathbb{1}_{(250,500]}(m) + 10^{-4}\,\mathbb{1}_{(500,\infty)}(m)]$, $\xi_i = 100$, $\mu(x) = r\,x$, $\sigma(x) = \operatorname{diag}(x_1, \ldots, x_d)\,\mathfrak{S}^*$, that

$$\mathcal{X}^{m-1,j,(i)}_{t_n} = \exp\Big(\big[r - \tfrac{1}{2}\|\varsigma_i\|^2_{\mathbb{R}^d}\big]t_n + \langle \varsigma_i, W^{m-1,j}_{t_n}\rangle_{\mathbb{R}^d}\Big)\,\xi_i, \tag{4.128}$$

and that

$$\begin{aligned} g(s,x) = {}& -e^{-rs}\max\left\{K_1 - \frac{1}{d}\left[\sum_{k=1}^{d} x_k\right], 0\right\} + e^{-rs}\max\left\{K_2 - \frac{1}{d}\left[\sum_{k=1}^{d} x_k\right], 0\right\} \\ & + e^{-rs}\max\left\{\frac{1}{d}\left[\sum_{k=1}^{d} x_k\right] - K_3, 0\right\} - e^{-rs}\max\left\{\frac{1}{d}\left[\sum_{k=1}^{d} x_k\right] - K_4, 0\right\}. \end{aligned} \tag{4.129}$$

The random variable $\mathcal{P}$ provides approximations for the price

$$\sup\left\{\mathbb{E}\big[g(\tau, X_\tau)\big] : \substack{\tau\colon \Omega\to[0,T] \text{ is an} \\ \mathbb{F}\text{-stopping time}}\right\}. \tag{4.130}$$

Table 4.10 shows approximations for the mean and for the standard deviation of $\mathcal{P}$, a lower bound for the price (4.130) according to Kohler, Krzyżak, & Todorovic [207, Figure 4.5 in Section 4] (cf. also Kohler [204, Figure 2 in Section 3] and, for an upper bound for the price (4.130), Kohler, Krzyżak, & Walk [208, Figure 4.2 in Section 4]), and the average runtime in seconds needed for calculating one realisation of $\mathcal{P}$. Since the mean of $\mathcal{P}$ is also a lower bound for the price (4.130), a higher value indicates a better approximation for the price (4.130) (cf. Table 4.10). The approximative calculations of the mean and of the standard deviation of $\mathcal{P}$ as well as the computation of the average runtime for calculating one realisation of $\mathcal{P}$ in Table 4.10 each are based on 10 independent realisations of $\mathcal{P}$, which have been obtained from an implementation in Python.

Mean of $\mathcal{P}$	Standard deviation of $\mathcal{P}$	Lower bound in [207]	Runtime in sec. for one realisation of $\mathcal{P}$
11.794	0.004	11.75	46.4

Table 4.10: Numerical simulations of the algorithm in Framework 4.2 for pricing the American strangle spread basket option from the example in Subsection 4.3.4.2.

4.3.4.3 A put basket option in Dupire's local volatility model

In this subsection we test the algorithm in Framework 4.2 on the example of pricing an American put basket option on five stocks in Dupire's local volatility model. This example is taken from Labart & Lelong [219, Subsection 6.3] with the modification that we also consider the case where the underlying stocks do not pay any dividends.

Assume Framework 4.3, let $L = 10$, $r = 0.05 = 5\%$, $\delta \in \{0\%, 10\%\}$, $K = 100$, assume for all $i \in \{1, \dots, d\}$, $x \in \mathbb{R}^d$ that $\xi_i = 100$ and $\mu(x) = (r - \delta)\, x$, let $\beta\colon [0,T] \times \mathbb{R} \to \mathbb{R}$ and $\boldsymbol{\sigma}\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}^{d\times d}$ be the functions which satisfy for all $t \in [0,T]$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ that

$$\beta(t, x_1) = 0.6\, e^{-0.05\sqrt{t}}\big(1.2 - e^{-0.1\, t - 0.001(e^{rt}x_1 - \xi_1)^2}\big)\, x_1 \tag{4.131}$$

and $\boldsymbol{\sigma}(t,x) = \operatorname{diag}(\beta(t,x_1), \beta(t,x_2), \dots, \beta(t,x_d))$, let $S = (S^{(1)}, \dots, S^{(d)})\colon [0,T]\times\Omega \to \mathbb{R}^d$ be an $\mathscr{F}$-adapted stochastic process with continuous sample paths which satisfies that for all $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that

$$S_t = \xi + \int_0^t \mu(S_s)\, \mathrm{d}s + \int_0^t \boldsymbol{\sigma}(s, S_s)\, \mathrm{d}W^{0,1}_s, \tag{4.132}$$

let $\mathcal{Y}^{m,j} = (\mathcal{Y}^{m,j,(1)}, \dots, \mathcal{Y}^{m,j,(d)})\colon [0,T]\times\Omega \to \mathbb{R}^d$, $j \in \mathbb{N}$, $m \in \mathbb{N}_0$, be the stochastic processes which satisfy for all $m \in \mathbb{N}_0$, $j \in \mathbb{N}$, $\ell \in \{0, 1, \dots, L-1\}$, $t \in \big[\frac{\ell T}{L}, \frac{(\ell+1)T}{L}\big]$, $i \in \{1, \dots, d\}$ that $\mathcal{Y}^{m,j,(i)}_0 = \log(\xi_i)$ and

$$\begin{aligned} \mathcal{Y}^{m,j,(i)}_t = \mathcal{Y}^{m,j(i)}_{\ell T/L} &+ \big(t - \tfrac{\ell T}{L}\big)\Big(r - \delta - \tfrac{1}{2}\big[\beta\big(\tfrac{\ell T}{L}, \exp\big(\mathcal{Y}^{m,j,(i)}_{\ell T/L}\big)\big)\big]^2\Big) \\ &+ \big(\tfrac{tL}{T} - \ell\big)\beta\big(\tfrac{\ell T}{L}, \exp\big(\mathcal{Y}^{m,j,(i)}_{\ell T/L}\big)\big)\big(W^{m,j,(i)}_{(\ell+1)T/L} - W^{m,j,(i)}_{\ell T/L}\big), \end{aligned} \tag{4.133}$$

let $\mathbb{F} = (\mathbb{F}_t)_{t\in[0,T]}$ be the filtration generated by S, let $\mathfrak{F} = (\mathfrak{F}_t)_{t\in[0,T]}$ be the filtration generated by $\mathcal{Y}^{0,1}$, and assume for all $m, j \in \mathbb{N}$, $n \in \{0, 1, \ldots, N\}$, $s \in [0, T]$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ that $T = 1$, $d = 5$, $M = 1200$, $\mathcal{X}^{m-1,j}_{t_n} = \mathcal{Y}^{m-1,j}_{t_n}$, $J_0 = 4\,096\,000$, $J_m = 8192$, $\varepsilon = 10^{-8}$, $\gamma_m = 5\,[10^{-2}\,\mathbb{1}_{[1,400]}(m) + 10^{-3}\,\mathbb{1}_{(400,800]}(m) + 10^{-4}\,\mathbb{1}_{(800,\infty)}(m)]$, and

$$g(s, x) = e^{-rs} \max\left\{K - \frac{1}{d}\left[\sum_{i=1}^{d} \exp(x_i)\right], 0\right\}. \tag{4.134}$$

The random variable $\mathcal{P}$ provides approximations for the price

$$\sup\left\{\mathbb{E}\big[g(\tau, \mathcal{Y}^{0,1}_\tau)\big] : \substack{\tau\colon \Omega\to[0,T] \text{ is an} \\ \mathfrak{F}\text{-stopping time}}\right\}, \tag{4.135}$$

which, in turn, is an approximation for the price

$$\sup\left\{\mathbb{E}\left[e^{-r\tau} \max\left\{K - \frac{1}{d}\left[\sum_{i=1}^{d} S^{(i)}_\tau\right], 0\right\}\right] : \substack{\tau\colon \Omega\to[0,T] \text{ is an} \\ \mathbb{F}\text{-stopping time}}\right\}. \tag{4.136}$$

In Table 4.11 we show approximations for the mean and for the standard deviation of $\mathcal{P}$ and the average runtime in seconds needed for calculating one realisation of $\mathcal{P}$ for $(\delta, N) \in \{0\%, 10\%\} \times \{5, 10, 50, 100\}$. For each case the calculations of the results in Table 4.11 are based on 10 independent realisations of $\mathcal{P}$, which have been obtained from an implementation in Python. According to [219, Subsection 6.3], the value 6.30 is an approximation for a to (4.135) corresponding price in the case $\delta = 10\%$. Furthermore, the to (4.135) corresponding European put basket option price $\mathbb{E}\big[g(T, \mathcal{Y}^{0,1}_T)\big]$ has been approximatively calculated using a Monte Carlo approximation based on 10^{10} realisations of the random variable $\Omega \ni \omega \mapsto g(T, \mathcal{Y}^{0,1}_T(\omega)) \in \mathbb{R}$, which resulted in the value 1.741 in the case $\delta = 0\%$ and in the value 6.304 in the case $\delta = 10\%$.

Dividend rate δ	Time discretisation parameter N	Mean of $\mathcal{P}$	Standard deviation of $\mathcal{P}$	Runtime in sec. for one realisation of $\mathcal{P}$
0%	5	1.935	0.001	12.7
0%	10	1.978	0.001	21.3
0%	50	1.975	0.002	69.2
0%	100	1.971	0.002	137.5
10%	5	6.301	0.004	12.8
10%	10	6.303	0.003	21.4
10%	50	6.304	0.003	69.2
10%	100	6.303	0.004	137.5

Table 4.11: Numerical simulations of the algorithm in Framework 4.2 for pricing the American put basket option in Dupire's local volatility model from the example in Subsection 4.3.4.3. The corresponding European put basket option price is approximately equal to the value 1.741 in the case $\delta = 0\%$ and to the value 6.304 in the case $\delta = 10\%$.

4.3.4.4 A path-dependent financial derivative

In this subsection we test the algorithm in Framework 4.2 on the example of pricing a specific path-dependent financial derivative contingent on prices of a single underlying stock in the Black–Scholes model, which is formulated as a 100-dimensional optimal stopping problem. This example is taken from Tsitsiklis & Van Roy [294, Section IV] with the modification that we consider a finite instead of an infinite time horizon.

Assume Framework 4.3, let $r = 0.0004 = 0.04\%$, $\beta = 0.02 = 2\%$, let $\mathcal{W}^{m,j} \colon [0,\infty) \times \Omega \to \mathbb{R}$, $j \in \mathbb{N}$, $m \in \mathbb{N}_0$, be independent $\mathbb{P}$-standard Brownian motions with continuous sample paths, let $S^{m,j} \colon [-100,\infty) \times \Omega \to \mathbb{R}$, $j \in \mathbb{N}$, $m \in \mathbb{N}_0$, and $\mathcal{Y}^{m,j} \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^{100}$, $j \in \mathbb{N}$, $m \in \mathbb{N}_0$, be the stochastic processes which satisfy for all $m, n \in \mathbb{N}_0$, $j \in \mathbb{N}$, $t \in [-100,\infty)$ that $S^{m,j}_t = \exp\!\big(\big[r - \tfrac{1}{2}\beta^2\big](t+100) + \beta\, \mathcal{W}^{m,j}_{t+100}\big)\, \xi_1$ and

$$\begin{aligned} \mathcal{Y}^{m,j}_n &= \Big(\tfrac{S^{m,j}_{n-99}}{S^{m,j}_{n-100}}, \tfrac{S^{m,j}_{n-98}}{S^{m,j}_{n-100}}, \ldots, \tfrac{S^{m,j}_{n}}{S^{m,j}_{n-100}}\Big) \\ &= \Big(\exp\!\big(\big[r - \tfrac{1}{2}\beta^2\big] + \beta\,[\mathcal{W}^{m,j}_{n+1} - \mathcal{W}^{m,j}_{n}]\big), \exp\!\big(2\,\big[r - \tfrac{1}{2}\beta^2\big] + \beta\,[\mathcal{W}^{m,j}_{n+2} - \mathcal{W}^{m,j}_{n}]\big), \\ &\qquad \ldots, \exp\!\big(100\,\big[r - \tfrac{1}{2}\beta^2\big] + \beta\,[\mathcal{W}^{m,j}_{n+100} - \mathcal{W}^{m,j}_{n}]\big)\Big), \end{aligned} \tag{4.137}$$

let $\mathbb{F} = (\mathbb{F}_n)_{n\in\mathbb{N}_0}$ be the filtration generated by $\mathcal{Y}^{0,1}$, and assume for all $m, j \in \mathbb{N}$, $n \in \{0,1,\ldots,N\}$, $s \in [0,T]$, $x = (x_1,\ldots,x_d) \in \mathbb{R}^d$ that $T \in \mathbb{N}$, $d = 100$, $N = T$, $M = 1200\, \mathbb{1}_{[1,150]}(T) + 1500\, \mathbb{1}_{(150,250]}(T) + 3000\, \mathbb{1}_{(250,\infty)}(T)$, $\mathcal{X}^{m-1,j}_n = \mathcal{Y}^{m-1,j}_n$, $J_0 = 4\,096\,000$, $J_m = 8192\, \mathbb{1}_{[1,150]}(T) + 4096\, \mathbb{1}_{(150,250]}(T) + 512\, \mathbb{1}_{(250,\infty)}(T)$, $\varepsilon = 10^{-8}$, $\gamma_m = 5\,[10^{-2}\, \mathbb{1}_{[1,M/3]}(m) + 10^{-3}\, \mathbb{1}_{(M/3,2M/3]}(m) + 10^{-4}\, \mathbb{1}_{(2M/3,\infty)}(m)]$, and $g(s,x) = e^{-rs}\, x_{100}$. The random variable $\mathcal{P}$ provides approximations for the real number

$$\sup\!\left\{ \mathbb{E}\!\left[e^{-r\tau} \tfrac{S^{0,1}_{\tau}}{S^{0,1}_{\tau-100}} \right] : \substack{\tau\colon \Omega \to \{0,1,\ldots,T\} \text{ is an} \\ (\mathbb{F}_n)_{n\in\{0,1,\ldots,T\}}\text{-stopping time}} \right\}. \tag{4.138}$$

In Table 4.12 we show approximations for the mean and for the standard deviation of $\mathcal{P}$ and the average runtime in seconds needed for calculating one realisation of $\mathcal{P}$ for $T \in \{100, 150, 200, 250, 1000\}$. For each case the calculations of the results in Table 4.12 are based on 10 independent realisations of $\mathcal{P}$, which have been obtained from an implementation in Python. Note that in this example time is measured in days and that, roughly speaking, (4.138) corresponds to the price of a financial derivative which, if the

Time horizon T	Mean of $\mathcal{P}$	Standard deviation of $\mathcal{P}$	Runtime in sec. for one realisation of $\mathcal{P}$
100	1.2721	0.0001	475.5
150	1.2821	0.0001	724.8
200	1.2894	0.0002	653.1
250	1.2959	0.0001	838.7
1000	1.3002	0.0006	1680.1

Table 4.12: Numerical simulations of the algorithm in Framework 4.2 for pricing the path-dependent financial derivative from the example in Subsection 4.3.4.4. According to [294, Subsection IV.D], the value 1.282 is a lower bound for the price (4.139).

holder decides to exercise, pays off the amount given by the ratio between the current underlying stock price and the underlying stock price 100 days before (cf. [294, Section IV] for more details). According to [294, Subsection IV.D], the value 1.282 is a lower bound for the price

$$\sup\left\{\mathbb{E}\left[e^{-r\tau}\frac{S^{0,1}_{\tau}}{S^{0,1}_{\tau-100}}\right] : {\scriptstyle \tau\colon \Omega\to\mathbb{N}_0 \text{ is an} \atop \mathbb{F}\text{-stopping time}}\right\}, \tag{4.139}$$

which corresponds to the price (4.138) in the case of an infinite time horizon. Since the mean of $\mathcal{P}$ is a lower bound for the price (4.138), which, in turn, is a lower bound for the price (4.139), a higher value indicates a better approximation for the price (4.139). In addition, observe that the price (4.138) is non-decreasing in T. While in our numerical simulations the approximate value of the mean of $\mathcal{P}$ is less or equal than 1.282 for comparatively small time horizons, i.e., for $T \leq 150$, it is already higher for slightly larger time horizons, i.e., for $T \geq 200$ (cf. Table 4.12).

Chapter 5

Overall error analysis for the training of deep neural networks via stochastic gradient descent with random initialisation

The content of this chapter is a slightly modified extract of the preprint Jentzen & Welti [196].

In this chapter we establish an overall error analysis of deep learning based empirical risk minimisation with quadratic loss function in the probabilistically strong sense (cf. Section 1.4 in Chapter 1). The main result of this chapter, Theorem 5.41 in Subsection 5.5.2, provides a strong convergence estimate for the overall error arising when the underlying deep neural networks (DNNs) are trained using, for example, a general stochastic approximation algorithm with random initialisation. Theorem 1.4 in Section 1.4 is a consequence of Theorem 5.41 and specialises it, in particular, to the case where stochastic gradient descent (SGD) with random initialisation is the employed optimisation method. Parts of our derivation of Theorems 5.41 and 1.4, respectively, are inspired by Beck, Jentzen, & Kuckuck [27], Berner, Grohs, & Jentzen [47], and Cucker & Smale [91].

This chapter is structured as follows. Section 5.1 recalls some basic definitions related to DNNs and thereby introduces the corresponding notation we use throughout this chapter. In Section 5.2 we examine the approximation error and, in particular, establish a convergence result for the approximation of Lipschitz continuous functions by DNNs. The following section, Section 5.3, contains our strong convergence analysis of the generalisation error. In Section 5.4, in turn, we address the optimisation error and derive in connection with this strong convergence rates for the Minimum Monte Carlo method. Finally, we combine in Section 5.5 a decomposition of the overall error (cf. Subsection 5.5.1) with our results for the different error sources from Sections 5.2, 5.3, and 5.4 to prove strong convergence results for the overall error. The employed optimisation method is initially allowed to be a general stochastic approximation algorithm with random initialisation (cf. Subsection 5.5.2) and is afterwards specialised to the setting of SGD with random initialisation (cf. Subsection 5.5.3).

5.1 Basics on DNNs

In this section we present the mathematical description of DNNs which we use throughout this chapter. It is a vectorised description in the sense that all the weights and biases associated to the DNN under consideration are collected in a single parameter vector $\theta \in \mathbb{R}^{\mathbf{d}}$ with $\mathbf{d} \in \mathbb{N} = \{1, 2, 3, \ldots\}$ sufficiently large (cf. Definitions 5.2 and 5.8). The content of this section is taken from Beck, Jentzen, & Kuckuck [27, Section 2.1] and is based on well-known material from the scientific literature; see, e.g., Beck et al. [22], Beck, E, & Jentzen [23], Berner, Grohs, & Jentzen [47], E, Han, & Jentzen [110], Goodfellow, Bengio, & Courville [145], and Grohs et al. [149]. In particular, Definition 5.1 is [27, Definition 2.1] (cf., e.g., (25) in [23]), Definition 5.2 is [27, Definition 2.2] (cf., e.g., (26) in [23]), Definition 5.3 is [27, Definition 2.3] (cf., e.g., [149, Definition 2.2]), and Definitions 5.4, 5.5, 5.6, 5.7, and 5.8 are [27, Definitions 2.4, 2.5, 2.6, 2.7, and 2.8] (cf., e.g., [47, Setting 2.5] and [145, Section 6.3]).

5.1.1 Vectorised description of DNNs

Definition 5.1 (Affine function). Let $\mathbf{d}, m, n \in \mathbb{N}$, $s \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}$, $\theta = (\theta_1, \theta_2, \ldots, \theta_{\mathbf{d}}) \in \mathbb{R}^{\mathbf{d}}$ satisfy $\mathbf{d} \geq s + mn + m$. Then we denote by $\mathcal{A}^{\theta,s}_{m,n} \colon \mathbb{R}^n \to \mathbb{R}^m$ the function which satisfies for all $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ that

$$\begin{aligned} \mathcal{A}^{\theta,s}_{m,n}(x) &= \begin{pmatrix} \theta_{s+1} & \theta_{s+2} & \cdots & \theta_{s+n} \\ \theta_{s+n+1} & \theta_{s+n+2} & \cdots & \theta_{s+2n} \\ \theta_{s+2n+1} & \theta_{s+2n+2} & \cdots & \theta_{s+3n} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{s+(m-1)n+1} & \theta_{s+(m-1)n+2} & \cdots & \theta_{s+mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \theta_{s+mn+1} \\ \theta_{s+mn+2} \\ \theta_{s+mn+3} \\ \vdots \\ \theta_{s+mn+m} \end{pmatrix} \\ &= \left(\left[\sum_{i=1}^{n} \theta_{s+i} x_i \right] + \theta_{s+mn+1}, \left[\sum_{i=1}^{n} \theta_{s+n+i} x_i \right] + \theta_{s+mn+2}, \ldots, \left[\sum_{i=1}^{n} \theta_{s+(m-1)n+i} x_i \right] + \theta_{s+mn+m} \right). \end{aligned} \tag{5.1}$$

Definition 5.2 (Fully connected feedforward artificial neural network). Let $\mathbf{d}, \mathbf{L}, \mathbf{l}_0, \mathbf{l}_1, \ldots, \mathbf{l}_{\mathbf{L}} \in \mathbb{N}$, $s \in \mathbb{N}_0$, $\theta \in \mathbb{R}^{\mathbf{d}}$ satisfy $\mathbf{d} \geq s + \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1} + 1)$ and let $\mathbf{a}_i \colon \mathbb{R}^{\mathbf{l}_i} \to \mathbb{R}^{\mathbf{l}_i}$, $i \in \{1, 2, \ldots, \mathbf{L}\}$, be functions. Then we denote by $\mathcal{N}^{\theta,s,\mathbf{l}_0}_{\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_{\mathbf{L}}} \colon \mathbb{R}^{\mathbf{l}_0} \to \mathbb{R}^{\mathbf{l}_{\mathbf{L}}}$ the function which satisfies for all $x \in \mathbb{R}^{\mathbf{l}_0}$ that

$$\begin{aligned} \big(\mathcal{N}^{\theta,s,\mathbf{l}_0}_{\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_{\mathbf{L}}}\big)(x) = \big(\mathbf{a}_{\mathbf{L}} \circ \mathcal{A}^{\theta, s+\sum_{i=1}^{\mathbf{L}-1} \mathbf{l}_i(\mathbf{l}_{i-1}+1)}_{\mathbf{l}_{\mathbf{L}},\mathbf{l}_{\mathbf{L}-1}} \circ \mathbf{a}_{\mathbf{L}-1} \circ \mathcal{A}^{\theta, s+\sum_{i=1}^{\mathbf{L}-2} \mathbf{l}_i(\mathbf{l}_{i-1}+1)}_{\mathbf{l}_{\mathbf{L}-1},\mathbf{l}_{\mathbf{L}-2}} \circ \ldots \\ \ldots \circ \mathbf{a}_2 \circ \mathcal{A}^{\theta, s+\mathbf{l}_1(\mathbf{l}_0+1)}_{\mathbf{l}_2,\mathbf{l}_1} \circ \mathbf{a}_1 \circ \mathcal{A}^{\theta,s}_{\mathbf{l}_1,\mathbf{l}_0}\big)(x) \end{aligned} \tag{5.2}$$

(cf. Definition 5.1).

5.1.2 Activation functions

Definition 5.3 (Multidimensional version). Let $d \in \mathbb{N}$ and let $\mathbf{a} \colon \mathbb{R} \to \mathbb{R}$ be a function. Then we denote by $\mathfrak{M}_{\mathbf{a},d} \colon \mathbb{R}^d \to \mathbb{R}^d$ the function which satisfies for all $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that

$$\mathfrak{M}_{\mathbf{a},d}(x) = (\mathbf{a}(x_1), \mathbf{a}(x_2), \ldots, \mathbf{a}(x_d)). \tag{5.3}$$

Definition 5.4 (Rectifier function)**.** We denote by $\mathfrak{r}\colon \mathbb{R} \to \mathbb{R}$ the function which satisfies for all $x \in \mathbb{R}$ that

$$\mathfrak{r}(x) = \max\{x, 0\}. \tag{5.4}$$

Definition 5.5 (Multidimensional rectifier function)**.** Let $d \in \mathbb{N}$. Then we denote by $\mathfrak{R}_d\colon \mathbb{R}^d \to \mathbb{R}^d$ the function given by

$$\mathfrak{R}_d = \mathfrak{M}_{\mathfrak{r},d} \tag{5.5}$$

(cf. Definitions 5.3 and 5.4).

Definition 5.6 (Clipping function)**.** Let $u \in [-\infty, \infty)$, $v \in (u, \infty]$. Then we denote by $\mathfrak{c}_{u,v}\colon \mathbb{R} \to \mathbb{R}$ the function which satisfies for all $x \in \mathbb{R}$ that

$$\mathfrak{c}_{u,v}(x) = \max\{u, \min\{x, v\}\}. \tag{5.6}$$

Definition 5.7 (Multidimensional clipping function)**.** Let $d \in \mathbb{N}$, $u \in [-\infty, \infty)$, $v \in (u, \infty]$. Then we denote by $\mathfrak{C}_{u,v,d}\colon \mathbb{R}^d \to \mathbb{R}^d$ the function given by

$$\mathfrak{C}_{u,v,d} = \mathfrak{M}_{\mathfrak{c}_{u,v},d} \tag{5.7}$$

(cf. Definitions 5.3 and 5.6).

5.1.3 Rectified DNNs

Definition 5.8 (Rectified clipped DNN)**.** Let $\mathbf{d}, \mathbf{L} \in \mathbb{N}$, $u \in [-\infty, \infty)$, $v \in (u, \infty]$, $\mathbf{l} = (\mathbf{l}_0, \mathbf{l}_1, \ldots, \mathbf{l}_\mathbf{L}) \in \mathbb{N}^{\mathbf{L}+1}$, $\theta \in \mathbb{R}^\mathbf{d}$ satisfy $\mathbf{d} \geq \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1} + 1)$. Then we denote by $\mathscr{N}_{u,v}^{\theta,\mathbf{l}}\colon \mathbb{R}^{\mathbf{l}_0} \to \mathbb{R}^{\mathbf{l}_\mathbf{L}}$ the function which satisfies for all $x \in \mathbb{R}^{\mathbf{l}_0}$ that

$$\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(x) = \begin{cases} \big(\mathcal{N}^{\theta,0,\mathbf{l}_0}_{\mathfrak{C}_{u,v,\mathbf{l}_\mathbf{L}}}\big)(x) & : \mathbf{L} = 1 \\ \big(\mathcal{N}^{\theta,0,\mathbf{l}_0}_{\mathfrak{R}_{\mathbf{l}_1},\mathfrak{R}_{\mathbf{l}_2},\ldots,\mathfrak{R}_{\mathbf{l}_{\mathbf{L}-1}},\mathfrak{C}_{u,v,\mathbf{l}_\mathbf{L}}}\big)(x) & : \mathbf{L} > 1 \end{cases} \tag{5.8}$$

(cf. Definitions 5.2, 5.5, and 5.7).

5.2 Analysis of the approximation error

This section is devoted to establishing a convergence result for the approximation of Lipschitz continuous functions by DNNs (cf. Proposition 5.13). More precisely, Proposition 5.13 establishes that a Lipschitz continuous function defined on a d-dimensional hypercube for $d \in \mathbb{N}$ can be approximated by DNNs with convergence rate $1/d$ with respect to a parameter $A \in (0, \infty)$ that bounds the architecture size (that is, depth and width) of the approximating DNN from below. Key ingredients of the proof of Proposition 5.13 are Beck, Jentzen, & Kuckuck [27, Corollary 3.8] as well as the elementary covering number estimate in Lemma 5.11. In order to improve the accessibility of Lemma 5.11, we recall the definition of covering numbers associated to a metric space in Definition 5.10, which is [27, Definition 3.11]. Lemma 5.11 provides upper bounds for the covering numbers of hypercubes equipped with the metric induced by the p-norm (cf. Definition 5.9) for $p \in [1, \infty]$ and is a generalisation of Berner, Grohs, & Jentzen [47, Lemma 2.7] (cf.

Cucker & Smale [91, Proposition 5] and [27, Proposition 3.12]). Furthermore, we present in Lemma 5.12 an elementary upper bound for the error arising when Lipschitz continuous functions defined on a hypercube are approximated by certain DNNs. Additional DNN approximation results can be found, e.g., in [11, 16, 17, 49, 50, 63, 64, 74, 76, 78, 92, 105, 116, 120–122, 128, 147–154, 159, 160, 163, 173–176, 180, 194, 218, 226, 243–245, 249, 252, 255, 256, 258–260, 263, 268, 275, 277, 280, 284, 285, 288, 298, 309, 310] and the references therein.

5.2.1 A covering number estimate

Definition 5.9 (p-norm)**.** We denote by $\|\cdot\|_p\colon \big(\bigcup_{d=1}^{\infty}\mathbb{R}^d\big)\to[0,\infty)$, $p\in[1,\infty]$, the functions which satisfy for all $p\in[1,\infty)$, $d\in\mathbb{N}$, $\theta=(\theta_1,\theta_2,\ldots,\theta_d)\in\mathbb{R}^d$ that

$$\|\theta\|_p=\left(\sum_{i=1}^{d}|\theta_i|^p\right)^{1/p}\qquad\text{and}\qquad\|\theta\|_\infty=\max_{i\in\{1,2,\ldots,d\}}|\theta_i|. \tag{5.9}$$

Definition 5.10 (Covering number)**.** Let (E,δ) be a metric space and let $r\in[0,\infty]$. Then we denote by $\mathcal{C}_{(E,\delta),r}\in\mathbb{N}_0\cup\{\infty\}$ (we denote by $\mathcal{C}_{E,r}\in\mathbb{N}_0\cup\{\infty\}$) the extended real number given by

$$\mathcal{C}_{(E,\delta),r}=\inf\!\left(\left\{n\in\mathbb{N}_0\colon\left[\exists\,A\subseteq E\colon\begin{pmatrix}(|A|\le n)\wedge(\forall\,x\in E\colon\\ \exists\,a\in A\colon\delta(a,x)\le r)\end{pmatrix}\right]\right\}\cup\{\infty\}\right). \tag{5.10}$$

Lemma 5.11. *Let $d\in\mathbb{N}$, $a\in\mathbb{R}$, $b\in(a,\infty)$, $r\in(0,\infty)$, for every $p\in[1,\infty]$ let $\delta_p\colon([a,b]^d)\times([a,b]^d)\to[0,\infty)$ satisfy for all $x,y\in[a,b]^d$ that $\delta_p(x,y)=\|x-y\|_p$, and let $\lceil\cdot\rceil\colon[0,\infty)\to\mathbb{N}_0$ satisfy for all $x\in[0,\infty)$ that $\lceil x\rceil=\min([x,\infty)\cap\mathbb{N}_0)$ (cf. Definition 5.9). Then*

(i) it holds for all $p\in[1,\infty)$ that

$$\mathcal{C}_{([a,b]^d,\delta_p),r}\le\left(\left\lceil\tfrac{d^{1/p}(b-a)}{2r}\right\rceil\right)^d\le\begin{cases}1 & \colon r\ge d(b-a)/2\\ \left(\tfrac{d(b-a)}{r}\right)^d & \colon r<d(b-a)/2\end{cases} \tag{5.11}$$

and

(ii) it holds that

$$\mathcal{C}_{([a,b]^d,\delta_\infty),r}\le\left(\left\lceil\tfrac{b-a}{2r}\right\rceil\right)^d\le\begin{cases}1 & \colon r\ge (b-a)/2\\ \left(\tfrac{b-a}{r}\right)^d & \colon r<(b-a)/2\end{cases} \tag{5.12}$$

(cf. Definition 5.10).

Proof of Lemma 5.11. Throughout this proof let $(\mathfrak{N}_p)_{p\in[1,\infty]}\subseteq\mathbb{N}$ satisfy for all $p\in[1,\infty)$ that

$$\mathfrak{N}_p=\left\lceil\tfrac{d^{1/p}(b-a)}{2r}\right\rceil\qquad\text{and}\qquad\mathfrak{N}_\infty=\left\lceil\tfrac{b-a}{2r}\right\rceil, \tag{5.13}$$

for every $N\in\mathbb{N}$, $i\in\{1,2,\ldots,N\}$ let $g_{N,i}\in[a,b]$ be given by $g_{N,i}=a+(i-1/2)(b-a)/N$, and for every $p\in[1,\infty]$ let $A_p\subseteq[a,b]^d$ be given by $A_p=\{g_{\mathfrak{N}_p,1},g_{\mathfrak{N}_p,2},\ldots,g_{\mathfrak{N}_p,\mathfrak{N}_p}\}^d$. Observe that it holds for all $N\in\mathbb{N}$, $i\in\{1,2,\ldots,N\}$, $x\in[a+(i-1)(b-a)/N,g_{N,i}]$ that

$$|x-g_{N,i}|=a+\tfrac{(i-1/2)(b-a)}{N}-x\le a+\tfrac{(i-1/2)(b-a)}{N}-\left(a+\tfrac{(i-1)(b-a)}{N}\right)=\tfrac{b-a}{2N}. \tag{5.14}$$

In addition, note that it holds for all $N \in \mathbb{N}$, $i \in \{1, 2, \ldots, N\}$, $x \in [g_{N,i}, a + {}^{i(b-a)}/_{N}]$ that

$$|x - g_{N,i}| = x - \left(a + \tfrac{(i-1/2)(b-a)}{N}\right) \leq a + \tfrac{i(b-a)}{N} - \left(a + \tfrac{(i-1/2)(b-a)}{N}\right) = \tfrac{b-a}{2N}. \tag{5.15}$$

Combining (5.14) and (5.15) implies for all $N \in \mathbb{N}$, $i \in \{1, 2, \ldots, N\}$, $x \in [a + {}^{(i-1)(b-a)}/_{N}, a + {}^{i(b-a)}/_{N}]$ that $|x - g_{N,i}| \leq {}^{(b-a)}/_{(2N)}$. This proves that for every $N \in \mathbb{N}$, $x \in [a, b]$ there exists $y \in \{g_{N,1}, g_{N,2}, \ldots, g_{N,N}\}$ such that

$$|x - y| \leq \tfrac{b-a}{2N}. \tag{5.16}$$

This, in turn, establishes that for every $p \in [1, \infty)$, $x = (x_1, x_2, \ldots, x_d) \in [a, b]^d$ there exists $y = (y_1, y_2, \ldots, y_d) \in A_p$ such that

$$\delta_p(x, y) = \|x-y\|_p = \left(\sum_{i=1}^{d} |x_i - y_i|^p\right)^{1/p} \leq \left(\sum_{i=1}^{d} \tfrac{(b-a)^p}{(2\mathfrak{N}_p)^p}\right)^{1/p} = \tfrac{d^{1/p}(b-a)}{2\mathfrak{N}_p} \leq \tfrac{d^{1/p}(b-a)2r}{2d^{1/p}(b-a)} = r. \tag{5.17}$$

Furthermore, again (5.16) shows that for every $x = (x_1, x_2, \ldots, x_d) \in [a, b]^d$ there exists $y = (y_1, y_2, \ldots, y_d) \in A_\infty$ such that

$$\delta_\infty(x, y) = \|x - y\|_\infty = \max_{i \in \{1,2,\ldots,d\}} |x_i - y_i| \leq \tfrac{b-a}{2\mathfrak{N}_\infty} \leq \tfrac{(b-a)2r}{2(b-a)} = r. \tag{5.18}$$

Note that (5.17), (5.13), and the fact that $\forall\, x \in [0, \infty) \colon \lceil x \rceil \leq \mathbb{1}_{(0,1]}(x) + 2x\mathbb{1}_{(1,\infty)}(x) = \mathbb{1}_{(0,r]}(rx) + 2x\mathbb{1}_{(r,\infty)}(rx)$ yield for all $p \in [1, \infty)$ that

$$\begin{aligned} \mathcal{C}_{([a,b]^d, \delta_p), r} &\leq |A_p| = (\mathfrak{N}_p)^d = \left(\left\lceil \tfrac{d^{1/p}(b-a)}{2r} \right\rceil\right)^d \leq \left(\left\lceil \tfrac{d(b-a)}{2r} \right\rceil\right)^d \\ &\leq \left(\mathbb{1}_{(0,r]}\left(\tfrac{d(b-a)}{2}\right) + \tfrac{2d(b-a)}{2r} \mathbb{1}_{(r,\infty)}\left(\tfrac{d(b-a)}{2}\right)\right)^d \\ &= \mathbb{1}_{(0,r]}\left(\tfrac{d(b-a)}{2}\right) + \left(\tfrac{d(b-a)}{r}\right)^d \mathbb{1}_{(r,\infty)}\left(\tfrac{d(b-a)}{2}\right). \end{aligned} \tag{5.19}$$

This proves (i). In addition, (5.18), (5.13), and again the fact that $\forall\, x \in [0, \infty) \colon \lceil x \rceil \leq \mathbb{1}_{(0,r]}(rx) + 2x\mathbb{1}_{(r,\infty)}(rx)$ demonstrate that

$$\mathcal{C}_{([a,b]^d, \delta_\infty), r} \leq |A_\infty| = (\mathfrak{N}_\infty)^d = \left(\left\lceil \tfrac{b-a}{2r} \right\rceil\right)^d \leq \mathbb{1}_{(0,r]}\left(\tfrac{b-a}{2}\right) + \left(\tfrac{b-a}{r}\right)^d \mathbb{1}_{(r,\infty)}\left(\tfrac{b-a}{2}\right). \tag{5.20}$$

This implies (ii) and thus completes the proof of Lemma 5.11. □

5.2.2 Convergence rates for the approximation error

Lemma 5.12. *Let $d, \mathbf{d}, \mathbf{L} \in \mathbb{N}$, $L, a \in \mathbb{R}$, $b \in (a, \infty)$, $u \in [-\infty, \infty)$, $v \in (u, \infty]$, $\mathbf{l} = (\mathbf{l}_0, \mathbf{l}_1, \ldots, \mathbf{l}_\mathbf{L}) \in \mathbb{N}^{\mathbf{L}+1}$, assume $\mathbf{l}_0 = d$, $\mathbf{l}_\mathbf{L} = 1$, and $\mathbf{d} \geq \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1} + 1)$, and let $f \colon [a, b]^d \to ([u, v] \cap \mathbb{R})$ satisfy for all $x, y \in [a, b]^d$ that $|f(x) - f(y)| \leq L\|x - y\|_1$ (cf. Definition 5.9). Then there exists $\vartheta \in \mathbb{R}^\mathbf{d}$ such that $\|\vartheta\|_\infty \leq \sup_{x \in [a,b]^d} |f(x)|$ and*

$$\sup\nolimits_{x \in [a,b]^d} |\mathscr{N}_{u,v}^{\vartheta, \mathbf{l}}(x) - f(x)| \leq \frac{dL(b-a)}{2} \tag{5.21}$$

(cf. Definition 5.8).

Proof of Lemma 5.12. Throughout this proof let $\mathfrak{d} \in \mathbb{N}$ be given by $\mathfrak{d} = \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1}+1)$, let $\mathbf{m} = (\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_d) \in [a,b]^d$ satisfy for all $i \in \{1,2,\ldots,d\}$ that $\mathbf{m}_i = {(a+b)}/{2}$, and let $\vartheta = (\vartheta_1, \vartheta_2, \ldots, \vartheta_{\mathbf{d}}) \in \mathbb{R}^{\mathbf{d}}$ satisfy for all $i \in \{1,2,\ldots,\mathbf{d}\} \setminus \{\mathfrak{d}\}$ that $\vartheta_i = 0$ and $\vartheta_{\mathfrak{d}} = f(\mathbf{m})$. Observe that the assumption that $\mathbf{l}_{\mathbf{L}} = 1$ and the fact that $\forall\, i \in \{1,2,\ldots,\mathfrak{d}-1\} \colon \vartheta_i = 0$ show for all $x = (x_1, x_2, \ldots, x_{\mathbf{l}_{\mathbf{L}-1}}) \in \mathbb{R}^{\mathbf{l}_{\mathbf{L}-1}}$ that

$$\begin{aligned}\mathcal{A}_{1,\mathbf{l}_{\mathbf{L}-1}}^{\vartheta,\sum_{i=1}^{\mathbf{L}-1}\mathbf{l}_i(\mathbf{l}_{i-1}+1)}(x) &= \left[\sum_{i=1}^{\mathbf{l}_{\mathbf{L}-1}} \vartheta_{[\sum_{i=1}^{\mathbf{L}-1}\mathbf{l}_i(\mathbf{l}_{i-1}+1)]+i}\, x_i\right] + \vartheta_{[\sum_{i=1}^{\mathbf{L}-1}\mathbf{l}_i(\mathbf{l}_{i-1}+1)]+\mathbf{l}_{\mathbf{L}-1}+1} \\ &= \left[\sum_{i=1}^{\mathbf{l}_{\mathbf{L}-1}} \vartheta_{[\sum_{i=1}^{\mathbf{L}}\mathbf{l}_i(\mathbf{l}_{i-1}+1)]-(\mathbf{l}_{\mathbf{L}-1}-i+1)}\, x_i\right] + \vartheta_{\sum_{i=1}^{\mathbf{L}}\mathbf{l}_i(\mathbf{l}_{i-1}+1)} \\ &= \left[\sum_{i=1}^{\mathbf{l}_{\mathbf{L}-1}} \vartheta_{\mathfrak{d}-(\mathbf{l}_{\mathbf{L}-1}-i+1)}\, x_i\right] + \vartheta_{\mathfrak{d}} = \vartheta_{\mathfrak{d}} = f(\mathbf{m})\end{aligned} \tag{5.22}$$

(cf. Definition 5.1). Combining this with the fact that $f(\mathbf{m}) \in [u,v]$ ensures for all $x \in \mathbb{R}^{\mathbf{l}_{\mathbf{L}-1}}$ that

$$\begin{aligned}&\big(\mathfrak{C}_{u,v,\mathbf{l}_{\mathbf{L}}} \circ \mathcal{A}_{\mathbf{l}_{\mathbf{L}},\mathbf{l}_{\mathbf{L}-1}}^{\vartheta,\sum_{i=1}^{\mathbf{L}-1}\mathbf{l}_i(\mathbf{l}_{i-1}+1)}\big)(x) = \big(\mathfrak{C}_{u,v,1} \circ \mathcal{A}_{1,\mathbf{l}_{\mathbf{L}-1}}^{\vartheta,\sum_{i=1}^{\mathbf{L}-1}\mathbf{l}_i(\mathbf{l}_{i-1}+1)}\big)(x) = \mathfrak{c}_{u,v}(f(\mathbf{m})) \\ &= \max\{u, \min\{f(\mathbf{m}), v\}\} = \max\{u, f(\mathbf{m})\} = f(\mathbf{m})\end{aligned} \tag{5.23}$$

(cf. Definitions 5.6 and 5.7). This proves for all $x \in \mathbb{R}^d$ that

$$\mathscr{N}_{u,v}^{\vartheta,\mathbf{l}}(x) = f(\mathbf{m}). \tag{5.24}$$

In addition, note that it holds for all $x \in [a, \mathbf{m}_1]$, $\mathfrak{x} \in [\mathbf{m}_1, b]$ that $|\mathbf{m}_1 - x| = \mathbf{m}_1 - x = {(a+b)}/{2} - x \leq {(a+b)}/{2} - a = {(b-a)}/{2}$ and $|\mathbf{m}_1 - \mathfrak{x}| = \mathfrak{x} - \mathbf{m}_1 = \mathfrak{x} - {(a+b)}/{2} \leq b - {(a+b)}/{2} = {(b-a)}/{2}$. The assumption that $\forall\, x,y \in [a,b]^d \colon |f(x) - f(y)| \leq L\|x-y\|_1$ and (5.24) hence demonstrate for all $x = (x_1, x_2, \ldots, x_d) \in [a,b]^d$ that

$$\begin{aligned}|\mathscr{N}_{u,v}^{\vartheta,\mathbf{l}}(x) - f(x)| &= |f(\mathbf{m}) - f(x)| \leq L\|\mathbf{m} - x\|_1 = L\sum_{i=1}^{d}|\mathbf{m}_i - x_i| \\ &= L\sum_{i=1}^{d}|\mathbf{m}_1 - x_i| \leq \sum_{i=1}^{d}\frac{L(b-a)}{2} = \frac{dL(b-a)}{2}.\end{aligned} \tag{5.25}$$

This and the fact that $\|\vartheta\|_\infty = \max_{i\in\{1,2,\ldots,\mathbf{d}\}}|\vartheta_i| = |f(\mathbf{m})| \leq \sup_{x\in[a,b]^d}|f(x)|$ complete the proof of Lemma 5.12. □

Proposition 5.13. *Let $d, \mathbf{d}, \mathbf{L} \in \mathbb{N}$, $A \in (0,\infty)$, $L, a \in \mathbb{R}$, $b \in (a,\infty)$, $u \in [-\infty,\infty)$, $v \in (u,\infty]$, $\mathbf{l} = (\mathbf{l}_0, \mathbf{l}_1, \ldots, \mathbf{l}_{\mathbf{L}}) \in \mathbb{N}^{\mathbf{L}+1}$, assume $\mathbf{L} \geq {A\mathbb{1}_{(6^d,\infty)}(A)}/{(2d)} + 1$, $\mathbf{l}_0 = d$, $\mathbf{l}_1 \geq A\mathbb{1}_{(6^d,\infty)}(A)$, $\mathbf{l}_{\mathbf{L}} = 1$, and $\mathbf{d} \geq \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1}+1)$, assume for all $i \in \{2,3,\ldots\} \cap [0,\mathbf{L})$ that $\mathbf{l}_i \geq \mathbb{1}_{(6^d,\infty)}(A)\max\{{A}/{d} - 2i + 3, 2\}$, and let $f \colon [a,b]^d \to ([u,v] \cap \mathbb{R})$ satisfy for all $x,y \in [a,b]^d$ that $|f(x)-f(y)| \leq L\|x-y\|_1$ (cf. Definition 5.9). Then there exists $\vartheta \in \mathbb{R}^{\mathbf{d}}$ such that $\|\vartheta\|_\infty \leq \max\{1, L, |a|, |b|, 2[\sup_{x\in[a,b]^d}|f(x)|]\}$ and*

$$\sup_{x\in[a,b]^d}|\mathscr{N}_{u,v}^{\vartheta,\mathbf{l}}(x) - f(x)| \leq \frac{3dL(b-a)}{A^{1/d}} \tag{5.26}$$

(cf. Definition 5.8).

Proof of Proposition 5.13. Throughout this proof assume w.l.o.g. that $A > 6^d$ (cf. Lemma 5.12), let $\mathfrak{N} \in \mathbb{N}$ be given by

$$\mathfrak{N} = \max\Big\{\mathfrak{n} \in \mathbb{N} \colon \mathfrak{n} \le \big(\tfrac{A}{2d}\big)^{1/d}\Big\}, \tag{5.27}$$

let $r \in (0,\infty)$ be given by $r = d(b-a)/(2\mathfrak{N})$, let $\delta \colon ([a,b]^d) \times ([a,b]^d) \to [0,\infty)$ satisfy for all $x, y \in [a,b]^d$ that $\delta(x,y) = \|x-y\|_1$, let $\mathscr{D} \subseteq [a,b]^d$ satisfy $|\mathscr{D}| = \max\{2, \mathcal{C}_{([a,b]^d,\delta),r}\}$ and

$$\sup\nolimits_{x\in[a,b]^d} \inf\nolimits_{y\in\mathscr{D}} \delta(x,y) \le r \tag{5.28}$$

(cf. Definition 5.10), and let $\lceil\cdot\rceil \colon [0,\infty) \to \mathbb{N}_0$ satisfy for all $x \in [0,\infty)$ that $\lceil x \rceil = \min([x,\infty) \cap \mathbb{N}_0)$. Note that it holds for all $\mathfrak{d} \in \mathbb{N}$ that

$$2\mathfrak{d} \le 2 \cdot 2^{\mathfrak{d}-1} = 2^{\mathfrak{d}}. \tag{5.29}$$

This implies that $3^d = 6^d/2^d \le A/(2d)$. Equation (5.27) hence demonstrates that

$$2 \le \tfrac{2}{3}\big(\tfrac{A}{2d}\big)^{1/d} = \big(\tfrac{A}{2d}\big)^{1/d} - \tfrac{1}{3}\big(\tfrac{A}{2d}\big)^{1/d} \le \big(\tfrac{A}{2d}\big)^{1/d} - 1 < \mathfrak{N}. \tag{5.30}$$

This and (i) in Lemma 5.11 (with $\delta_1 \leftarrow \delta$, $p \leftarrow 1$ in the notation of (i) in Lemma 5.11) establish that

$$|\mathscr{D}| = \max\{2, \mathcal{C}_{([a,b]^d,\delta),r}\} \le \max\Big\{2, \Big(\Big\lceil \tfrac{d(b-a)}{2r} \Big\rceil\Big)^d\Big\} = \max\{2, (\lceil \mathfrak{N} \rceil)^d\} = \mathfrak{N}^d. \tag{5.31}$$

Combining this with (5.27) proves that

$$4 \le 2d|\mathscr{D}| \le 2d\mathfrak{N}^d \le \tfrac{2dA}{2d} = A. \tag{5.32}$$

The fact that $\mathbf{L} \ge A\mathbb{1}_{(6^d,\infty)}(A)/(2d) + 1 = A/(2d) + 1$ hence yields that $|\mathscr{D}| \le A/(2d) \le \mathbf{L} - 1$. This, (5.32), and the facts that $\mathbf{l}_1 \ge A\mathbb{1}_{(6^d,\infty)}(A) = A$ and $\forall\, i \in \{2,3,\ldots\} \cap [0,\mathbf{L}) = \{2,3,\ldots,\mathbf{L}-1\} \colon \mathbf{l}_i \ge \mathbb{1}_{(6^d,\infty)}(A) \max\{A/d - 2i + 3, 2\} = \max\{A/d - 2i + 3, 2\}$ imply for all $i \in \{2,3,\ldots,|\mathscr{D}|\}$ that

$$\mathbf{L} \ge |\mathscr{D}| + 1, \quad \mathbf{l}_1 \ge A \ge 2d|\mathscr{D}|, \quad \text{and} \quad \mathbf{l}_i \ge A/d - 2i + 3 \ge 2|\mathscr{D}| - 2i + 3. \tag{5.33}$$

In addition, the fact that $\forall\, i \in \{2,3,\ldots\} \cap [0,\mathbf{L}) \colon \mathbf{l}_i \ge \max\{A/d - 2i + 3, 2\}$ ensures for all $i \in \mathbb{N} \cap (|\mathscr{D}|, \mathbf{L})$ that

$$\mathbf{l}_i \ge 2. \tag{5.34}$$

Furthermore, observe that it holds for all $x = (x_1, x_2, \ldots, x_d), y = (y_1, y_2, \ldots, y_d) \in [a,b]^d$ that

$$|f(x) - f(y)| \le L\|x-y\|_1 = L\Bigg[\sum_{i=1}^d |x_i - y_i|\Bigg]. \tag{5.35}$$

This, the assumptions that $\mathbf{l}_0 = d$, $\mathbf{l}_{\mathbf{L}} = 1$, and $\mathbf{d} \ge \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1}+1)$, (5.33)–(5.34), and Beck, Jentzen, & Kuckuck [27, Corollary 3.8] (with $d \leftarrow d$, $\mathfrak{d} \leftarrow \mathbf{d}$, $\mathfrak{L} \leftarrow \mathbf{L}$, $L \leftarrow L$, $u \leftarrow u$, $v \leftarrow v$, $D \leftarrow [a,b]^d$, $f \leftarrow f$, $\mathcal{M} \leftarrow \mathscr{D}$, $l \leftarrow \mathbf{l}$ in the notation of [27, Corollary 3.8]) show that there exists $\vartheta \in \mathbb{R}^{\mathbf{d}}$ such that $\|\vartheta\|_\infty \le \max\{1, L, \sup_{x\in\mathscr{D}}\|x\|_\infty, 2[\sup_{x\in\mathscr{D}}|f(x)|]\}$ and

$$\begin{aligned}\sup_{x\in[a,b]^d} |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - f(x)| &\le 2L\Bigg[\sup_{x=(x_1,x_2,\ldots,x_d)\in[a,b]^d} \Bigg(\inf_{y=(y_1,y_2,\ldots,y_d)\in\mathscr{D}} \sum_{i=1}^d |x_i - y_i|\Bigg)\Bigg] \\ &= 2L\Bigg[\sup_{x\in[a,b]^d} \inf_{y\in\mathscr{D}} \|x-y\|_1\Bigg] = 2L\Bigg[\sup_{x\in[a,b]^d} \inf_{y\in\mathscr{D}} \delta(x,y)\Bigg].\end{aligned} \tag{5.36}$$

Note that this demonstrates that

$$\|\vartheta\|_\infty \le \max\{1, L, |a|, |b|, 2[\sup_{x\in[a,b]^d}|f(x)|]\}. \tag{5.37}$$

Moreover, (5.36) and (5.28)–(5.30) prove that

$$\begin{aligned}&\sup_{x\in[a,b]^d}|\mathscr{N}^{\vartheta,1}_{u,v}(x) - f(x)| \le 2L\big[\sup_{x\in[a,b]^d}\inf_{y\in\mathscr{D}}\delta(x,y)\big] \le 2Lr\\ &= \frac{dL(b-a)}{\mathfrak{N}} \le \frac{dL(b-a)}{\frac{2}{3}\big(\frac{A}{2d}\big)^{1/d}} = \frac{(2d)^{1/d}3dL(b-a)}{2A^{1/d}} \le \frac{3dL(b-a)}{A^{1/d}}.\end{aligned} \tag{5.38}$$

Combining this with (5.37) completes the proof of Proposition 5.13. □

5.3 Analysis of the generalisation error

In this section we consider the *worst-case* generalisation error arising in deep learning based empirical risk minimisation with quadratic loss function for DNNs with a fixed architecture and weights and biases bounded in size by a fixed constant (cf. Corollary 5.28 in Subsection 5.3.3). We prove that this worst-case generalisation error converges in the probabilistically strong sense with rate $1/2$ (up to a logarithmic factor) with respect to the number of samples used for calculating the empirical risk and that the constant in the corresponding upper bound for the worst-case generalisation error scales favourably (i.e., only very moderately) in terms of depth and width of the DNNs employed; cf. (ii) in Corollary 5.28. Corollary 5.28 is a consequence of the main result of this section, Proposition 5.27 in Subsection 5.3.3, which provides a similar conclusion in a more general setting. The proofs of Proposition 5.27 and Corollary 5.28, respectively, rely on the tools developed in the two preceding subsections, Subsections 5.3.1 and 5.3.2.

On the one hand, Subsection 5.3.1 provides an essentially well-known estimate for the L^p-error of Monte Carlo-type approximations; cf. Corollary 5.18. Corollary 5.18 is a consequence of the well-known result stated here as Proposition 5.17, which, in turn, follows directly from, e.g., Cox et al. [85, Corollary 5.11] (with $M \leftarrow M$, $q \leftarrow 2$, $(E, \|\cdot\|_E) \leftarrow (\mathbb{R}^d, \|\cdot\|_2|_{\mathbb{R}^d})$, $(\Omega, \mathscr{F}, \mathbb{P}) \leftarrow (\Omega, \mathcal{F}, \mathbb{P})$, $(\xi_j)_{j\in\{1,2,\dots,M\}} \leftarrow (X_j)_{j\in\{1,2,\dots,M\}}$, $p \leftarrow p$ in the notation of [85, Corollary 5.11] and Proposition 5.17, respectively). In the proof of Corollary 5.18 we also apply Lemma 5.16, which is Grohs et al. [148, Lemma 2.2]. In order to make the statements of Lemma 5.16 and Proposition 5.17 more accessible for the reader, we recall in Definition 5.14 (cf., e.g., [85, Definition 5.1]) the notion of a Rademacher family and in Definition 5.15 (cf., e.g., [85, Definition 5.4] or Gonon et al. [144, Definition 2.1]) the notion of the p-Kahane–Khintchine constant.

On the other hand, we derive in Subsection 5.3.2 uniform L^p-estimates for Lipschitz continuous random fields with a separable metric space as index set (cf. Lemmas 5.23 and 5.24 and Corollary 5.25). These estimates are uniform in the sense that the supremum over the index set is *inside* the expectation belonging to the L^p-norm, which is necessary since we intend to prove error bounds for the *worst-case* generalisation error, as illustrated above. One of the elementary but crucial arguments in our derivation of such uniform L^p-estimates is given in Lemma 5.22 (cf. Lemma 5.21). Roughly speaking, Lemma 5.22 illustrates how the L^p-norm of a supremum can be bounded from above by the supremum of certain L^p-norms, where the L^p-norms are integrating over a general measure space and where the suprema are taken over a general (bounded) separable metric space. Furthermore, the elementary and well-known Lemmas 5.19 and 5.20, respectively, follow immediately from Beck, Jentzen, & Kuckuck [27, (ii) in Lemma 3.13 and (ii) in Lemma 3.14] and ensure that the mathematical statements of Lemmas 5.21, 5.22, and 5.23 do indeed make sense.

The results in Subsections 5.3.2 and 5.3.3 are in parts inspired by [27, Subsection 3.2] and we refer, e.g., to [18, 47, 91, 113–115, 156, 240, 281, 297] and the references therein for further results on the generalisation error.

5.3.1 Monte Carlo estimates

Definition 5.14 (Rademacher family)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let J be a set. Then we say that $(r_j)_{j \in J}$ is a $\mathbb{P}$-Rademacher family if and only if it holds that $r_j \colon \Omega \to \{-1, 1\}$, $j \in J$, are independent random variables with $\forall\, j \in J \colon \mathbb{P}(r_j = 1) = \mathbb{P}(r_j = -1)$.

Definition 5.15 (p-Kahane–Khintchine constant)**.** Let $p \in (0, \infty)$. Then we denote by $\mathfrak{K}_p \in (0, \infty]$ the extended real number given by

$$\mathfrak{K}_p = \sup \left\{ c \in [0,\infty) \colon \left[\begin{array}{c} \exists\, \mathbb{R}\text{-Banach space } (E, \|\cdot\|_E) \colon \\ \exists\, \text{probability space } (\Omega, \mathcal{F}, \mathbb{P}) \colon \\ \exists\, \mathbb{P}\text{-Rademacher family } (r_j)_{j \in \mathbb{N}} \colon \\ \exists\, k \in \mathbb{N} \colon \exists\, x_1, x_2, \dots, x_k \in E \setminus \{0\} \colon \\ \Big(\mathbb{E}\Big[\big\|\sum_{j=1}^k r_j x_j\big\|_E^p\Big]\Big)^{1/p} = c \Big(\mathbb{E}\Big[\big\|\sum_{j=1}^k r_j x_j\big\|_E^2\Big]\Big)^{1/2} \end{array} \right] \right\} \tag{5.39}$$

(cf. Definition 5.14).

Lemma 5.16. *It holds for all $p \in [2, \infty)$ that $\mathfrak{K}_p \leq \sqrt{p-1} < \infty$ (cf. Definition 5.15).*

Proposition 5.17. *Let $d, M \in \mathbb{N}$, $p \in [2, \infty)$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X_j \colon \Omega \to \mathbb{R}^d$, $j \in \{1, 2, \dots, M\}$, be independent random variables, and assume $\max_{j \in \{1,2,\dots,M\}} \mathbb{E}[\|X_j\|_2] < \infty$ (cf. Definition 5.9). Then*

$$\left(\mathbb{E}\left[\left\|\left[\sum_{j=1}^M X_j\right] - \mathbb{E}\left[\sum_{j=1}^M X_j\right]\right\|_2^p\right]\right)^{1/p} \leq 2\mathfrak{K}_p \left[\sum_{j=1}^M \left(\mathbb{E}\big[\|X_j - \mathbb{E}[X_j]\|_2^p\big]\right)^{2/p}\right]^{1/2} \tag{5.40}$$

(cf. Definition 5.15 and Lemma 5.16).

Corollary 5.18. *Let $d, M \in \mathbb{N}$, $p \in [2, \infty)$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X_j \colon \Omega \to \mathbb{R}^d$, $j \in \{1, 2, \dots, M\}$, be independent random variables, and assume $\max_{j \in \{1,2,\dots,M\}} \mathbb{E}[\|X_j\|_2] < \infty$ (cf. Definition 5.9). Then*

$$\left(\mathbb{E}\left[\left\|\frac{1}{M}\left[\sum_{j=1}^M X_j\right] - \mathbb{E}\left[\frac{1}{M}\sum_{j=1}^M X_j\right]\right\|_2^p\right]\right)^{1/p} \leq \frac{2\sqrt{p-1}}{\sqrt{M}} \left[\max_{j \in \{1,2,\dots,M\}} \left(\mathbb{E}\big[\|X_j - \mathbb{E}[X_j]\|_2^p\big]\right)^{1/p}\right]. \tag{5.41}$$

Proof of Corollary 5.18. Observe that Proposition 5.17 and Lemma 5.16 imply that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left\|\frac{1}{M}\left[\sum_{j=1}^{M} X_j\right] - \mathbb{E}\left[\frac{1}{M}\sum_{j=1}^{M} X_j\right]\right\|_2^p\right]\right)^{1/p} \\
&= \frac{1}{M}\left(\mathbb{E}\left[\left\|\left[\sum_{j=1}^{M} X_j\right] - \mathbb{E}\left[\sum_{j=1}^{M} X_j\right]\right\|_2^p\right]\right)^{1/p} \\
&\leq \frac{2\mathfrak{K}_p}{M}\left[\sum_{j=1}^{M}\left(\mathbb{E}\left[\|X_j - \mathbb{E}[X_j]\|_2^p\right]\right)^{2/p}\right]^{1/2} \\
&\leq \frac{2\mathfrak{K}_p}{M}\left[M\left(\max_{j\in\{1,2,\ldots,M\}}\left(\mathbb{E}\left[\|X_j - \mathbb{E}[X_j]\|_2^p\right]\right)^{2/p}\right)\right]^{1/2} \\
&= \frac{2\mathfrak{K}_p}{\sqrt{M}}\left[\max_{j\in\{1,2,\ldots,M\}}\left(\mathbb{E}\left[\|X_j - \mathbb{E}[X_j]\|_2^p\right]\right)^{1/p}\right] \\
&\leq \frac{2\sqrt{p-1}}{\sqrt{M}}\left[\max_{j\in\{1,2,\ldots,M\}}\left(\mathbb{E}\left[\|X_j - \mathbb{E}[X_j]\|_2^p\right]\right)^{1/p}\right]
\end{aligned} \tag{5.42}$$

(cf. Definition 5.15). The proof of Corollary 5.18 is thus complete. □

5.3.2 Uniform strong error estimates for random fields

Lemma 5.19. *Let $(E, \mathscr{E})$ be a separable topological space, assume $E \neq \emptyset$, let $(\Omega, \mathcal{F})$ be a measurable space, let $f_x\colon \Omega \to \mathbb{R}$, $x \in E$, be $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable functions, and assume for all $\omega \in \Omega$ that $E \ni x \mapsto f_x(\omega) \in \mathbb{R}$ is a continuous function. Then it holds that the function*

$$\Omega \ni \omega \mapsto \sup\nolimits_{x\in E} f_x(\omega) \in \mathbb{R} \cup \{\infty\} \tag{5.43}$$

is $\mathcal{F}/\mathcal{B}(\mathbb{R}\cup\{\infty\})$-measurable.

Lemma 5.20. *Let (E, δ) be a separable metric space, assume $E \neq \emptyset$, let $L \in \mathbb{R}$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $Z_x\colon \Omega \to \mathbb{R}$, $x \in E$, be random variables, and assume for all $x, y \in E$ that $\mathbb{E}[|Z_x|] < \infty$ and $|Z_x - Z_y| \leq L\delta(x, y)$. Then it holds that the function*

$$\Omega \ni \omega \mapsto \sup\nolimits_{x\in E}|Z_x(\omega) - \mathbb{E}[Z_x]| \in [0, \infty] \tag{5.44}$$

is $\mathcal{F}/\mathcal{B}([0, \infty])$-measurable.

Lemma 5.21. *Let (E, δ) be a separable metric space, let $N \in \mathbb{N}$, $p, L, r_1, r_2, \ldots, r_N \in [0, \infty)$, $z_1, z_2, \ldots, z_N \in E$ satisfy $E \subseteq \bigcup_{i=1}^{N}\{x \in E\colon \delta(x, z_i) \leq r_i\}$, let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $Z_x\colon \Omega \to \mathbb{R}$, $x \in E$, be $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable functions, and assume for all $\omega \in \Omega$, $x, y \in E$ that $|Z_x(\omega) - Z_y(\omega)| \leq L\delta(x, y)$. Then*

$$\int_\Omega \sup_{x\in E} |Z_x(\omega)|^p \,\mu(\mathrm{d}\omega) \leq \sum_{i=1}^{N} \int_\Omega (Lr_i + |Z_{z_i}(\omega)|)^p \,\mu(\mathrm{d}\omega) \tag{5.45}$$

(cf. Lemma 5.19).

Proof of Lemma 5.21. Throughout this proof let $B_1, B_2, \ldots, B_N \subseteq E$ satisfy for all $i \in \{1, 2, \ldots, N\}$ that $B_i = \{x \in E\colon \delta(x, z_i) \leq r_i\}$. Note that the fact that $E = \bigcup_{i=1}^{N} B_i$ shows for all $\omega \in \Omega$ that

$$\sup\nolimits_{x\in E}|Z_x(\omega)| = \sup\nolimits_{x\in(\bigcup_{i=1}^{N} B_i)}|Z_x(\omega)| = \max\nolimits_{i\in\{1,2,\ldots,N\}} \sup\nolimits_{x\in B_i}|Z_x(\omega)|. \tag{5.46}$$

This establishes that

$$\begin{aligned}\int_\Omega \sup_{x\in E}|Z_x(\omega)|^p\,\mu(\mathrm{d}\omega) &= \int_\Omega \max_{i\in\{1,2,\ldots,N\}}\sup_{x\in B_i}|Z_x(\omega)|^p\,\mu(\mathrm{d}\omega)\\ &\le \int_\Omega \sum_{i=1}^N \sup_{x\in B_i}|Z_x(\omega)|^p\,\mu(\mathrm{d}\omega) = \sum_{i=1}^N \int_\Omega \sup_{x\in B_i}|Z_x(\omega)|^p\,\mu(\mathrm{d}\omega).\end{aligned} \tag{5.47}$$

Furthermore, the assumption that $\forall\,\omega\in\Omega,\ x,y\in E\colon |Z_x(\omega)-Z_y(\omega)|\le L\delta(x,y)$ implies for all $\omega\in\Omega$, $i\in\{1,2,\ldots,N\}$, $x\in B_i$ that

$$\begin{aligned}|Z_x(\omega)| = |Z_x(\omega)-Z_{z_i}(\omega)+Z_{z_i}(\omega)| &\le |Z_x(\omega)-Z_{z_i}(\omega)|+|Z_{z_i}(\omega)|\\ &\le L\delta(x,z_i)+|Z_{z_i}(\omega)| \le Lr_i+|Z_{z_i}(\omega)|.\end{aligned} \tag{5.48}$$

Combining this with (5.47) proves that

$$\int_\Omega \sup_{x\in E}|Z_x(\omega)|^p\,\mu(\mathrm{d}\omega) \le \sum_{i=1}^N \int_\Omega (Lr_i+|Z_{z_i}(\omega)|)^p\,\mu(\mathrm{d}\omega). \tag{5.49}$$

The proof of Lemma 5.21 is thus complete. □

Lemma 5.22. *Let $p,L,r\in(0,\infty)$, let (E,δ) be a separable metric space, let $(\Omega,\mathcal{F},\mu)$ be a measure space, assume $E\neq\emptyset$ and $\mu(\Omega)\neq 0$, let $Z_x\colon\Omega\to\mathbb{R}$, $x\in E$, be $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable functions, and assume for all $\omega\in\Omega$, $x,y\in E$ that $|Z_x(\omega)-Z_y(\omega)|\le L\delta(x,y)$. Then*

$$\int_\Omega \sup_{x\in E}|Z_x(\omega)|^p\,\mu(\mathrm{d}\omega) \le \mathcal{C}_{(E,\delta),r}\left[\sup_{x\in E}\int_\Omega (Lr+|Z_x(\omega)|)^p\,\mu(\mathrm{d}\omega)\right] \tag{5.50}$$

(cf. Definition 5.10 and Lemma 5.19).

Proof of Lemma 5.22. Throughout this proof assume w.l.o.g. that $\mathcal{C}_{(E,\delta),r}<\infty$, let $N\in\mathbb{N}$ be given by $N=\mathcal{C}_{(E,\delta),r}$, and let $z_1,z_2,\ldots,z_N\in E$ satisfy $E\subseteq\bigcup_{i=1}^N\{x\in E\colon \delta(x,z_i)\le r\}$. Note that Lemma 5.21 (with $r_1\leftarrow r$, $r_2\leftarrow r$, …, $r_N\leftarrow r$ in the notation of Lemma 5.21) establishes that

$$\begin{aligned}\int_\Omega \sup_{x\in E}|Z_x(\omega)|^p\,\mu(\mathrm{d}\omega) &\le \sum_{i=1}^N \int_\Omega (Lr+|Z_{z_i}(\omega)|)^p\,\mu(\mathrm{d}\omega)\\ &\le \sum_{i=1}^N\left[\sup_{x\in E}\int_\Omega (Lr+|Z_x(\omega)|)^p\,\mu(\mathrm{d}\omega)\right] = N\left[\sup_{x\in E}\int_\Omega (Lr+|Z_x(\omega)|)^p\,\mu(\mathrm{d}\omega)\right].\end{aligned} \tag{5.51}$$

The proof of Lemma 5.22 is thus complete. □

Lemma 5.23. *Let $p\in[1,\infty)$, $L,r\in(0,\infty)$, let (E,δ) be a separable metric space, assume $E\neq\emptyset$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $Z_x\colon\Omega\to\mathbb{R}$, $x\in E$, be random variables, and assume for all $x,y\in E$ that $\mathbb{E}[|Z_x|]<\infty$ and $|Z_x-Z_y|\le L\delta(x,y)$. Then*

$$\left(\mathbb{E}\big[\sup\nolimits_{x\in E}|Z_x-\mathbb{E}[Z_x]|^p\big]\right)^{1/p} \le (\mathcal{C}_{(E,\delta),r})^{1/p}\left[2Lr+\sup\nolimits_{x\in E}\left(\mathbb{E}\big[|Z_x-\mathbb{E}[Z_x]|^p\big]\right)^{1/p}\right] \tag{5.52}$$

(cf. Definition 5.10 and Lemma 5.20).

Proof of Lemma 5.23. Throughout this proof let $Y_x\colon \Omega \to \mathbb{R}$, $x \in E$, satisfy for all $x \in E$, $\omega \in \Omega$ that $Y_x(\omega) = Z_x(\omega) - \mathbb{E}[Z_x]$. Note that it holds for all $\omega \in \Omega$, $x, y \in E$ that

$$\begin{aligned} |Y_x(\omega) - Y_y(\omega)| &= |(Z_x(\omega) - \mathbb{E}[Z_x]) - (Z_y(\omega) - \mathbb{E}[Z_y])| \\ &\le |Z_x(\omega) - Z_y(\omega)| + |\mathbb{E}[Z_x] - \mathbb{E}[Z_y]| \le L\delta(x,y) + \mathbb{E}[|Z_x - Z_y|] \\ &\le 2L\delta(x,y). \end{aligned} \tag{5.53}$$

Combining this with Lemma 5.22 (with $L \leftarrow 2L$, $(\Omega, \mathcal{F}, \mu) \leftarrow (\Omega, \mathcal{F}, \mathbb{P})$, $(Z_x)_{x\in E} \leftarrow (Y_x)_{x\in E}$ in the notation of Lemma 5.22) implies that

$$\begin{aligned} &\big(\mathbb{E}\big[\sup\nolimits_{x\in E}|Z_x - \mathbb{E}[Z_x]|^p\big]\big)^{1/p} = \big(\mathbb{E}\big[\sup\nolimits_{x\in E}|Y_x|^p\big]\big)^{1/p} \\ &\le (\mathcal{C}_{(E,\delta),r})^{1/p}\Big[\sup\nolimits_{x\in E}\big(\mathbb{E}\big[(2Lr + |Y_x|)^p\big]\big)^{1/p}\Big] \\ &\le (\mathcal{C}_{(E,\delta),r})^{1/p}\Big[2Lr + \sup\nolimits_{x\in E}\big(\mathbb{E}\big[|Y_x|^p\big]\big)^{1/p}\Big] \\ &= (\mathcal{C}_{(E,\delta),r})^{1/p}\Big[2Lr + \sup\nolimits_{x\in E}\big(\mathbb{E}\big[|Z_x - \mathbb{E}[Z_x]|^p\big]\big)^{1/p}\Big]. \end{aligned} \tag{5.54}$$

The proof of Lemma 5.23 is thus complete. ☐

Lemma 5.24. *Let $M \in \mathbb{N}$, $p \in [2,\infty)$, $L, r \in (0,\infty)$, let (E,δ) be a separable metric space, assume $E \neq \emptyset$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, for every $x \in E$ let $Y_{x,j}\colon \Omega \to \mathbb{R}$, $j \in \{1,2,\ldots,M\}$, be independent random variables, assume for all $x, y \in E$, $j \in \{1,2,\ldots,M\}$ that $\mathbb{E}[|Y_{x,j}|] < \infty$ and $|Y_{x,j} - Y_{y,j}| \le L\delta(x,y)$, and let $Z_x\colon \Omega \to \mathbb{R}$, $x \in E$, satisfy for all $x \in E$ that*

$$Z_x = \frac{1}{M}\left[\sum_{j=1}^{M} Y_{x,j}\right]. \tag{5.55}$$

Then

(i) it holds for all $x \in E$ that $\mathbb{E}[|Z_x|] < \infty$,

(ii) it holds that the function $\Omega \ni \omega \mapsto \sup_{x\in E}|Z_x(\omega) - \mathbb{E}[Z_x]| \in [0,\infty]$ is $\mathcal{F}/\mathcal{B}([0,\infty])$-measurable, and

(iii) it holds that

$$\begin{aligned} &\big(\mathbb{E}\big[\sup\nolimits_{x\in E}|Z_x - \mathbb{E}[Z_x]|^p\big]\big)^{1/p} \\ &\le 2(\mathcal{C}_{(E,\delta),r})^{1/p}\Big[Lr + \tfrac{\sqrt{p-1}}{\sqrt{M}}\Big(\sup\nolimits_{x\in E}\max\nolimits_{j\in\{1,2,\ldots,M\}}\big(\mathbb{E}\big[|Y_{x,j} - \mathbb{E}[Y_{x,j}]|^p\big]\big)^{1/p}\Big)\Big] \end{aligned} \tag{5.56}$$

(cf. Definition 5.10).

Proof of Lemma 5.24. Note that the assumption that $\forall\, x \in E, j \in \{1,2,\ldots,M\}\colon \mathbb{E}[|Y_{x,j}|] < \infty$ implies for all $x \in E$ that

$$\mathbb{E}[|Z_x|] = \mathbb{E}\left[\frac{1}{M}\left|\sum_{j=1}^{M} Y_{x,j}\right|\right] \le \frac{1}{M}\left[\sum_{j=1}^{M}\mathbb{E}[|Y_{x,j}|]\right] \le \max_{j\in\{1,2,\ldots,M\}}\mathbb{E}[|Y_{x,j}|] < \infty. \tag{5.57}$$

This proves (i). Next observe that the assumption that $\forall\, x, y \in E, j \in \{1,2,\ldots,M\}\colon |Y_{x,j} - Y_{y,j}| \le L\delta(x,y)$ demonstrates for all $x, y \in E$ that

$$|Z_x - Z_y| = \frac{1}{M}\left|\left[\sum_{j=1}^{M} Y_{x,j}\right] - \left[\sum_{j=1}^{M} Y_{y,j}\right]\right| \le \frac{1}{M}\left[\sum_{j=1}^{M}|Y_{x,j} - Y_{y,j}|\right] \le L\delta(x,y). \tag{5.58}$$

Combining this with (i) and Lemma 5.20 establishes (ii). It thus remains to show (iii). For this note that (i), (5.58), and Lemma 5.23 yield that

$$\big(\mathbb{E}\big[\sup_{x\in E}|Z_x-\mathbb{E}[Z_x]|^p\big]\big)^{1/p}\leq(\mathcal{C}_{(E,\delta),r})^{1/p}\Big[2Lr+\sup_{x\in E}\big(\mathbb{E}\big[|Z_x-\mathbb{E}[Z_x]|^p\big]\big)^{1/p}\Big]. \tag{5.59}$$

Moreover, (5.57) and Corollary 5.18 (with $d\leftarrow 1$, $(X_j)_{j\in\{1,2,\ldots,M\}}\leftarrow(Y_{x,j})_{j\in\{1,2,\ldots,M\}}$ for $x\in E$ in the notation of Corollary 5.18) prove for all $x\in E$ that

$$\begin{aligned}\big(\mathbb{E}\big[|Z_x-\mathbb{E}[Z_x]|^p\big]\big)^{1/p}&=\left(\mathbb{E}\left[\left|\frac{1}{M}\left[\sum_{j=1}^M Y_{x,j}\right]-\mathbb{E}\left[\frac{1}{M}\sum_{j=1}^M Y_{x,j}\right]\right|^p\right]\right)^{1/p}\\&\leq\frac{2\sqrt{p-1}}{\sqrt{M}}\left[\max_{j\in\{1,2,\ldots,M\}}\big(\mathbb{E}\big[|Y_{x,j}-\mathbb{E}[Y_{x,j}]|^p\big]\big)^{1/p}\right].\end{aligned} \tag{5.60}$$

This and (5.59) imply that

$$\begin{aligned}&\big(\mathbb{E}\big[\sup_{x\in E}|Z_x-\mathbb{E}[Z_x]|^p\big]\big)^{1/p}\\&\leq(\mathcal{C}_{(E,\delta),r})^{1/p}\Big[2Lr+\tfrac{2\sqrt{p-1}}{\sqrt{M}}\Big(\sup_{x\in E}\max_{j\in\{1,2,\ldots,M\}}\big(\mathbb{E}\big[|Y_{x,j}-\mathbb{E}[Y_{x,j}]|^p\big]\big)^{1/p}\Big)\Big]\\&=2(\mathcal{C}_{(E,\delta),r})^{1/p}\Big[Lr+\tfrac{\sqrt{p-1}}{\sqrt{M}}\Big(\sup_{x\in E}\max_{j\in\{1,2,\ldots,M\}}\big(\mathbb{E}\big[|Y_{x,j}-\mathbb{E}[Y_{x,j}]|^p\big]\big)^{1/p}\Big)\Big].\end{aligned} \tag{5.61}$$

The proof of Lemma 5.24 is thus complete. □

Corollary 5.25. *Let $M\in\mathbb{N}$, $p\in[2,\infty)$, $L,C\in(0,\infty)$, let (E,δ) be a separable metric space, assume $E\neq\emptyset$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, for every $x\in E$ let $Y_{x,j}\colon\Omega\to\mathbb{R}$, $j\in\{1,2,\ldots,M\}$, be independent random variables, assume for all $x,y\in E$, $j\in\{1,2,\ldots,M\}$ that $\mathbb{E}[|Y_{x,j}|]<\infty$ and $|Y_{x,j}-Y_{y,j}|\leq L\delta(x,y)$, and let $Z_x\colon\Omega\to\mathbb{R}$, $x\in E$, satisfy for all $x\in E$ that*

$$Z_x=\frac{1}{M}\left[\sum_{j=1}^M Y_{x,j}\right]. \tag{5.62}$$

Then

(i) it holds for all $x\in E$ that $\mathbb{E}[|Z_x|]<\infty$,

(ii) it holds that the function $\Omega\ni\omega\mapsto\sup_{x\in E}|Z_x(\omega)-\mathbb{E}[Z_x]|\in[0,\infty]$ is $\mathcal{F}/\mathcal{B}([0,\infty])$-measurable, and

(iii) it holds that

$$\begin{aligned}&\big(\mathbb{E}\big[\sup_{x\in E}|Z_x-\mathbb{E}[Z_x]|^p\big]\big)^{1/p}\\&\leq\tfrac{2\sqrt{p-1}}{\sqrt{M}}\Big(\mathcal{C}_{(E,\delta),\frac{C\sqrt{p-1}}{L\sqrt{M}}}\Big)^{1/p}\Big[C+\sup_{x\in E}\max_{j\in\{1,2,\ldots,M\}}\big(\mathbb{E}\big[|Y_{x,j}-\mathbb{E}[Y_{x,j}]|^p\big]\big)^{1/p}\Big]\end{aligned} \tag{5.63}$$

(cf. Definition 5.10).

Proof of Corollary 5.25. Note that Lemma 5.24 shows (i) and (ii). In addition, Lemma 5.24 (with $r\leftarrow C\sqrt{p-1}/(L\sqrt{M})$ in the notation of Lemma 5.24) ensures that

$$\begin{aligned}&\big(\mathbb{E}\big[\sup_{x\in E}|Z_x-\mathbb{E}[Z_x]|^p\big]\big)^{1/p}\\&\leq 2\Big(\mathcal{C}_{(E,\delta),\frac{C\sqrt{p-1}}{L\sqrt{M}}}\Big)^{1/p}\Big[L\tfrac{C\sqrt{p-1}}{L\sqrt{M}}+\tfrac{\sqrt{p-1}}{\sqrt{M}}\Big(\sup_{x\in E}\max_{j\in\{1,2,\ldots,M\}}\big(\mathbb{E}\big[|Y_{x,j}-\mathbb{E}[Y_{x,j}]|^p\big]\big)^{1/p}\Big)\Big]\\&=\tfrac{2\sqrt{p-1}}{\sqrt{M}}\Big(\mathcal{C}_{(E,\delta),\frac{C\sqrt{p-1}}{L\sqrt{M}}}\Big)^{1/p}\Big[C+\sup_{x\in E}\max_{j\in\{1,2,\ldots,M\}}\big(\mathbb{E}\big[|Y_{x,j}-\mathbb{E}[Y_{x,j}]|^p\big]\big)^{1/p}\Big].\end{aligned} \tag{5.64}$$

This establishes (iii) and thus completes the proof of Corollary 5.25. □

5.3.3 Strong convergence rates for the generalisation error

Lemma 5.26. *Let $M \in \mathbb{N}$, $p \in [2,\infty)$, $L, C, b \in (0,\infty)$, let (E,δ) be a separable metric space, assume $E \neq \emptyset$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X_{x,j} \colon \Omega \to \mathbb{R}$, $j \in \{1,2,\dots,M\}$, $x \in E$, and $Y_j \colon \Omega \to \mathbb{R}$, $j \in \{1,2,\dots,M\}$, be functions, assume for every $x \in E$ that $(X_{x,j}, Y_j)$, $j \in \{1,2,\dots,M\}$, are i.i.d. random variables, assume for all $x, y \in E$, $j \in \{1,2,\dots,M\}$ that $|X_{x,j} - Y_j| \leq b$ and $|X_{x,j} - X_{y,j}| \leq L\delta(x,y)$, let $\mathbf{R} \colon E \to [0,\infty)$ satisfy for all $x \in E$ that $\mathbf{R}(x) = \mathbb{E}[|X_{x,1} - Y_1|^2]$, and let $\mathcal{R} \colon E \times \Omega \to [0,\infty)$ satisfy for all $x \in E$, $\omega \in \Omega$ that*

$$\mathcal{R}(x,\omega) = \frac{1}{M}\left[\sum_{j=1}^{M} |X_{x,j}(\omega) - Y_j(\omega)|^2\right]. \tag{5.65}$$

Then

(i) it holds that the function $\Omega \ni \omega \mapsto \sup_{x\in E} |\mathcal{R}(x,\omega) - \mathbf{R}(x)| \in [0,\infty]$ is $\mathcal{F}/\mathcal{B}([0,\infty])$-measurable and

(ii) it holds that

$$\left(\mathbb{E}\left[\sup_{x\in E} |\mathcal{R}(x) - \mathbf{R}(x)|^p\right]\right)^{1/p} \leq \left(\mathcal{C}_{(E,\delta),\frac{Cb\sqrt{p-1}}{2L\sqrt{M}}}\right)^{1/p} \left[\frac{2(C+1)b^2\sqrt{p-1}}{\sqrt{M}}\right] \tag{5.66}$$

(cf. Definition 5.10).

Proof of Lemma 5.26. Throughout this proof let $\mathcal{Y}_{x,j} \colon \Omega \to \mathbb{R}$, $j \in \{1,2,\dots,M\}$, $x \in E$, satisfy for all $x \in E$, $j \in \{1,2,\dots,M\}$ that $\mathcal{Y}_{x,j} = |X_{x,j} - Y_j|^2$. Note that the assumption that for every $x \in E$ it holds that $(X_{x,j}, Y_j)$, $j \in \{1,2,\dots,M\}$, are i.i.d. random variables ensures for all $x \in E$ that

$$\mathbb{E}[\mathcal{R}(x)] = \frac{1}{M}\left[\sum_{j=1}^{M} \mathbb{E}\big[|X_{x,j} - Y_j|^2\big]\right] = \frac{M\,\mathbb{E}\big[|X_{x,1} - Y_1|^2\big]}{M} = \mathbf{R}(x). \tag{5.67}$$

Furthermore, the assumption that $\forall\, x \in E, j \in \{1,2,\dots,M\} \colon |X_{x,j} - Y_j| \leq b$ shows for all $x \in E$, $j \in \{1,2,\dots,M\}$ that

$$\mathbb{E}[|\mathcal{Y}_{x,j}|] = \mathbb{E}[|X_{x,j} - Y_j|^2] \leq b^2 < \infty, \tag{5.68}$$

$$\mathcal{Y}_{x,j} - \mathbb{E}[\mathcal{Y}_{x,j}] = |X_{x,j} - Y_j|^2 - \mathbb{E}\big[|X_{x,j} - Y_j|^2\big] \leq |X_{x,j} - Y_j|^2 \leq b^2, \tag{5.69}$$

and

$$\mathbb{E}[\mathcal{Y}_{x,j}] - \mathcal{Y}_{x,j} = \mathbb{E}\big[|X_{x,j} - Y_j|^2\big] - |X_{x,j} - Y_j|^2 \leq \mathbb{E}\big[|X_{x,j} - Y_j|^2\big] \leq b^2. \tag{5.70}$$

Combining (5.68)–(5.70) implies for all $x \in E$, $j \in \{1,2,\dots,M\}$ that

$$\left(\mathbb{E}\big[|\mathcal{Y}_{x,j} - \mathbb{E}[\mathcal{Y}_{x,j}]|^p\big]\right)^{1/p} \leq \left(\mathbb{E}\big[b^{2p}\big]\right)^{1/p} = b^2. \tag{5.71}$$

Moreover, note that the assumptions that $\forall\, x, y \in E, j \in \{1,2,\dots,M\} \colon [|X_{x,j} - Y_j| \leq b$ and $|X_{x,j} - X_{y,j}| \leq L\delta(x,y)]$ and the fact that $\forall\, x_1, x_2, y \in \mathbb{R} \colon (x_1 - y)^2 - (x_2 - y)^2 = (x_1 - x_2)((x_1 - y) + (x_2 - y))$ establish for all $x, y \in E$, $j \in \{1,2,\dots,M\}$ that

$$\begin{aligned} |\mathcal{Y}_{x,j} - \mathcal{Y}_{y,j}| &= |(X_{x,j} - Y_j)^2 - (X_{y,j} - Y_j)^2| \\ &\leq |X_{x,j} - X_{y,j}|(|X_{x,j} - Y_j| + |X_{y,j} - Y_j|) \\ &\leq 2b|X_{x,j} - X_{y,j}| \leq 2bL\delta(x,y). \end{aligned} \tag{5.72}$$

Combining this, (5.67), (5.68), and the fact that for every $x \in E$ it holds that $\mathcal{Y}_{x,j}$, $j \in \{1, 2, \ldots, M\}$, are independent random variables with Corollary 5.25 (with $L \leftarrow 2bL$, $C \leftarrow Cb^2$, $(Y_{x,j})_{x\in E, j\in\{1,2,\ldots,M\}} \leftarrow (\mathcal{Y}_{x,j})_{x\in E, j\in\{1,2,\ldots,M\}}$, $(Z_x)_{x\in E} \leftarrow (\Omega \ni \omega \mapsto \mathcal{R}(x, \omega) \in \mathbb{R})_{x\in E}$ in the notation of Corollary 5.25) and (5.71) proves (i) and

$$\begin{aligned}
&\big(\mathbb{E}\big[\sup_{x\in E}|\mathcal{R}(x) - \mathbf{R}(x)|^p\big]\big)^{1/p} = \big(\mathbb{E}\big[\sup_{x\in E}|\mathcal{R}(x) - \mathbb{E}[\mathcal{R}(x)]|^p\big]\big)^{1/p} \\
&\leq \tfrac{2\sqrt{p-1}}{\sqrt{M}}\Big(\mathcal{C}_{(E,\delta),\frac{Cb^2\sqrt{p-1}}{2bL\sqrt{M}}}\Big)^{1/p}\Big[Cb^2 + \sup_{x\in E}\max_{j\in\{1,2,\ldots,M\}}\big(\mathbb{E}\big[|\mathcal{Y}_{x,j} - \mathbb{E}[\mathcal{Y}_{x,j}]|^p\big]\big)^{1/p}\Big] \\
&\leq \tfrac{2\sqrt{p-1}}{\sqrt{M}}\Big(\mathcal{C}_{(E,\delta),\frac{Cb\sqrt{p-1}}{2L\sqrt{M}}}\Big)^{1/p}[Cb^2 + b^2] = \Big(\mathcal{C}_{(E,\delta),\frac{Cb\sqrt{p-1}}{2L\sqrt{M}}}\Big)^{1/p}\left[\frac{2(C+1)b^2\sqrt{p-1}}{\sqrt{M}}\right].
\end{aligned} \tag{5.73}$$

This shows (ii) and thus completes the proof of Lemma 5.26. ☐

Proposition 5.27. *Let $d, \mathbf{d}, M \in \mathbb{N}$, $L, b \in (0, \infty)$, $\alpha \in \mathbb{R}$, $\beta \in (\alpha, \infty)$, $D \subseteq \mathbb{R}^d$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X_j \colon \Omega \to D$, $j \in \{1, 2, \ldots, M\}$, and $Y_j \colon \Omega \to \mathbb{R}$, $j \in \{1, 2, \ldots, M\}$, be functions, assume that (X_j, Y_j), $j \in \{1, 2, \ldots, M\}$, are i.i.d. random variables, let $f = (f_\theta)_{\theta\in[\alpha,\beta]^{\mathbf{d}}} \colon [\alpha, \beta]^{\mathbf{d}} \to C(D, \mathbb{R})$ be a function, assume for all $\theta, \vartheta \in [\alpha, \beta]^{\mathbf{d}}$, $j \in \{1, 2, \ldots, M\}$, $x \in D$ that $|f_\theta(X_j) - Y_j| \leq b$ and $|f_\theta(x) - f_\vartheta(x)| \leq L\|\theta - \vartheta\|_\infty$, let $\mathbf{R} \colon [\alpha, \beta]^{\mathbf{d}} \to [0, \infty)$ satisfy for all $\theta \in [\alpha, \beta]^{\mathbf{d}}$ that $\mathbf{R}(\theta) = \mathbb{E}[|f_\theta(X_1) - Y_1|^2]$, and let $\mathcal{R} \colon [\alpha, \beta]^{\mathbf{d}} \times \Omega \to [0, \infty)$ satisfy for all $\theta \in [\alpha, \beta]^{\mathbf{d}}$, $\omega \in \Omega$ that*

$$\mathcal{R}(\theta, \omega) = \frac{1}{M}\left[\sum_{j=1}^{M}|f_\theta(X_j(\omega)) - Y_j(\omega)|^2\right] \tag{5.74}$$

(cf. Definition 5.9). Then

(i) it holds that the function $\Omega \ni \omega \mapsto \sup_{\theta\in[\alpha,\beta]^{\mathbf{d}}}|\mathcal{R}(\theta, \omega) - \mathbf{R}(\theta)| \in [0, \infty]$ is $\mathcal{F}/\mathcal{B}([0, \infty])$-measurable and

(ii) it holds for all $p \in (0, \infty)$ that

$$\begin{aligned}
&\big(\mathbb{E}\big[\sup_{\theta\in[\alpha,\beta]^{\mathbf{d}}}|\mathcal{R}(\theta) - \mathbf{R}(\theta)|^p\big]\big)^{1/p} \\
&\leq \inf_{C,\varepsilon\in(0,\infty)}\left[\frac{2(C+1)b^2\max\{1, [2\sqrt{M}L(\beta-\alpha)(Cb)^{-1}]^\varepsilon\}\sqrt{\max\{1, p, \mathbf{d}/\varepsilon\}}}{\sqrt{M}}\right] \\
&\leq \inf_{C\in(0,\infty)}\left[\frac{2(C+1)b^2\sqrt{e\max\{1, p, \mathbf{d}\ln(4ML^2(\beta-\alpha)^2(Cb)^{-2})\}}}{\sqrt{M}}\right].
\end{aligned} \tag{5.75}$$

Proof of Proposition 5.27. Throughout this proof let $p \in (0, \infty)$, let $(\kappa_C)_{C\in(0,\infty)} \subseteq (0, \infty)$ satisfy for all $C \in (0, \infty)$ that $\kappa_C = {}^{2\sqrt{M}L(\beta-\alpha)}/{}_{(Cb)}$, let $\mathcal{X}_{\theta,j} \colon \Omega \to \mathbb{R}$, $j \in \{1, 2, \ldots, M\}$, $\theta \in [\alpha, \beta]^{\mathbf{d}}$, satisfy for all $\theta \in [\alpha, \beta]^{\mathbf{d}}$, $j \in \{1, 2, \ldots, M\}$ that $\mathcal{X}_{\theta,j} = f_\theta(X_j)$, and let $\delta \colon ([\alpha, \beta]^{\mathbf{d}}) \times ([\alpha, \beta]^{\mathbf{d}}) \to [0, \infty)$ satisfy for all $\theta, \vartheta \in [\alpha, \beta]^{\mathbf{d}}$ that $\delta(\theta, \vartheta) = \|\theta - \vartheta\|_\infty$. First of all, note that the assumption that $\forall\, \theta \in [\alpha, \beta]^{\mathbf{d}}, j \in \{1, 2, \ldots, M\} \colon |f_\theta(X_j) - Y_j| \leq b$ implies for all $\theta \in [\alpha, \beta]^{\mathbf{d}}$, $j \in \{1, 2, \ldots, M\}$ that

$$|\mathcal{X}_{\theta,j} - Y_j| = |f_\theta(X_j) - Y_j| \leq b. \tag{5.76}$$

In addition, the assumption that $\forall\, \theta, \vartheta \in [\alpha, \beta]^{\mathbf{d}}$, $x \in D\colon |f_\theta(x) - f_\vartheta(x)| \le L\|\theta - \vartheta\|_\infty$ ensures for all $\theta, \vartheta \in [\alpha, \beta]^{\mathbf{d}}$, $j \in \{1, 2, \dots, M\}$ that

$$|\mathcal{X}_{\theta,j} - \mathcal{X}_{\vartheta,j}| = |f_\theta(X_j) - f_\vartheta(X_j)| \le \sup\nolimits_{x\in D}|f_\theta(x) - f_\vartheta(x)| \le L\|\theta - \vartheta\|_\infty = L\delta(\theta, \vartheta). \tag{5.77}$$

Combining this, (5.76), and the fact that for every $\theta \in [\alpha, \beta]^{\mathbf{d}}$ it holds that $(\mathcal{X}_{\theta,j}, Y_j)$, $j \in \{1, 2, \dots, M\}$, are i.i.d. random variables with Lemma 5.26 (with $p \leftarrow q$, $C \leftarrow C$, $(E, \delta) \leftarrow ([\alpha, \beta]^{\mathbf{d}}, \delta)$, $(X_{x,j})_{x\in E,\, j\in\{1,2,\dots,M\}} \leftarrow (\mathcal{X}_{\theta,j})_{\theta\in[\alpha,\beta]^{\mathbf{d}},\, j\in\{1,2,\dots,M\}}$ for $q \in [2, \infty)$, $C \in (0, \infty)$ in the notation of Lemma 5.26) demonstrates for all $C \in (0, \infty)$, $q \in [2, \infty)$ that the function $\Omega \ni \omega \mapsto \sup_{\theta\in[\alpha,\beta]^{\mathbf{d}}}|\mathcal{R}(\theta, \omega) - \mathbf{R}(\theta)| \in [0, \infty]$ is $\mathcal{F}/\mathcal{B}([0, \infty])$-measurable and

$$\big(\mathbb{E}\big[\sup\nolimits_{\theta\in[\alpha,\beta]^{\mathbf{d}}}|\mathcal{R}(\theta) - \mathbf{R}(\theta)|^q\big]\big)^{1/q} \le \Big(\mathcal{C}_{([\alpha,\beta]^{\mathbf{d}},\delta),\frac{Cb\sqrt{q-1}}{2L\sqrt{M}}}\Big)^{1/q}\left[\frac{2(C+1)b^2\sqrt{q-1}}{\sqrt{M}}\right] \tag{5.78}$$

(cf. Definition 5.10). This finishes the proof of (i). Next observe that (ii) in Lemma 5.11 (with $d \leftarrow \mathbf{d}$, $a \leftarrow \alpha$, $b \leftarrow \beta$, $r \leftarrow r$ for $r \in (0, \infty)$ in the notation of Lemma 5.11) shows for all $r \in (0, \infty)$ that

$$\begin{aligned}
\mathcal{C}_{([\alpha,\beta]^{\mathbf{d}},\delta),r} &\le \mathbb{1}_{[0,r]}\big(\tfrac{\beta-\alpha}{2}\big) + \big(\tfrac{\beta-\alpha}{r}\big)^{\mathbf{d}}\mathbb{1}_{(r,\infty)}\big(\tfrac{\beta-\alpha}{2}\big)\\
&\le \max\Big\{1, \big(\tfrac{\beta-\alpha}{r}\big)^{\mathbf{d}}\Big\}\big(\mathbb{1}_{[0,r]}\big(\tfrac{\beta-\alpha}{2}\big) + \mathbb{1}_{(r,\infty)}\big(\tfrac{\beta-\alpha}{2}\big)\big)\\
&= \max\Big\{1, \big(\tfrac{\beta-\alpha}{r}\big)^{\mathbf{d}}\Big\}.
\end{aligned} \tag{5.79}$$

This yields for all $C \in (0, \infty)$, $q \in [2, \infty)$ that

$$\begin{aligned}
\Big(\mathcal{C}_{([\alpha,\beta]^{\mathbf{d}},\delta),\frac{Cb\sqrt{q-1}}{2L\sqrt{M}}}\Big)^{1/q} &\le \max\bigg\{1, \Big(\tfrac{2(\beta-\alpha)L\sqrt{M}}{Cb\sqrt{q-1}}\Big)^{\frac{\mathbf{d}}{q}}\bigg\}\\
&\le \max\bigg\{1, \Big(\tfrac{2(\beta-\alpha)L\sqrt{M}}{Cb}\Big)^{\frac{\mathbf{d}}{q}}\bigg\} = \max\Big\{1, (\kappa_C)^{\frac{\mathbf{d}}{q}}\Big\}.
\end{aligned} \tag{5.80}$$

Jensen's inequality and (5.78) hence prove for all $C, \varepsilon \in (0, \infty)$ that

$$\begin{aligned}
&\big(\mathbb{E}\big[\sup\nolimits_{\theta\in[\alpha,\beta]^{\mathbf{d}}}|\mathcal{R}(\theta) - \mathbf{R}(\theta)|^p\big]\big)^{1/p}\\
&\le \big(\mathbb{E}\big[\sup\nolimits_{\theta\in[\alpha,\beta]^{\mathbf{d}}}|\mathcal{R}(\theta) - \mathbf{R}(\theta)|^{\max\{2,p,\mathbf{d}/\varepsilon\}}\big]\big)^{\frac{1}{\max\{2,p,\mathbf{d}/\varepsilon\}}}\\
&\le \max\Big\{1, (\kappa_C)^{\frac{\mathbf{d}}{\max\{2,p,\mathbf{d}/\varepsilon\}}}\Big\}\frac{2(C+1)b^2\sqrt{\max\{2, p, \mathbf{d}/\varepsilon\} - 1}}{\sqrt{M}}\\
&= \max\big\{1, (\kappa_C)^{\min\{\mathbf{d}/2,\mathbf{d}/p,\varepsilon\}}\big\}\frac{2(C+1)b^2\sqrt{\max\{1, p-1, \mathbf{d}/\varepsilon - 1\}}}{\sqrt{M}}\\
&\le \frac{2(C+1)b^2\max\{1, (\kappa_C)^\varepsilon\}\sqrt{\max\{1, p, \mathbf{d}/\varepsilon\}}}{\sqrt{M}}.
\end{aligned} \tag{5.81}$$

Next note that the fact that $\forall\, a \in (1, \infty)\colon a^{1/(2\ln(a))} = e^{\ln(a)/(2\ln(a))} = e^{1/2} = \sqrt{e} \ge 1$ ensures for all $C \in (0,\infty)$ with $\kappa_C > 1$ that

$$\begin{aligned}
&\inf_{\varepsilon\in(0,\infty)}\left[\frac{2(C+1)b^2\max\{1,(\kappa_C)^\varepsilon\}\sqrt{\max\{1,p,{}^{\mathbf{d}}\!/\!_\varepsilon\}}}{\sqrt{M}}\right]\\
&\le \frac{2(C+1)b^2\max\{1,(\kappa_C)^{1/(2\ln(\kappa_C))}\}\sqrt{\max\{1,p,2\mathbf{d}\ln(\kappa_C)\}}}{\sqrt{M}}\\
&= \frac{2(C+1)b^2\sqrt{e\max\{1,p,\mathbf{d}\ln([\kappa_C]^2)\}}}{\sqrt{M}}.
\end{aligned} \tag{5.82}$$

In addition, observe that it holds for all $C \in (0,\infty)$ with $\kappa_C \le 1$ that

$$\begin{aligned}
&\inf_{\varepsilon\in(0,\infty)}\left[\frac{2(C+1)b^2\max\{1,(\kappa_C)^\varepsilon\}\sqrt{\max\{1,p,{}^{\mathbf{d}}\!/\!_\varepsilon\}}}{\sqrt{M}}\right]\\
&= \inf_{\varepsilon\in(0,\infty)}\left[\frac{2(C+1)b^2\sqrt{\max\{1,p,{}^{\mathbf{d}}\!/\!_\varepsilon\}}}{\sqrt{M}}\right] \le \frac{2(C+1)b^2\sqrt{\max\{1,p\}}}{\sqrt{M}}\\
&\le \frac{2(C+1)b^2\sqrt{e\max\{1,p,\mathbf{d}\ln([\kappa_C]^2)\}}}{\sqrt{M}}.
\end{aligned} \tag{5.83}$$

Combining (5.81) with (5.82) and (5.83) demonstrates that

$$\begin{aligned}
&\big(\mathbb{E}\big[\sup\nolimits_{\theta\in[\alpha,\beta]^{\mathbf{d}}}|\mathcal{R}(\theta)-\mathbf{R}(\theta)|^p\big]\big)^{1/p}\\
&\le \inf_{C,\varepsilon\in(0,\infty)}\left[\frac{2(C+1)b^2\max\{1,(\kappa_C)^\varepsilon\}\sqrt{\max\{1,p,{}^{\mathbf{d}}\!/\!_\varepsilon\}}}{\sqrt{M}}\right]\\
&= \inf_{C,\varepsilon\in(0,\infty)}\left[\frac{2(C+1)b^2\max\{1,[2\sqrt{M}L(\beta-\alpha)(Cb)^{-1}]^\varepsilon\}\sqrt{\max\{1,p,{}^{\mathbf{d}}\!/\!_\varepsilon\}}}{\sqrt{M}}\right]\\
&\le \inf_{C\in(0,\infty)}\left[\frac{2(C+1)b^2\sqrt{e\max\{1,p,\mathbf{d}\ln([\kappa_C]^2)\}}}{\sqrt{M}}\right]\\
&= \inf_{C\in(0,\infty)}\left[\frac{2(C+1)b^2\sqrt{e\max\{1,p,\mathbf{d}\ln(4ML^2(\beta-\alpha)^2(Cb)^{-2})\}}}{\sqrt{M}}\right].
\end{aligned} \tag{5.84}$$

This establishes (ii) and thus completes the proof of Proposition 5.27. □

Corollary 5.28. *Let $d,\mathbf{d},\mathbf{L},M \in \mathbb{N}$, $B,b \in [1,\infty)$, $u \in \mathbb{R}$, $v \in [u+1,\infty)$, $\mathbf{l} = (\mathbf{l}_0,\mathbf{l}_1,\ldots,\mathbf{l}_\mathbf{L}) \in \mathbb{N}^{\mathbf{L}+1}$, $D \subseteq [-b,b]^d$, assume $\mathbf{l}_0 = d$, $\mathbf{l}_\mathbf{L} = 1$, and $\mathbf{d} \ge \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1}+1)$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $X_j\colon \Omega \to D$, $j \in \{1,2,\ldots,M\}$, and $Y_j\colon \Omega \to [u,v]$, $j \in \{1,2,\ldots,M\}$, be functions, assume that (X_j,Y_j), $j \in \{1,2,\ldots,M\}$, are i.i.d. random variables, let $\mathbf{R}\colon [-B,B]^\mathbf{d} \to [0,\infty)$ satisfy for all $\theta \in [-B,B]^\mathbf{d}$ that $\mathbf{R}(\theta) = \mathbb{E}[|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_1)-Y_1|^2]$, and let $\mathcal{R}\colon [-B,B]^\mathbf{d}\times\Omega \to [0,\infty)$ satisfy for all $\theta \in [-B,B]^\mathbf{d}$, $\omega \in \Omega$ that*

$$\mathcal{R}(\theta,\omega) = \frac{1}{M}\left[\sum_{j=1}^{M}|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_j(\omega))-Y_j(\omega)|^2\right] \tag{5.85}$$

(cf. Definition 5.8). Then

(i) it holds that the function $\Omega \ni \omega \mapsto \sup_{\theta\in[-B,B]^{\mathbf{d}}}|\mathcal{R}(\theta,\omega) - \mathbf{R}(\theta)| \in [0,\infty]$ *is* $\mathcal{F}/\mathcal{B}([0,\infty])$*-measurable and*

(ii) it holds for all $p \in (0,\infty)$ *that*

$$\begin{aligned}
&\big(\mathbb{E}\big[\sup_{\theta\in[-B,B]^{\mathbf{d}}}|\mathcal{R}(\theta) - \mathbf{R}(\theta)|^p\big]\big)^{1/p} \\
&\leq \frac{9(v-u)^2\mathbf{L}(\|\mathbf{l}\|_\infty+1)\sqrt{\max\{p,\ln(4(Mb)^{1/\mathbf{L}}(\|\mathbf{l}\|_\infty+1)B)\}}}{\sqrt{M}} \\
&\leq \frac{9(v-u)^2\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{p,\ln(3MBb)\}}{\sqrt{M}}
\end{aligned} \tag{5.86}$$

(cf. Definition 5.9).

Proof of Corollary 5.28. Throughout this proof let $\mathfrak{d} \in \mathbb{N}$ be given by $\mathfrak{d} = \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1}+1)$, let $L \in (0,\infty)$ be given by $L = b\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}B^{\mathbf{L}-1}$, let $f = (f_\theta)_{\theta\in[-B,B]^{\mathfrak{d}}}\colon [-B,B]^{\mathfrak{d}} \to C(D,\mathbb{R})$ satisfy for all $\theta \in [-B,B]^{\mathfrak{d}}$, $x \in D$ that $f_\theta(x) = \mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x)$, let $\mathscr{R}\colon [-B,B]^{\mathfrak{d}} \to [0,\infty)$ satisfy for all $\theta \in [-B,B]^{\mathfrak{d}}$ that $\mathscr{R}(\theta) = \mathbb{E}[|f_\theta(X_1) - Y_1|^2] = \mathbb{E}[|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_1) - Y_1|^2]$, and let $R\colon [-B,B]^{\mathfrak{d}} \times \Omega \to [0,\infty)$ satisfy for all $\theta \in [-B,B]^{\mathfrak{d}}$, $\omega \in \Omega$ that

$$R(\theta,\omega) = \frac{1}{M}\Bigg[\sum_{j=1}^{M}|f_\theta(X_j(\omega)) - Y_j(\omega)|^2\Bigg] = \frac{1}{M}\Bigg[\sum_{j=1}^{M}|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_j(\omega)) - Y_j(\omega)|^2\Bigg]. \tag{5.87}$$

Note that the fact that $\forall\, \theta \in \mathbb{R}^{\mathfrak{d}}, x \in \mathbb{R}^d\colon \mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x) \in [u,v]$ and the assumption that $\forall\, j \in \{1,2,\dots,M\}\colon Y_j(\Omega) \subseteq [u,v]$ imply for all $\theta \in [-B,B]^{\mathfrak{d}}$, $j \in \{1,2,\dots,M\}$ that

$$|f_\theta(X_j) - Y_j| = |\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_j) - Y_j| \leq \sup_{y_1,y_2\in[u,v]}|y_1 - y_2| = v - u. \tag{5.88}$$

Moreover, the assumptions that $D \subseteq [-b,b]^d$, $\mathbf{l}_0 = d$, and $\mathbf{l}_{\mathbf{L}} = 1$, Beck, Jentzen, & Kuckuck [27, Corollary 2.37] (with $a \leftarrow -b$, $b \leftarrow b$, $u \leftarrow u$, $v \leftarrow v$, $d \leftarrow \mathfrak{d}$, $L \leftarrow \mathbf{L}$, $l \leftarrow \mathbf{l}$ in the notation of [27, Corollary 2.37]), and the assumptions that $b \geq 1$ and $B \geq 1$ ensure for all $\theta, \vartheta \in [-B,B]^{\mathfrak{d}}$, $x \in D$ that

$$\begin{aligned}
|f_\theta(x) - f_\vartheta(x)| &\leq \sup_{y\in[-b,b]^d}|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(y) - \mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(y)| \\
&\leq \mathbf{L}\max\{1,b\}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}(\max\{1,\|\theta\|_\infty,\|\vartheta\|_\infty\})^{\mathbf{L}-1}\|\theta-\vartheta\|_\infty \\
&\leq b\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}B^{\mathbf{L}-1}\|\theta-\vartheta\|_\infty = L\|\theta-\vartheta\|_\infty.
\end{aligned} \tag{5.89}$$

Furthermore, the facts that $\mathbf{d} \geq \mathfrak{d}$ and $\forall\, \theta = (\theta_1,\theta_2,\dots,\theta_{\mathbf{d}}) \in \mathbb{R}^{\mathbf{d}}\colon \mathscr{N}^{\theta,\mathbf{l}}_{u,v} = \mathscr{N}^{(\theta_1,\theta_2,\dots,\theta_{\mathfrak{d}}),\mathbf{l}}_{u,v}$ prove for all $\omega \in \Omega$ that

$$\sup_{\theta\in[-B,B]^{\mathbf{d}}}|\mathcal{R}(\theta,\omega) - \mathbf{R}(\theta)| = \sup_{\theta\in[-B,B]^{\mathfrak{d}}}|R(\theta,\omega) - \mathscr{R}(\theta)|. \tag{5.90}$$

Next observe that (5.88), (5.89), Proposition 5.27 (with $\mathbf{d} \leftarrow \mathfrak{d}$, $b \leftarrow v-u$, $\alpha \leftarrow -B$, $\beta \leftarrow B$, $\mathbf{R} \leftarrow \mathscr{R}$, $\mathcal{R} \leftarrow R$ in the notation of Proposition 5.27), and the facts that $v - u \geq (u+1) - u = 1$ and $\mathfrak{d} \leq \mathbf{L}\|\mathbf{l}\|_\infty(\|\mathbf{l}\|_\infty+1) \leq \mathbf{L}(\|\mathbf{l}\|_\infty+1)^2$ demonstrate for all $p \in (0,\infty)$ that the function $\Omega \ni \omega \mapsto \sup_{\theta\in[-B,B]^{\mathfrak{d}}}|R(\theta,\omega) - \mathscr{R}(\theta)| \in [0,\infty]$ is $\mathcal{F}/\mathcal{B}([0,\infty])$-measurable and

$$\begin{aligned}
&\big(\mathbb{E}\big[\sup_{\theta\in[-B,B]^{\mathfrak{d}}}|R(\theta) - \mathscr{R}(\theta)|^p\big]\big)^{1/p} \\
&\leq \inf_{C\in(0,\infty)}\left[\frac{2(C+1)(v-u)^2\sqrt{e\max\{1,p,\mathfrak{d}\ln(4ML^2(2B)^2(C[v-u])^{-2})\}}}{\sqrt{M}}\right] \\
&\leq \inf_{C\in(0,\infty)}\left[\frac{2(C+1)(v-u)^2\sqrt{e\max\{1,p,\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\ln(2^4ML^2B^2C^{-2})\}}}{\sqrt{M}}\right].
\end{aligned} \tag{5.91}$$

This and (5.90) establish (i). In addition, combining (5.90)–(5.91) with the fact that $2^6\mathbf{L}^2 \leq 2^6 \cdot 2^{2(\mathbf{L}-1)} = 2^{4+2\mathbf{L}} \leq 2^{4\mathbf{L}+2\mathbf{L}} = 2^{6\mathbf{L}}$ and the facts that $3 \geq e$, $B \geq 1$, $\mathbf{L} \geq 1$, $M \geq 1$, and $b \geq 1$ shows for all $p \in (0, \infty)$ that

$$\begin{aligned}
&\big(\mathbb{E}\big[\sup\nolimits_{\theta\in[-B,B]^{\mathbf{d}}}|\mathcal{R}(\theta) - \mathbf{R}(\theta)|^p\big]\big)^{1/p} = \big(\mathbb{E}\big[\sup\nolimits_{\theta\in[-B,B]^{\mathfrak{d}}}|R(\theta) - \mathscr{R}(\theta)|^p\big]\big)^{1/p} \\
&\leq \frac{2(1/2+1)(v-u)^2\sqrt{e\max\{1, p, \mathbf{L}(\|\mathbf{l}\|_\infty + 1)^2\ln(2^4 M L^2 B^2 2^2)\}}}{\sqrt{M}} \\
&= \frac{3(v-u)^2\sqrt{e\max\{p, \mathbf{L}(\|\mathbf{l}\|_\infty + 1)^2\ln(2^6 M b^2 \mathbf{L}^2(\|\mathbf{l}\|_\infty + 1)^{2\mathbf{L}} B^{2\mathbf{L}})\}}}{\sqrt{M}} \\
&\leq \frac{3(v-u)^2\sqrt{e\max\{p, 3\mathbf{L}^2(\|\mathbf{l}\|_\infty + 1)^2\ln([2^{6\mathbf{L}} M b^2(\|\mathbf{l}\|_\infty + 1)^{2\mathbf{L}} B^{2\mathbf{L}}]^{1/(3\mathbf{L})})\}}}{\sqrt{M}} \\
&\leq \frac{3(v-u)^2\sqrt{3\max\{p, 3\mathbf{L}^2(\|\mathbf{l}\|_\infty + 1)^2\ln(2^2(Mb^2)^{1/(3\mathbf{L})}(\|\mathbf{l}\|_\infty + 1)B)\}}}{\sqrt{M}} \\
&\leq \frac{9(v-u)^2\mathbf{L}(\|\mathbf{l}\|_\infty + 1)\sqrt{\max\{p, \ln(4(Mb)^{1/\mathbf{L}}(\|\mathbf{l}\|_\infty + 1)B)\}}}{\sqrt{M}}.
\end{aligned} \tag{5.92}$$

Furthermore, note that the fact that $\forall\, n \in \mathbb{N}\colon n \leq 2^{n-1}$ and the fact that $\|\mathbf{l}\|_\infty \geq 1$ imply that

$$4(\|\mathbf{l}\|_\infty + 1) \leq 2^2 \cdot 2^{(\|\mathbf{l}\|_\infty+1)-1} = 2^3 \cdot 2^{(\|\mathbf{l}\|_\infty+1)-2} \leq 3^2 \cdot 3^{(\|\mathbf{l}\|_\infty+1)-2} = 3^{(\|\mathbf{l}\|_\infty+1)}. \tag{5.93}$$

This demonstrates for all $p \in (0, \infty)$ that

$$\begin{aligned}
&\frac{9(v-u)^2\mathbf{L}(\|\mathbf{l}\|_\infty + 1)\sqrt{\max\{p, \ln(4(Mb)^{1/\mathbf{L}}(\|\mathbf{l}\|_\infty + 1)B)\}}}{\sqrt{M}} \\
&\leq \frac{9(v-u)^2\mathbf{L}(\|\mathbf{l}\|_\infty + 1)\sqrt{\max\{p, (\|\mathbf{l}\|_\infty + 1)\ln([3^{(\|\mathbf{l}\|_\infty+1)}(Mb)^{1/\mathbf{L}}B]^{1/(\|\mathbf{l}\|_\infty+1)})\}}}{\sqrt{M}} \\
&\leq \frac{9(v-u)^2\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^2\max\{p, \ln(3MBb)\}}{\sqrt{M}}.
\end{aligned} \tag{5.94}$$

Combining this with (5.92) shows (ii). The proof of Corollary 5.28 is thus complete. □

5.4 Analysis of the optimisation error

The main result of this section, Proposition 5.34, establishes that the optimisation error of the Minimum Monte Carlo method applied to a Lipschitz continuous random field with a $\mathbf{d}$-dimensional hypercube as index set, where $\mathbf{d} \in \mathbb{N}$, converges in the probabilistically strong sense with rate $1/\mathbf{d}$ with respect to the number of samples used, provided that the sample indices are continuous uniformly drawn from the index hypercube (cf. (ii) in Proposition 5.34). We refer to Beck, Jentzen, & Kuckuck [27, Lemmas 3.22 and 3.23] for analogous results for convergence in probability instead of strong convergence and to Beck et al. [22, Lemma 3.5] for a related result. Corollary 5.36 below specialises Proposition 5.34 to the case where the empirical risk from deep learning based empirical risk minimisation with quadratic loss function indexed by a hypercube of DNN parameter vectors plays the role of the random field under consideration. In the proof of Corollary 5.36 we make use of the elementary and well-known fact that this choice for the random field is indeed Lipschitz continuous, which is the assertion of Lemma 5.35. Further results on the optimisation error in the context of stochastic approximation can be found, e.g., in [10, 12, 44, 72, 102, 103, 108, 109, 124, 161, 190, 195, 200, 222, 282, 311, 313] and the references therein.

The proof of the main result of this section, Proposition 5.34, crucially relies (cf. Lemma 5.33) on the complementary distribution function formula (cf., e.g., Elbrächter et al. [120, Lemma 2.2]) and the elementary estimate for the beta function given in Corollary 5.32. In order to prove Corollary 5.32, we first collect a few basic facts about the gamma and the beta function in the elementary and well-known Lemma 5.29 and derive from these in Proposition 5.31 further elementary and essentially well-known properties of the gamma function. In particular, the inequalities in (5.96) in Proposition 5.31 are slightly reformulated versions of the well-known inequalities called *Wendel's double inequality* (cf. Wendel [308]) or *Gautschi's double inequality* (cf. Gautschi [130]); cf., e.g., Qi [264, Subsections 2.1 and 2.4].

5.4.1 Properties of the gamma and the beta function

Lemma 5.29. *Let $\Gamma\colon (0,\infty)\to(0,\infty)$ satisfy for all $x\in(0,\infty)$ that $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}\,\mathrm{d}t$ and let $\mathbb{B}\colon (0,\infty)^2\to(0,\infty)$ satisfy for all $x,y\in(0,\infty)$ that $\mathbb{B}(x,y)=\int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t$. Then*

(i) it holds for all $x\in(0,\infty)$ that $\Gamma(x+1)=x\,\Gamma(x)$,

(ii) it holds that $\Gamma(1)=\Gamma(2)=1$, and

(iii) it holds for all $x,y\in(0,\infty)$ that $\mathbb{B}(x,y)=\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$.

Lemma 5.30. *It holds for all $\alpha,x\in[0,1]$ that $(1-x)^\alpha\le 1-\alpha x$.*

Proof of Lemma 5.30. Note that the fact that for every $y\in[0,\infty)$ it holds that the function $[0,\infty)\ni z\mapsto y^z\in[0,\infty)$ is a convex function implies for all $\alpha,x\in[0,1]$ that

$$\begin{aligned}(1-x)^\alpha &= (1-x)^{\alpha\cdot 1+(1-\alpha)\cdot 0}\\ &\le \alpha(1-x)^1+(1-\alpha)(1-x)^0\\ &= \alpha-\alpha x+1-\alpha = 1-\alpha x.\end{aligned} \tag{5.95}$$

The proof of Lemma 5.30 is thus complete. □

Proposition 5.31. *Let $\Gamma\colon (0,\infty)\to(0,\infty)$ satisfy for all $x\in(0,\infty)$ that $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}\,\mathrm{d}t$ and let $\llcorner\cdot\lrcorner\colon(0,\infty)\to\mathbb{N}_0$ satisfy for all $x\in(0,\infty)$ that $\llcorner x\lrcorner=\max([0,x)\cap\mathbb{N}_0)$. Then*

(i) it holds that $\Gamma\colon(0,\infty)\to(0,\infty)$ is a convex function,

(ii) it holds for all $x\in(0,\infty)$ that $\Gamma(x+1)=x\,\Gamma(x)\le x^{\llcorner x\lrcorner}\le\max\{1,x^x\}$,

(iii) it holds for all $x\in(0,\infty)$, $\alpha\in[0,1]$ that

$$(\max\{x+\alpha-1,0\})^\alpha\le\frac{x}{(x+\alpha)^{1-\alpha}}\le\frac{\Gamma(x+\alpha)}{\Gamma(x)}\le x^\alpha, \tag{5.96}$$

and

(iv) it holds for all $x \in (0,\infty)$, $\alpha \in [0,\infty)$ that

$$(\max\{x+\min\{\alpha-1,0\},0\})^\alpha \le \frac{\Gamma(x+\alpha)}{\Gamma(x)} \le (x+\max\{\alpha-1,0\})^\alpha. \tag{5.97}$$

Proof of Proposition 5.31. First, observe that the fact that for every $t \in (0,\infty)$ it holds that the function $\mathbb{R} \ni x \mapsto t^x \in (0,\infty)$ is a convex function implies for all $x, y \in (0,\infty)$, $\alpha \in [0,1]$ that

$$\begin{aligned}\Gamma(\alpha x+(1-\alpha)y) &= \int_0^\infty t^{\alpha x+(1-\alpha)y-1}e^{-t}\,\mathrm{d}t = \int_0^\infty t^{\alpha x+(1-\alpha)y}t^{-1}e^{-t}\,\mathrm{d}t \\ &\le \int_0^\infty (\alpha t^x+(1-\alpha)t^y)t^{-1}e^{-t}\,\mathrm{d}t \\ &= \alpha\int_0^\infty t^{x-1}e^{-t}\,\mathrm{d}t + (1-\alpha)\int_0^\infty t^{y-1}e^{-t}\,\mathrm{d}t \\ &= \alpha\,\Gamma(x)+(1-\alpha)\Gamma(y).\end{aligned} \tag{5.98}$$

This shows (i).

Second, note that (ii) in Lemma 5.29 and (i) establish for all $\alpha \in [0,1]$ that

$$\Gamma(\alpha+1) = \Gamma(\alpha\cdot 2+(1-\alpha)\cdot 1) \le \alpha\,\Gamma(2)+(1-\alpha)\Gamma(1) = \alpha+(1-\alpha) = 1. \tag{5.99}$$

This yields for all $x \in (0,1]$ that

$$\Gamma(x+1) \le 1 = x^{\llcorner x \lrcorner} = \max\{1, x^x\}. \tag{5.100}$$

Induction, (i) in Lemma 5.29, and the fact that $\forall\, x \in (0,\infty)\colon x - \llcorner x \lrcorner \in (0,1]$ hence ensure for all $x \in [1,\infty)$ that

$$\Gamma(x+1) = \left[\prod_{i=1}^{\llcorner x \lrcorner}(x-i+1)\right]\Gamma(x-\llcorner x \lrcorner+1) \le x^{\llcorner x \lrcorner}\Gamma(x-\llcorner x \lrcorner+1) \le x^{\llcorner x \lrcorner} \le x^x = \max\{1,x^x\}. \tag{5.101}$$

Combining this with again (i) in Lemma 5.29 and (5.100) establishes (ii).

Third, note that Hölder's inequality and (i) in Lemma 5.29 prove for all $x \in (0,\infty)$, $\alpha \in [0,1]$ that

$$\begin{aligned}\Gamma(x+\alpha) &= \int_0^\infty t^{x+\alpha-1}e^{-t}\,\mathrm{d}t = \int_0^\infty t^{\alpha x}e^{-\alpha t}t^{(1-\alpha)x-(1-\alpha)}e^{-(1-\alpha)t}\,\mathrm{d}t \\ &= \int_0^\infty [t^x e^{-t}]^\alpha [t^{x-1}e^{-t}]^{1-\alpha}\,\mathrm{d}t \\ &\le \left(\int_0^\infty t^x e^{-t}\,\mathrm{d}t\right)^\alpha \left(\int_0^\infty t^{x-1}e^{-t}\,\mathrm{d}t\right)^{1-\alpha} \\ &= [\Gamma(x+1)]^\alpha[\Gamma(x)]^{1-\alpha} = x^\alpha[\Gamma(x)]^\alpha[\Gamma(x)]^{1-\alpha} = x^\alpha\Gamma(x).\end{aligned} \tag{5.102}$$

This and again (i) in Lemma 5.29 demonstrate for all $x \in (0,\infty)$, $\alpha \in [0,1]$ that

$$x\,\Gamma(x) = \Gamma(x+1) = \Gamma(x+\alpha+(1-\alpha)) \le (x+\alpha)^{1-\alpha}\Gamma(x+\alpha). \tag{5.103}$$

Combining (5.102) and (5.103) yields for all $x \in (0, \infty)$, $\alpha \in [0, 1]$ that

$$\frac{x}{(x+\alpha)^{1-\alpha}} \leq \frac{\Gamma(x+\alpha)}{\Gamma(x)} \leq x^{\alpha}. \tag{5.104}$$

Furthermore, observe that (i) in Lemma 5.29 and (5.104) imply for all $x \in (0, \infty)$, $\alpha \in [0, 1]$ that

$$\frac{\Gamma(x+\alpha)}{\Gamma(x+1)} = \frac{\Gamma(x+\alpha)}{x\,\Gamma(x)} \leq x^{\alpha-1}. \tag{5.105}$$

This shows for all $\alpha \in [0, 1]$, $x \in (\alpha, \infty)$ that

$$\frac{\Gamma(x)}{\Gamma(x+(1-\alpha))} = \frac{\Gamma((x-\alpha)+\alpha)}{\Gamma((x-\alpha)+1)} \leq (x-\alpha)^{\alpha-1} = \frac{1}{(x-\alpha)^{1-\alpha}}. \tag{5.106}$$

This, in turn, ensures for all $\alpha \in [0, 1]$, $x \in (1-\alpha, \infty)$ that

$$(x+\alpha-1)^{\alpha} = (x-(1-\alpha))^{\alpha} \leq \frac{\Gamma(x+\alpha)}{\Gamma(x)}. \tag{5.107}$$

Next note that Lemma 5.30 proves for all $x \in (0, \infty)$, $\alpha \in [0, 1]$ that

$$\begin{aligned}
(\max\{x+\alpha-1, 0\})^{\alpha} &= (x+\alpha)^{\alpha}\left(\frac{\max\{x+\alpha-1, 0\}}{x+\alpha}\right)^{\alpha} \\
&= (x+\alpha)^{\alpha}\left(\max\left\{1-\frac{1}{x+\alpha}, 0\right\}\right)^{\alpha} \\
&\leq (x+\alpha)^{\alpha}\left(1-\frac{\alpha}{x+\alpha}\right) = (x+\alpha)^{\alpha}\left(\frac{x}{x+\alpha}\right) \\
&= \frac{x}{(x+\alpha)^{1-\alpha}}.
\end{aligned} \tag{5.108}$$

This and (5.104) establish (iii).

Fourth, we show (iv). For this let $\lfloor \cdot \rfloor \colon [0, \infty) \to \mathbb{N}_0$ satisfy for all $x \in [0, \infty)$ that $\lfloor x \rfloor = \max([0, x] \cap \mathbb{N}_0)$. Observe that induction, (i) in Lemma 5.29, the fact that $\forall\, \alpha \in [0, \infty) \colon \alpha - \lfloor \alpha \rfloor \in [0, 1)$, and (iii) demonstrate for all $x \in (0, \infty)$, $\alpha \in [0, \infty)$ that

$$\begin{aligned}
\frac{\Gamma(x+\alpha)}{\Gamma(x)} &= \left[\prod_{i=1}^{\lfloor \alpha \rfloor}(x+\alpha-i)\right]\frac{\Gamma(x+\alpha-\lfloor \alpha \rfloor)}{\Gamma(x)} \leq \left[\prod_{i=1}^{\lfloor \alpha \rfloor}(x+\alpha-i)\right]x^{\alpha-\lfloor \alpha \rfloor} \\
&\leq (x+\alpha-1)^{\lfloor \alpha \rfloor}x^{\alpha-\lfloor \alpha \rfloor} \\
&\leq (x+\max\{\alpha-1, 0\})^{\lfloor \alpha \rfloor}(x+\max\{\alpha-1, 0\})^{\alpha-\lfloor \alpha \rfloor} \\
&= (x+\max\{\alpha-1, 0\})^{\alpha}.
\end{aligned} \tag{5.109}$$

Furthermore, again the fact that $\forall\, \alpha \in [0, \infty) \colon \alpha - \lfloor \alpha \rfloor \in [0, 1)$, (iii), induction, and (i) in

Lemma 5.29 imply for all $x \in (0,\infty)$, $\alpha \in [0,\infty)$ that

$$\begin{aligned}
\frac{\Gamma(x+\alpha)}{\Gamma(x)} &= \frac{\Gamma(x+\lfloor\alpha\rfloor+\alpha-\lfloor\alpha\rfloor)}{\Gamma(x)} \\
&\geq (\max\{x+\lfloor\alpha\rfloor+\alpha-\lfloor\alpha\rfloor-1,0\})^{\alpha-\lfloor\alpha\rfloor}\left[\frac{\Gamma(x+\lfloor\alpha\rfloor)}{\Gamma(x)}\right] \\
&= (\max\{x+\alpha-1,0\})^{\alpha-\lfloor\alpha\rfloor}\left[\prod_{i=1}^{\lfloor\alpha\rfloor}(x+\lfloor\alpha\rfloor-i)\right]\frac{\Gamma(x)}{\Gamma(x)} \\
&\geq (\max\{x+\alpha-1,0\})^{\alpha-\lfloor\alpha\rfloor}x^{\lfloor\alpha\rfloor} \\
&= (\max\{x+\alpha-1,0\})^{\alpha-\lfloor\alpha\rfloor}(\max\{x,0\})^{\lfloor\alpha\rfloor} \\
&\geq (\max\{x+\min\{\alpha-1,0\},0\})^{\alpha-\lfloor\alpha\rfloor}(\max\{x+\min\{\alpha-1,0\},0\})^{\lfloor\alpha\rfloor} \\
&= (\max\{x+\min\{\alpha-1,0\},0\})^{\alpha}.
\end{aligned} \tag{5.110}$$

Combining this with (5.109) shows (iv). The proof of Proposition 5.31 is thus complete. □

Corollary 5.32. *Let $\mathbb{B}\colon (0,\infty)^2 \to (0,\infty)$ satisfy for all $x,y \in (0,\infty)$ that $\mathbb{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t$ and let $\Gamma\colon (0,\infty) \to (0,\infty)$ satisfy for all $x \in (0,\infty)$ that $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,\mathrm{d}t$. Then it holds for all $x,y \in (0,\infty)$ with $x+y>1$ that*

$$\frac{\Gamma(x)}{(y+\max\{x-1,0\})^x} \leq \mathbb{B}(x,y) \leq \frac{\Gamma(x)}{(y+\min\{x-1,0\})^x} \leq \frac{\max\{1,x^x\}}{x(y+\min\{x-1,0\})^x}. \tag{5.111}$$

Proof of Corollary 5.32. Note that (iii) in Lemma 5.29 ensures for all $x,y \in (0,\infty)$ that

$$\mathbb{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(y+x)}. \tag{5.112}$$

In addition, observe that it holds for all $x,y \in (0,\infty)$ with $x+y>1$ that $y+\min\{x-1,0\} > 0$. This and (iv) in Proposition 5.31 demonstrate for all $x,y \in (0,\infty)$ with $x+y>1$ that

$$0 < (y+\min\{x-1,0\})^x \leq \frac{\Gamma(y+x)}{\Gamma(y)} \leq (y+\max\{x-1,0\})^x. \tag{5.113}$$

Combining this with (5.112) and (ii) in Proposition 5.31 shows for all $x,y \in (0,\infty)$ with $x+y>1$ that

$$\frac{\Gamma(x)}{(y+\max\{x-1,0\})^x} \leq \mathbb{B}(x,y) \leq \frac{\Gamma(x)}{(y+\min\{x-1,0\})^x} \leq \frac{\max\{1,x^x\}}{x(y+\min\{x-1,0\})^x}. \tag{5.114}$$

The proof of Corollary 5.32 is thus complete. □

5.4.2 Strong convergence rates for the optimisation error

Lemma 5.33. *Let $K \in \mathbb{N}$, $p, L \in (0,\infty)$, let (E,δ) be a metric space, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $\mathcal{R}\colon E\times\Omega \to \mathbb{R}$ be a $(\mathcal{B}(E)\otimes\mathcal{F})/\mathcal{B}(\mathbb{R})$-measurable function, assume for all $x,y \in E$, $\omega \in \Omega$ that $|\mathcal{R}(x,\omega)-\mathcal{R}(y,\omega)| \leq L\delta(x,y)$, and let $X_k\colon \Omega \to E$, $k \in \{1,2,\ldots,K\}$, be i.i.d. random variables. Then it holds for all $x \in E$ that*

$$\mathbb{E}\big[\min\nolimits_{k\in\{1,2,\ldots,K\}}|\mathcal{R}(X_k)-\mathcal{R}(x)|^p\big] \leq L^p\int_0^\infty [\mathbb{P}(\delta(X_1,x) > \varepsilon^{1/p})]^K\,\mathrm{d}\varepsilon. \tag{5.115}$$

Proof of Lemma 5.33. Throughout this proof let $x \in E$ and let $Y\colon \Omega \to [0,\infty)$ be the function which satisfies for all $\omega \in \Omega$ that $Y(\omega) = \min_{k\in\{1,2,\ldots,K\}}[\delta(X_k(\omega), x)]^p$. Observe that the fact that Y is a random variable, the assumption that $\forall\, v, w \in E, \omega \in \Omega\colon |\mathcal{R}(v,\omega) - \mathcal{R}(w,\omega)| \leq L\delta(v,w)$, and the complementary distribution function formula (see, e.g., Elbrächter et al. [120, Lemma 2.2]) demonstrate that

$$\begin{aligned}
&\mathbb{E}\big[\min_{k\in\{1,2,\ldots,K\}}|\mathcal{R}(X_k) - \mathcal{R}(x)|^p\big] \leq L^p\, \mathbb{E}\big[\min_{k\in\{1,2,\ldots,K\}}[\delta(X_k, x)]^p\big] \\
&= L^p\, \mathbb{E}[Y] = L^p \int_0^\infty y\, \mathbb{P}_Y(\mathrm{d}y) = L^p \int_0^\infty \mathbb{P}_Y((\varepsilon,\infty))\, \mathrm{d}\varepsilon \\
&= L^p \int_0^\infty \mathbb{P}(Y > \varepsilon)\, \mathrm{d}\varepsilon = L^p \int_0^\infty \mathbb{P}\big(\min_{k\in\{1,2,\ldots,K\}}[\delta(X_k, x)]^p > \varepsilon\big)\, \mathrm{d}\varepsilon.
\end{aligned} \tag{5.116}$$

Moreover, the assumption that Θ_k, $k \in \{1, 2, \ldots, K\}$, are i.i.d. random variables shows for all $\varepsilon \in (0,\infty)$ that

$$\begin{aligned}
&\mathbb{P}\big(\min_{k\in\{1,2,\ldots,K\}}[\delta(X_k, x)]^p > \varepsilon\big) = \mathbb{P}\big(\forall\, k \in \{1,2,\ldots,K\}\colon [\delta(X_k,x)]^p > \varepsilon\big) \\
&= \prod_{k=1}^K \mathbb{P}([\delta(X_k, x)]^p > \varepsilon) = [\mathbb{P}([\delta(X_1,x)]^p > \varepsilon)]^K = [\mathbb{P}(\delta(X_1,x) > \varepsilon^{1/p})]^K.
\end{aligned} \tag{5.117}$$

Combining (5.116) with (5.117) proves (5.115). The proof of Lemma 5.33 is thus complete. ∎

Proposition 5.34. *Let $\mathbf{d}, K \in \mathbb{N}$, $L, \alpha \in \mathbb{R}$, $\beta \in (\alpha,\infty)$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{R}\colon [\alpha,\beta]^{\mathbf{d}} \times \Omega \to \mathbb{R}$ be a random field, assume for all $\theta, \vartheta \in [\alpha,\beta]^{\mathbf{d}}$, $\omega \in \Omega$ that $|\mathcal{R}(\theta,\omega) - \mathcal{R}(\vartheta,\omega)| \leq L\|\theta - \vartheta\|_\infty$, let $\Theta_k\colon \Omega \to [\alpha,\beta]^{\mathbf{d}}$, $k \in \{1,2,\ldots,K\}$, be i.i.d. random variables, and assume that Θ_1 is continuous uniformly distributed on $[\alpha,\beta]^{\mathbf{d}}$ (cf. Definition 5.9). Then*

(i) it holds that $\mathcal{R}$ is a $(\mathcal{B}([\alpha,\beta]^{\mathbf{d}}) \otimes \mathcal{F})/\mathcal{B}(\mathbb{R})$-measurable function and

(ii) it holds for all $\theta \in [\alpha,\beta]^{\mathbf{d}}$, $p \in (0,\infty)$ that

$$\begin{aligned}
&\big(\mathbb{E}\big[\min_{k\in\{1,2,\ldots,K\}}|\mathcal{R}(\Theta_k) - \mathcal{R}(\theta)|^p\big]\big)^{1/p} \\
&\leq \frac{L(\beta-\alpha)\max\{1, (p/\mathbf{d})^{1/\mathbf{d}}\}}{K^{1/\mathbf{d}}} \leq \frac{L(\beta-\alpha)\max\{1,p\}}{K^{1/\mathbf{d}}}.
\end{aligned} \tag{5.118}$$

Proof of Proposition 5.34. Throughout this proof assume w.l.o.g. that $L > 0$, let $\delta\colon ([\alpha,\beta]^{\mathbf{d}}) \times ([\alpha,\beta]^{\mathbf{d}}) \to [0,\infty)$ satisfy for all $\theta, \vartheta \in [\alpha,\beta]^{\mathbf{d}}$ that $\delta(\theta,\vartheta) = \|\theta - \vartheta\|_\infty$, let $\mathbb{B}\colon (0,\infty)^2 \to (0,\infty)$ satisfy for all $x, y \in (0,\infty)$ that $\mathbb{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t$, and let $\Theta_{1,1}, \Theta_{1,2}, \ldots, \Theta_{1,\mathbf{d}}\colon \Omega \to [\alpha,\beta]$ satisfy $\Theta_1 = (\Theta_{1,1}, \Theta_{1,2}, \ldots, \Theta_{1,\mathbf{d}})$. First of all, note that the assumption that $\forall\, \theta, \vartheta \in [\alpha,\beta]^{\mathbf{d}}, \omega \in \Omega\colon |\mathcal{R}(\theta,\omega) - \mathcal{R}(\vartheta,\omega)| \leq L\|\theta - \vartheta\|_\infty$ ensures for all $\omega \in \Omega$ that the function $[\alpha,\beta]^{\mathbf{d}} \ni \theta \mapsto \mathcal{R}(\theta,\omega) \in \mathbb{R}$ is continuous. Combining this with the fact that $([\alpha,\beta]^{\mathbf{d}}, \delta)$ is a separable metric space, the fact that for every $\theta \in [\alpha,\beta]^{\mathbf{d}}$ it holds that the function $\Omega \ni \omega \mapsto \mathcal{R}(\theta,\omega) \in \mathbb{R}$ is $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable, and, e.g., Aliprantis & Border [4, Lemma 4.51] (see also, e.g., Beck et al. [22, Lemma 2.4]) proves (i). Next observe that it holds for all $\theta \in [\alpha, \beta]$, $\varepsilon \in [0, \infty)$ that

$$\begin{aligned} &\min\{\theta+\varepsilon, \beta\} - \max\{\theta-\varepsilon, \alpha\} = \min\{\theta+\varepsilon, \beta\} + \min\{\varepsilon-\theta, -\alpha\} \\ &= \min\bigl\{\theta+\varepsilon+\min\{\varepsilon-\theta, -\alpha\}, \beta+\min\{\varepsilon-\theta, -\alpha\}\bigr\} \\ &= \min\bigl\{\min\{2\varepsilon, \theta-\alpha+\varepsilon\}, \min\{\beta-\theta+\varepsilon, \beta-\alpha\}\bigr\} \\ &\geq \min\bigl\{\min\{2\varepsilon, \alpha-\alpha+\varepsilon\}, \min\{\beta-\beta+\varepsilon, \beta-\alpha\}\bigr\} \\ &= \min\{2\varepsilon, \varepsilon, \varepsilon, \beta-\alpha\} = \min\{\varepsilon, \beta-\alpha\}. \end{aligned} \tag{5.119}$$

The assumption that Θ_1 is continuous uniformly distributed on $[\alpha, \beta]^{\mathbf{d}}$ hence shows for all $\theta = (\theta_1, \theta_2, \ldots, \theta_{\mathbf{d}}) \in [\alpha, \beta]^{\mathbf{d}}$, $\varepsilon \in [0, \infty)$ that

$$\begin{aligned} &\mathbb{P}(\|\Theta_1 - \theta\|_\infty \leq \varepsilon) = \mathbb{P}\bigl(\max_{i \in \{1,2,\ldots,\mathbf{d}\}} |\Theta_{1,i} - \theta_i| \leq \varepsilon\bigr) \\ &= \mathbb{P}\bigl(\forall\, i \in \{1,2,\ldots,\mathbf{d}\} \colon -\varepsilon \leq \Theta_{1,i} - \theta_i \leq \varepsilon\bigr) \\ &= \mathbb{P}\bigl(\forall\, i \in \{1,2,\ldots,\mathbf{d}\} \colon \theta_i - \varepsilon \leq \Theta_{1,i} \leq \theta_i + \varepsilon\bigr) \\ &= \mathbb{P}\bigl(\forall\, i \in \{1,2,\ldots,\mathbf{d}\} \colon \max\{\theta_i - \varepsilon, \alpha\} \leq \Theta_{1,i} \leq \min\{\theta_i + \varepsilon, \beta\}\bigr) \\ &= \mathbb{P}\bigl(\Theta_1 \in \bigl[\times_{i=1}^{\mathbf{d}} [\max\{\theta_i - \varepsilon, \alpha\}, \min\{\theta_i + \varepsilon, \beta\}]\bigr]\bigr) \\ &= \tfrac{1}{(\beta-\alpha)^{\mathbf{d}}} \prod_{i=1}^{\mathbf{d}} (\min\{\theta_i + \varepsilon, \beta\} - \max\{\theta_i - \varepsilon, \alpha\}) \\ &\geq \tfrac{1}{(\beta-\alpha)^{\mathbf{d}}} [\min\{\varepsilon, \beta-\alpha\}]^{\mathbf{d}} = \min\Bigl\{1, \tfrac{\varepsilon^{\mathbf{d}}}{(\beta-\alpha)^{\mathbf{d}}}\Bigr\}. \end{aligned} \tag{5.120}$$

Therefore, we obtain for all $\theta \in [\alpha, \beta]^{\mathbf{d}}$, $p \in (0, \infty)$, $\varepsilon \in [0, \infty)$ that

$$\begin{aligned} \mathbb{P}(\|\Theta_1 - \theta\|_\infty > \varepsilon^{1/p}) &= 1 - \mathbb{P}(\|\Theta_1 - \theta\|_\infty \leq \varepsilon^{1/p}) \\ &\leq 1 - \min\Bigl\{1, \tfrac{\varepsilon^{\mathbf{d}/p}}{(\beta-\alpha)^{\mathbf{d}}}\Bigr\} = \max\Bigl\{0, 1 - \tfrac{\varepsilon^{\mathbf{d}/p}}{(\beta-\alpha)^{\mathbf{d}}}\Bigr\}. \end{aligned} \tag{5.121}$$

This, (i), the assumption that $\forall\, \theta, \vartheta \in [\alpha, \beta]^{\mathbf{d}}$, $\omega \in \Omega \colon |\mathcal{R}(\theta, \omega) - \mathcal{R}(\vartheta, \omega)| \leq L\|\theta - \vartheta\|_\infty$, the assumption that Θ_k, $k \in \{1, 2, \ldots, K\}$, are i.i.d. random variables, and Lemma 5.33 (with $(E, \delta) \leftarrow ([\alpha, \beta]^{\mathbf{d}}, \delta)$, $(X_k)_{k \in \{1,2,\ldots,K\}} \leftarrow (\Theta_k)_{k \in \{1,2,\ldots,K\}}$ in the notation of Lemma 5.33) establish for all $\theta \in [\alpha, \beta]^{\mathbf{d}}$, $p \in (0, \infty)$ that

$$\begin{aligned} &\mathbb{E}\bigl[\min_{k \in \{1,2,\ldots,K\}} |\mathcal{R}(\Theta_k) - \mathcal{R}(\theta)|^p\bigr] \leq L^p \int_0^\infty [\mathbb{P}(\|\Theta_1 - \theta\|_\infty > \varepsilon^{1/p})]^K \, d\varepsilon \\ &\leq L^p \int_0^\infty \Bigl[\max\Bigl\{0, 1 - \tfrac{\varepsilon^{\mathbf{d}/p}}{(\beta-\alpha)^{\mathbf{d}}}\Bigr\}\Bigr]^K d\varepsilon = L^p \int_0^{(\beta-\alpha)^p} \Bigl(1 - \tfrac{\varepsilon^{\mathbf{d}/p}}{(\beta-\alpha)^{\mathbf{d}}}\Bigr)^K d\varepsilon \\ &= \tfrac{p}{\mathbf{d}} L^p (\beta-\alpha)^p \int_0^1 t^{p/\mathbf{d}-1} (1-t)^K \, dt = \tfrac{p}{\mathbf{d}} L^p (\beta-\alpha)^p \int_0^1 t^{p/\mathbf{d}-1} (1-t)^{K+1-1} \, dt \\ &= \tfrac{p}{\mathbf{d}} L^p (\beta-\alpha)^p \, \mathbb{B}(p/\mathbf{d}, K+1). \end{aligned} \tag{5.122}$$

Corollary 5.32 (with $x \leftarrow p/\mathbf{d}$, $y \leftarrow K+1$ for $p \in (0, \infty)$ in the notation of (5.111) in Corollary 5.32) hence demonstrates for all $\theta \in [\alpha, \beta]^{\mathbf{d}}$, $p \in (0, \infty)$ that

$$\begin{aligned} &\mathbb{E}\bigl[\min_{k \in \{1,2,\ldots,K\}} |\mathcal{R}(\Theta_k) - \mathcal{R}(\theta)|^p\bigr] \\ &\leq \frac{\frac{p}{\mathbf{d}} L^p (\beta-\alpha)^p \max\{1, (p/\mathbf{d})^{p/\mathbf{d}}\}}{\frac{p}{\mathbf{d}} (K + 1 + \min\{p/\mathbf{d} - 1, 0\})^{p/\mathbf{d}}} \leq \frac{L^p (\beta-\alpha)^p \max\{1, (p/\mathbf{d})^{p/\mathbf{d}}\}}{K^{p/\mathbf{d}}}. \end{aligned} \tag{5.123}$$

This implies for all $\theta \in [\alpha, \beta]^{\mathbf{d}}$, $p \in (0, \infty)$ that

$$\begin{aligned}&\big(\mathbb{E}\big[\min_{k \in \{1,2,\ldots,K\}} |\mathcal{R}(\Theta_k) - \mathcal{R}(\theta)|^p\big]\big)^{1/p} \\ &\leq \frac{L(\beta-\alpha)\max\{1, (p/\mathbf{d})^{1/\mathbf{d}}\}}{K^{1/\mathbf{d}}} \leq \frac{L(\beta-\alpha)\max\{1,p\}}{K^{1/\mathbf{d}}}.\end{aligned} \tag{5.124}$$

This shows (ii) and thus completes the proof of Proposition 5.34. □

Lemma 5.35. *Let $d, \mathbf{d}, \mathbf{L}, M \in \mathbb{N}$, $B, b \in [1, \infty)$, $u \in \mathbb{R}$, $v \in (u, \infty)$, $\mathbf{l} = (\mathbf{l}_0, \mathbf{l}_1, \ldots, \mathbf{l}_{\mathbf{L}}) \in \mathbb{N}^{\mathbf{L}+1}$, $D \subseteq [-b, b]^d$, assume $\mathbf{l}_0 = d$, $\mathbf{l}_{\mathbf{L}} = 1$, and $\mathbf{d} \geq \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1}+1)$, let Ω be a set, let $X_j \colon \Omega \to D$, $j \in \{1, 2, \ldots, M\}$, and $Y_j \colon \Omega \to [u, v]$, $j \in \{1, 2, \ldots, M\}$, be functions, and let $\mathcal{R} \colon [-B, B]^{\mathbf{d}} \times \Omega \to [0, \infty)$ satisfy for all $\theta \in [-B, B]^{\mathbf{d}}$, $\omega \in \Omega$ that*

$$\mathcal{R}(\theta, \omega) = \frac{1}{M}\left[\sum_{j=1}^{M} |\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(X_j(\omega)) - Y_j(\omega)|^2\right] \tag{5.125}$$

(cf. Definition 5.8). Then it holds for all $\theta, \vartheta \in [-B, B]^{\mathbf{d}}$, $\omega \in \Omega$ that

$$|\mathcal{R}(\theta, \omega) - \mathcal{R}(\vartheta, \omega)| \leq 2(v-u)b\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} B^{\mathbf{L}-1} \|\theta - \vartheta\|_\infty \tag{5.126}$$

(cf. Definition 5.9).

Proof of Lemma 5.35. Observe that the fact that $\forall\, x_1, x_2, y \in \mathbb{R} \colon (x_1-y)^2 - (x_2-y)^2 = (x_1-x_2)((x_1-y)+(x_2-y))$, the fact that $\forall\, \theta \in \mathbb{R}^{\mathbf{d}}$, $x \in \mathbb{R}^d \colon \mathscr{N}_{u,v}^{\theta,\mathbf{l}}(x) \in [u, v]$, and the assumption that $\forall\, j \in \{1, 2, \ldots, M\}$, $\omega \in \Omega \colon Y_j(\omega) \in [u, v]$ prove for all $\theta, \vartheta \in [-B, B]^{\mathbf{d}}$, $\omega \in \Omega$ that

$$\begin{aligned}&|\mathcal{R}(\theta, \omega) - \mathcal{R}(\vartheta, \omega)| \\ &= \frac{1}{M}\left|\left[\sum_{j=1}^{M} |\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(X_j(\omega)) - Y_j(\omega)|^2\right] - \left[\sum_{j=1}^{M} |\mathscr{N}_{u,v}^{\vartheta,\mathbf{l}}(X_j(\omega)) - Y_j(\omega)|^2\right]\right| \\ &\leq \frac{1}{M}\left[\sum_{j=1}^{M} \big|[\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(X_j(\omega)) - Y_j(\omega)]^2 - [\mathscr{N}_{u,v}^{\vartheta,\mathbf{l}}(X_j(\omega)) - Y_j(\omega)]^2\big|\right] \\ &= \frac{1}{M}\left[\sum_{j=1}^{M} \Big(\big|\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(X_j(\omega)) - \mathscr{N}_{u,v}^{\vartheta,\mathbf{l}}(X_j(\omega))\big| \right. \\ &\qquad \left. \cdot \big|[\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(X_j(\omega)) - Y_j(\omega)] + [\mathscr{N}_{u,v}^{\vartheta,\mathbf{l}}(X_j(\omega)) - Y_j(\omega)]\big|\Big)\right] \\ &\leq \frac{2}{M}\left[\sum_{j=1}^{M} \Big(\big[\sup\nolimits_{x \in D} |\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(x) - \mathscr{N}_{u,v}^{\vartheta,\mathbf{l}}(x)|\big]\big[\sup\nolimits_{y_1,y_2 \in [u,v]} |y_1 - y_2|\big]\Big)\right] \\ &= 2(v-u)\big[\sup\nolimits_{x \in D} |\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(x) - \mathscr{N}_{u,v}^{\vartheta,\mathbf{l}}(x)|\big].\end{aligned} \tag{5.127}$$

In addition, combining the assumptions that $D \subseteq [-b, b]^d$, $\mathbf{d} \geq \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1}+1)$, $\mathbf{l}_0 = d$, $\mathbf{l}_{\mathbf{L}} = 1$, $b \geq 1$, and $B \geq 1$ with Beck, Jentzen, & Kuckuck [27, Corollary 2.37] (with $a \leftarrow -b$, $b \leftarrow b$, $u \leftarrow u$, $v \leftarrow v$, $d \leftarrow \mathbf{d}$, $L \leftarrow \mathbf{L}$, $l \leftarrow \mathbf{l}$ in the notation of [27, Corollary 2.37]) shows for all $\theta, \vartheta \in [-B, B]^{\mathbf{d}}$ that

$$\begin{aligned}&\sup\nolimits_{x \in D} |\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(x) - \mathscr{N}_{u,v}^{\vartheta,\mathbf{l}}(x)| \leq \sup\nolimits_{x \in [-b,b]^d} |\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(x) - \mathscr{N}_{u,v}^{\vartheta,\mathbf{l}}(x)| \\ &\leq \mathbf{L}\max\{1, b\}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} (\max\{1, \|\theta\|_\infty, \|\vartheta\|_\infty\})^{\mathbf{L}-1} \|\theta - \vartheta\|_\infty \\ &\leq b\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} B^{\mathbf{L}-1} \|\theta - \vartheta\|_\infty.\end{aligned} \tag{5.128}$$

This and (5.127) imply for all $\theta, \vartheta \in [-B, B]^{\mathbf{d}}$, $\omega \in \Omega$ that

$$|\mathcal{R}(\theta, \omega) - \mathcal{R}(\vartheta, \omega)| \leq 2(v-u)b\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} B^{\mathbf{L}-1} \|\theta - \vartheta\|_\infty. \tag{5.129}$$

The proof of Lemma 5.35 is thus complete. □

Corollary 5.36. *Let $d, \mathbf{d}, \mathfrak{d}, \mathbf{L}, M, K \in \mathbb{N}$, $B, b \in [1, \infty)$, $u \in \mathbb{R}$, $v \in (u, \infty)$, $\mathbf{l} = (\mathbf{l}_0, \mathbf{l}_1, \ldots, \mathbf{l}_\mathbf{L}) \in \mathbb{N}^{\mathbf{L}+1}$, $D \subseteq [-b, b]^d$, assume $\mathbf{l}_0 = d$, $\mathbf{l}_\mathbf{L} = 1$, and $\mathbf{d} \geq \mathfrak{d} = \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1}+1)$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\Theta_k \colon \Omega \to [-B, B]^{\mathbf{d}}$, $k \in \{1, 2, \ldots, K\}$, be i.i.d. random variables, assume that Θ_1 is continuous uniformly distributed on $[-B, B]^{\mathbf{d}}$, let $X_j \colon \Omega \to D$, $j \in \{1, 2, \ldots, M\}$, and $Y_j \colon \Omega \to [u, v]$, $j \in \{1, 2, \ldots, M\}$, be random variables, and let $\mathcal{R} \colon [-B, B]^{\mathbf{d}} \times \Omega \to [0, \infty)$ satisfy for all $\theta \in [-B, B]^{\mathbf{d}}$, $\omega \in \Omega$ that*

$$\mathcal{R}(\theta, \omega) = \frac{1}{M}\left[\sum_{j=1}^{M} |\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(X_j(\omega)) - Y_j(\omega)|^2\right] \tag{5.130}$$

(cf. Definition 5.8). Then

(i) it holds that $\mathcal{R}$ is a $(\mathcal{B}([-B, B]^{\mathbf{d}}) \otimes \mathcal{F})/\mathcal{B}([0, \infty))$-measurable function and

(ii) it holds for all $\theta \in [-B, B]^{\mathbf{d}}$, $p \in (0, \infty)$ that

$$\begin{aligned}
&\left(\mathbb{E}\left[\min\nolimits_{k \in \{1,2,\ldots,K\}} |\mathcal{R}(\Theta_k) - \mathcal{R}(\theta)|^p\right]\right)^{1/p} \\
&\leq \frac{4(v-u)b\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} B^{\mathbf{L}} \sqrt{\max\{1, p/\mathfrak{d}\}}}{K^{1/\mathfrak{d}}} \leq \frac{4(v-u)b\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} B^{\mathbf{L}} \max\{1, p\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}}
\end{aligned} \tag{5.131}$$

(cf. Definition 5.9).

Proof of Corollary 5.36. Throughout this proof let $L \in \mathbb{R}$ be given by $L = 2(v-u)[b\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} B^{\mathbf{L}-1}]$, let $P \colon [-B, B]^{\mathbf{d}} \to [-B, B]^{\mathfrak{d}}$ satisfy for all $\theta = (\theta_1, \theta_2, \ldots, \theta_{\mathbf{d}}) \in [-B, B]^{\mathbf{d}}$ that $P(\theta) = (\theta_1, \theta_2, \ldots, \theta_{\mathfrak{d}})$, and let $R \colon [-B, B]^{\mathfrak{d}} \times \Omega \to \mathbb{R}$ satisfy for all $\theta \in [-B, B]^{\mathfrak{d}}$, $\omega \in \Omega$ that

$$R(\theta, \omega) = \frac{1}{M}\left[\sum_{j=1}^{M} |\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(X_j(\omega)) - Y_j(\omega)|^2\right]. \tag{5.132}$$

Note that the fact that $\forall\, \theta \in [-B, B]^{\mathbf{d}} \colon \mathscr{N}_{u,v}^{\theta,\mathbf{l}} = \mathscr{N}_{u,v}^{P(\theta),\mathbf{l}}$ implies for all $\theta \in [-B, B]^{\mathbf{d}}$, $\omega \in \Omega$ that

$$\begin{aligned}
\mathcal{R}(\theta, \omega) &= \frac{1}{M}\left[\sum_{j=1}^{M} |\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(X_j(\omega)) - Y_j(\omega)|^2\right] \\
&= \frac{1}{M}\left[\sum_{j=1}^{M} |\mathscr{N}_{u,v}^{P(\theta),\mathbf{l}}(X_j(\omega)) - Y_j(\omega)|^2\right] = R(P(\theta), \omega).
\end{aligned} \tag{5.133}$$

Furthermore, Lemma 5.35 (with $\mathbf{d} \leftarrow \mathfrak{d}$, $\mathcal{R} \leftarrow ([-B, B]^{\mathfrak{d}} \times \Omega \ni (\theta, \omega) \mapsto R(\theta, \omega) \in [0, \infty))$ in the notation of Lemma 5.35) demonstrates for all $\theta, \vartheta \in [-B, B]^{\mathfrak{d}}$, $\omega \in \Omega$ that

$$|R(\theta, \omega) - R(\vartheta, \omega)| \leq 2(v-u)b\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} B^{\mathbf{L}-1} \|\theta - \vartheta\|_\infty = L\|\theta - \vartheta\|_\infty. \tag{5.134}$$

Moreover, observe that the assumption that X_j, $j \in \{1, 2, \ldots, M\}$, and Y_j, $j \in \{1, 2, \ldots, M\}$, are random variables ensures that $R \colon [-B, B]^{\mathfrak{d}} \times \Omega \to \mathbb{R}$ is a random field. This, (5.134), the fact that $P\circ\Theta_k\colon \Omega\to[-B,B]^{\mathfrak{d}}$, $k\in\{1,2,\dots,K\}$, are i.i.d. random variables, the fact that $P\circ\Theta_1$ is continuous uniformly distributed on $[-B,B]^{\mathfrak{d}}$, and Proposition 5.34 (with $\mathbf{d}\leftarrow\mathfrak{d}$, $\alpha\leftarrow -B$, $\beta\leftarrow B$, $\mathcal{R}\leftarrow R$, $(\Theta_k)_{k\in\{1,2,\dots,K\}}\leftarrow(P\circ\Theta_k)_{k\in\{1,2,\dots,K\}}$ in the notation of Proposition 5.34) prove for all $\theta\in[-B,B]^{\mathbf{d}}$, $p\in(0,\infty)$ that R is a $(\mathcal{B}([-B,B]^{\mathfrak{d}})\otimes\mathcal{F})/\mathcal{B}(\mathbb{R})$-measurable function and

$$\begin{aligned}&\big(\mathbb{E}\big[\min\nolimits_{k\in\{1,2,\dots,K\}}|R(P(\Theta_k))-R(P(\theta))|^p\big]\big)^{1/p}\\&\le\frac{L(2B)\max\{1,(p/\mathfrak{d})^{1/\mathfrak{d}}\}}{K^{1/\mathfrak{d}}}=\frac{4(v-u)b\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}B^{\mathbf{L}}\max\{1,(p/\mathfrak{d})^{1/\mathfrak{d}}\}}{K^{1/\mathfrak{d}}}.\end{aligned}\tag{5.135}$$

The fact that P is a $\mathcal{B}([-B,B]^{\mathbf{d}})/\mathcal{B}([-B,B]^{\mathfrak{d}})$-measurable function and (5.133) hence show (i). In addition, (5.133), (5.135), and the fact that $2\le\mathfrak{d}=\sum_{i=1}^{\mathbf{L}}\mathbf{l}_i(\mathbf{l}_{i-1}+1)\le\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2$ yield for all $\theta\in[-B,B]^{\mathbf{d}}$, $p\in(0,\infty)$ that

$$\begin{aligned}&\big(\mathbb{E}\big[\min\nolimits_{k\in\{1,2,\dots,K\}}|\mathcal{R}(\Theta_k)-\mathcal{R}(\theta)|^p\big]\big)^{1/p}\\&=\big(\mathbb{E}\big[\min\nolimits_{k\in\{1,2,\dots,K\}}|R(P(\Theta_k))-R(P(\theta))|^p\big]\big)^{1/p}\\&\le\frac{4(v-u)b\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}B^{\mathbf{L}}\sqrt{\max\{1,p/\mathfrak{d}\}}}{K^{1/\mathfrak{d}}}\le\frac{4(v-u)b\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}B^{\mathbf{L}}\max\{1,p\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}}.\end{aligned}\tag{5.136}$$

This establishes (ii). The proof of Corollary 5.36 is thus complete. □

5.5 Analysis of the overall error

In Subsection 5.5.2 below we present the main result of this chapter, Theorem 5.41, that provides an estimate for the overall L^2-error arising in deep learning based empirical risk minimisation with quadratic loss function in the probabilistically strong sense and that covers the case where the underlying DNNs are trained using a general stochastic optimisation algorithm with random initialisation.

In order to prove Theorem 5.41, we require a link to combine the results from Sections 5.2, 5.3, and 5.4, which is given in Subsection 5.5.1 below. More specifically, Proposition 5.37 in Subsection 5.5.1 shows that the overall error can be decomposed into three different error sources: the *approximation error* (cf. Section 5.2), the *worst-case generalisation error* (cf. Section 5.3), and the *optimisation error* (cf. Section 5.4). Proposition 5.37 is a consequence of the well-known bias–variance decomposition (cf., e.g., Beck, Jentzen, & Kuckuck [27, Lemma 4.1] or Berner, Grohs, & Jentzen [47, Lemma 2.2]) and is very similar to [27, Lemma 4.3].

Thereafter, Subsection 5.5.2 is devoted to strong convergence results for deep learning based empirical risk minimisation with quadratic loss function where a general stochastic approximation algorithm with random initialisation is allowed to be the employed optimisation method. Apart from the main result (cf. Theorem 5.41), Subsection 5.5.2 also includes, on the one hand, Proposition 5.39, which combines the overall error decomposition (cf. Proposition 5.37) with our convergence result for the generalisation error (cf. Corollary 5.28 in Section 5.3) and our convergence result for the optimisation error (cf. Corollary 5.36 in Section 5.4), and, on the other hand, Corollary 5.42, which replaces the architecture parameter $A\in(0,\infty)$ in Theorem 5.41 (cf. Proposition 5.13) by the minimum of the depth parameter $\mathbf{L}\in\mathbb{N}$ and the hidden layer sizes $\mathbf{l}_1,\mathbf{l}_2,\dots,\mathbf{l}_{\mathbf{L}-1}\in\mathbb{N}$ of the trained DNN (cf. (5.174) in the proof of Corollary 5.42).

Finally, in Subsection 5.5.3 we present three more strong convergence results for the special case where SGD with random initialisation is the employed optimisation method. In particular, Corollary 5.43 specifies Corollary 5.42 to this special case, Corollary 5.44 provides a convergence estimate for the expectation of the L^1-distance between the trained DNN and the target function, and Corollary 5.45 reaches an analogous conclusion in a simplified setting.

5.5.1 Overall error decomposition

Proposition 5.37. *Let $d, \mathbf{d}, \mathbf{L}, M, K, N \in \mathbb{N}$, $B \in [0,\infty)$, $u \in \mathbb{R}$, $v \in (u,\infty)$, $\mathbf{l} = (\mathbf{l}_0, \mathbf{l}_1, \ldots, \mathbf{l}_\mathbf{L}) \in \mathbb{N}^{\mathbf{L}+1}$, $\mathbf{N} \subseteq \{0,1,\ldots,N\}$, $D \subseteq \mathbb{R}^d$, assume $0 \in \mathbf{N}$, $\mathbf{l}_0 = d$, $\mathbf{l}_\mathbf{L} = 1$, and $\mathbf{d} \geq \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1}+1)$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X_j \colon \Omega \to D$, $j \in \{1,2,\ldots,M\}$, and $Y_j \colon \Omega \to [u,v]$, $j \in \{1,2,\ldots,M\}$, be random variables, let $\mathcal{E} \colon D \to [u,v]$ be a $\mathcal{B}(D)/\mathcal{B}([u,v])$-measurable function, assume that it holds $\mathbb{P}$-a.s. that $\mathcal{E}(X_1) = \mathbb{E}[Y_1|X_1]$, let $\Theta_{k,n} \colon \Omega \to \mathbb{R}^\mathbf{d}$, $k,n \in \mathbb{N}_0$, satisfy $\big(\bigcup_{k=1}^\infty \Theta_{k,0}(\Omega)\big) \subseteq [-B,B]^\mathbf{d}$, let $\mathbf{R} \colon \mathbb{R}^\mathbf{d} \to [0,\infty)$ satisfy for all $\theta \in \mathbb{R}^\mathbf{d}$ that $\mathbf{R}(\theta) = \mathbb{E}[|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_1) - Y_1|^2]$, and let $\mathcal{R} \colon \mathbb{R}^\mathbf{d} \times \Omega \to [0,\infty)$ and $\mathbf{k} \colon \Omega \to (\mathbb{N}_0)^2$ satisfy for all $\theta \in \mathbb{R}^\mathbf{d}$, $\omega \in \Omega$ that*

$$\mathcal{R}(\theta,\omega) = \frac{1}{M}\left[\sum_{j=1}^{M} |\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_j(\omega)) - Y_j(\omega)|^2\right] \qquad \textit{and} \tag{5.137}$$

$$\mathbf{k}(\omega) \in \arg\min\nolimits_{(k,n)\in\{1,2,\ldots,K\}\times\mathbf{N},\, \|\Theta_{k,n}(\omega)\|_\infty \leq B} \mathcal{R}(\Theta_{k,n}(\omega),\omega) \tag{5.138}$$

(cf. Definitions 5.8 and 5.9). Then it holds for all $\vartheta \in [-B,B]^\mathbf{d}$ that

$$\begin{aligned} &\int_D |\mathscr{N}^{\Theta_\mathbf{k},\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x) \\ &\leq \big[\sup\nolimits_{x\in D} |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\big] + 2\big[\sup\nolimits_{\theta\in[-B,B]^\mathbf{d}} |\mathcal{R}(\theta) - \mathbf{R}(\theta)|\big] \\ &\quad + \min\nolimits_{(k,n)\in\{1,2,\ldots,K\}\times\mathbf{N},\, \|\Theta_{k,n}\|_\infty\leq B} |\mathcal{R}(\Theta_{k,n}) - \mathcal{R}(\vartheta)|. \end{aligned} \tag{5.139}$$

Proof of Proposition 5.37. Throughout this proof let $\mathscr{R} \colon \mathcal{L}^2(\mathbb{P}_{X_1};\mathbb{R}) \to [0,\infty)$ satisfy for all $f \in \mathcal{L}^2(\mathbb{P}_{X_1};\mathbb{R})$ that $\mathscr{R}(f) = \mathbb{E}[|f(X_1) - Y_1|^2]$. Observe that the assumption that $\forall\, \omega \in \Omega \colon Y_1(\omega) \in [u,v]$ and the fact that $\forall\, \theta \in \mathbb{R}^\mathbf{d}, x \in \mathbb{R}^d \colon \mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x) \in [u,v]$ ensure for all $\theta \in \mathbb{R}^\mathbf{d}$ that $\mathbb{E}[|Y_1|^2] \leq \max\{u^2,v^2\} < \infty$ and

$$\int_D |\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x) = \mathbb{E}\big[|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_1)|^2\big] \leq \max\{u^2,v^2\} < \infty. \tag{5.140}$$

The bias–variance decomposition (cf., e.g., Beck, Jentzen, & Kuckuck [27, (iii) in Lemma 4.1] with $(\Omega,\mathcal{F},\mathbb{P}) \leftarrow (\Omega,\mathcal{F},\mathbb{P})$, $(S,\mathcal{S}) \leftarrow (D,\mathcal{B}(D))$, $X \leftarrow X_1$, $Y \leftarrow (\Omega \ni \omega \mapsto Y_1(\omega) \in \mathbb{R})$, $\mathcal{E} \leftarrow \mathscr{R}$, $f \leftarrow \mathscr{N}^{\theta,\mathbf{l}}_{u,v}|_D$, $g \leftarrow \mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}|_D$ for $\theta,\vartheta \in \mathbb{R}^\mathbf{d}$ in the notation of [27, (iii) in Lemma 4.1]) hence proves for all $\theta,\vartheta \in \mathbb{R}^\mathbf{d}$ that

$$\begin{aligned} &\int_D |\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x) \\ &= \mathbb{E}\big[|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_1) - \mathcal{E}(X_1)|^2\big] = \mathbb{E}\big[|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_1) - \mathbb{E}[Y_1|X_1]|^2\big] \\ &= \mathbb{E}\big[|\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(X_1) - \mathbb{E}[Y_1|X_1]|^2\big] + \mathscr{R}(\mathscr{N}^{\theta,\mathbf{l}}_{u,v}|_D) - \mathscr{R}(\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}|_D) \\ &= \mathbb{E}\big[|\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(X_1) - \mathcal{E}(X_1)|^2\big] + \mathbb{E}\big[|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_1) - Y_1|^2\big] - \mathbb{E}\big[|\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(X_1) - Y_1|^2\big] \\ &= \int_D |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x) + \mathbf{R}(\theta) - \mathbf{R}(\vartheta). \end{aligned} \tag{5.141}$$

This implies for all $\theta, \vartheta \in \mathbb{R}^{\mathbf{d}}$ that

$$\begin{aligned}
&\int_D |\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x) \\
&= \int_D |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x) - [\mathcal{R}(\theta) - \mathbf{R}(\theta)] + \mathcal{R}(\vartheta) - \mathbf{R}(\vartheta) + \mathcal{R}(\theta) - \mathcal{R}(\vartheta) \\
&\leq \int_D |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x) + |\mathcal{R}(\theta) - \mathbf{R}(\theta)| + |\mathcal{R}(\vartheta) - \mathbf{R}(\vartheta)| + \mathcal{R}(\theta) - \mathcal{R}(\vartheta) \\
&\leq \int_D |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x) + 2\big[\max\nolimits_{\eta \in \{\theta,\vartheta\}} |\mathcal{R}(\eta) - \mathbf{R}(\eta)|\big] + \mathcal{R}(\theta) - \mathcal{R}(\vartheta).
\end{aligned} \tag{5.142}$$

Next note that the fact that $\forall\, \omega \in \Omega \colon \|\Theta_{\mathbf{k}(\omega)}(\omega)\|_\infty \leq B$ ensures for all $\omega \in \Omega$ that $\Theta_{\mathbf{k}(\omega)}(\omega) \in [-B, B]^{\mathbf{d}}$. Combining (5.142) with (5.138) hence establishes for all $\vartheta \in [-B, B]^{\mathbf{d}}$ that

$$\begin{aligned}
&\int_D |\mathscr{N}^{\Theta_{\mathbf{k}},\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x) \\
&\leq \int_D |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x) + 2\big[\sup\nolimits_{\theta \in [-B,B]^{\mathbf{d}}} |\mathcal{R}(\theta) - \mathbf{R}(\theta)|\big] + \mathcal{R}(\Theta_{\mathbf{k}}) - \mathcal{R}(\vartheta) \\
&= \int_D |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x) + 2\big[\sup\nolimits_{\theta \in [-B,B]^{\mathbf{d}}} |\mathcal{R}(\theta) - \mathbf{R}(\theta)|\big] \\
&\quad + \min\nolimits_{(k,n) \in \{1,2,\dots,K\} \times \mathbf{N},\, \|\Theta_{k,n}\|_\infty \leq B} [\mathcal{R}(\Theta_{k,n}) - \mathcal{R}(\vartheta)] \\
&\leq \big[\sup\nolimits_{x \in D} |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\big] + 2\big[\sup\nolimits_{\theta \in [-B,B]^{\mathbf{d}}} |\mathcal{R}(\theta) - \mathbf{R}(\theta)|\big] \\
&\quad + \min\nolimits_{(k,n) \in \{1,2,\dots,K\} \times \mathbf{N},\, \|\Theta_{k,n}\|_\infty \leq B} |\mathcal{R}(\Theta_{k,n}) - \mathcal{R}(\vartheta)|.
\end{aligned} \tag{5.143}$$

The proof of Proposition 5.37 is thus complete. □

5.5.2 Overall strong error analysis for the training of DNNs

Lemma 5.38. *Let $d, \mathbf{d}, \mathbf{L} \in \mathbb{N}$, $p \in [0, \infty)$, $u \in [-\infty, \infty)$, $v \in (u, \infty]$, $\mathbf{l} = (\mathbf{l}_0, \mathbf{l}_1, \dots, \mathbf{l}_{\mathbf{L}}) \in \mathbb{N}^{\mathbf{L}+1}$, $D \subseteq \mathbb{R}^d$, assume $\mathbf{l}_0 = d$, $\mathbf{l}_{\mathbf{L}} = 1$, and $\mathbf{d} \geq \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1} + 1)$, let $\mathcal{E} \colon D \to \mathbb{R}$ be a $\mathcal{B}(D)/\mathcal{B}(\mathbb{R})$-measurable function, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X \colon \Omega \to D$, $\mathbf{k} \colon \Omega \to (\mathbb{N}_0)^2$, and $\Theta_{k,n} \colon \Omega \to \mathbb{R}^{\mathbf{d}}$, $k, n \in \mathbb{N}_0$, be random variables. Then*

(i) it holds that the function $\mathbb{R}^{\mathbf{d}} \times \mathbb{R}^d \ni (\theta, x) \mapsto \mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x) \in \mathbb{R}$ is $(\mathcal{B}(\mathbb{R}^{\mathbf{d}}) \otimes \mathcal{B}(\mathbb{R}^d))/\mathcal{B}(\mathbb{R})$-measurable,

(ii) it holds that the function $\Omega \ni \omega \mapsto \Theta_{\mathbf{k}(\omega)}(\omega) \in \mathbb{R}^{\mathbf{d}}$ is $\mathcal{F}/\mathcal{B}(\mathbb{R}^{\mathbf{d}})$-measurable, and

(iii) it holds that the function

$$\Omega \ni \omega \mapsto \int_D |\mathscr{N}^{\Theta_{\mathbf{k}(\omega)}(\omega),\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^p \, \mathbb{P}_X(\mathrm{d}x) \in [0, \infty] \tag{5.144}$$

is $\mathcal{F}/\mathcal{B}([0, \infty])$-measurable

(cf. Definition 5.8).

Proof of Lemma 5.38. First, observe that Beck, Jentzen, & Kuckuck [27, Corollary 2.37] (with $a \leftarrow -\|x\|_\infty$, $b \leftarrow \|x\|_\infty$, $u \leftarrow u$, $v \leftarrow v$, $d \leftarrow \mathbf{d}$, $L \leftarrow \mathbf{L}$, $l \leftarrow \mathbf{l}$ for $x \in \mathbb{R}^d$ in the notation of [27, Corollary 2.37]) demonstrates for all $x \in \mathbb{R}^d$, $\theta, \vartheta \in \mathbb{R}^{\mathbf{d}}$ that

$$\begin{aligned} |\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x) - \mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x)| &\leq \sup\nolimits_{y\in[-\|x\|_\infty,\|x\|_\infty]^{\mathbf{l}_0}} |\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(y) - \mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(y)| \\ &\leq \mathbf{L}\max\{1,\|x\|_\infty\}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}}(\max\{1,\|\theta\|_\infty,\|\vartheta\|_\infty\})^{\mathbf{L}-1}\|\theta-\vartheta\|_\infty \end{aligned} \tag{5.145}$$

(cf. Definition 5.9). This implies for all $x \in \mathbb{R}^d$ that the function

$$\mathbb{R}^{\mathbf{d}} \ni \theta \mapsto \mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x) \in \mathbb{R} \tag{5.146}$$

is continuous. In addition, the fact that $\forall\, \theta \in \mathbb{R}^{\mathbf{d}}\colon \mathscr{N}^{\theta,\mathbf{l}}_{u,v} \in C(\mathbb{R}^d,\mathbb{R})$ ensures for all $\theta \in \mathbb{R}^{\mathbf{d}}$ that the function $\mathbb{R}^d \ni x \mapsto \mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x) \in \mathbb{R}$ is $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R})$-measurable. This, (5.146), the fact that $(\mathbb{R}^{\mathbf{d}}, \|\cdot\|_\infty|_{\mathbb{R}^{\mathbf{d}}})$ is a separable normed $\mathbb{R}$-vector space, and, e.g., Aliprantis & Border [4, Lemma 4.51] (see also, e.g., Beck et al. [22, Lemma 2.4]) show (i).

Second, we prove (ii). For this let $\Xi\colon \Omega \to \mathbb{R}^{\mathbf{d}}$ satisfy for all $\omega \in \Omega$ that $\Xi(\omega) = \Theta_{\mathbf{k}(\omega)}(\omega)$. Observe that the assumption that $\Theta_{k,n}\colon \Omega \to \mathbb{R}^{\mathbf{d}}$, $k,n \in \mathbb{N}_0$, and $\mathbf{k}\colon \Omega \to (\mathbb{N}_0)^2$ are random variables establishes for all $U \in \mathcal{B}(\mathbb{R}^{\mathbf{d}})$ that

$$\begin{aligned} \Xi^{-1}(U) &= \{\omega \in \Omega\colon \Xi(\omega) \in U\} = \{\omega \in \Omega\colon \Theta_{\mathbf{k}(\omega)}(\omega) \in U\} \\ &= \big\{\omega \in \Omega\colon \big[\exists\, k,n \in \mathbb{N}_0\colon ([\Theta_{k,n}(\omega) \in U] \wedge [\mathbf{k}(\omega) = (k,n)])\big]\big\} \\ &= \bigcup_{k=0}^\infty \bigcup_{n=0}^\infty \big(\{\omega \in \Omega\colon \Theta_{k,n}(\omega) \in U\} \cap \{\omega \in \Omega\colon \mathbf{k}(\omega) = (k,n)\}\big) \\ &= \bigcup_{k=0}^\infty \bigcup_{n=0}^\infty \big([(\Theta_{k,n})^{-1}(U)] \cap [\mathbf{k}^{-1}(\{(k,n)\})]\big) \in \mathcal{F}. \end{aligned} \tag{5.147}$$

This implies (ii).

Third, note that (i)–(ii) yield that the function $\Omega \times \mathbb{R}^d \ni (\omega, x) \mapsto \mathscr{N}^{\Theta_{\mathbf{k}(\omega)}(\omega),\mathbf{l}}_{u,v}(x) \in \mathbb{R}$ is $(\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^d))/\mathcal{B}(\mathbb{R})$-measurable. This and the assumption that $\mathcal{E}\colon D \to \mathbb{R}$ is $\mathcal{B}(D)/\mathcal{B}(\mathbb{R})$-measurable demonstrate that the function $\Omega \times D \ni (\omega,x) \mapsto |\mathscr{N}^{\Theta_{\mathbf{k}(\omega)}(\omega),\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^p \in [0,\infty)$ is $(\mathcal{F}\otimes\mathcal{B}(D))/\mathcal{B}([0,\infty))$-measurable. Tonelli's theorem hence establishes (iii). The proof of Lemma 5.38 is thus complete. □

Proposition 5.39. *Let $d, \mathbf{d}, \mathbf{L}, M, K, N \in \mathbb{N}$, $b, c \in [1,\infty)$, $B \in [c,\infty)$, $u \in \mathbb{R}$, $v \in (u,\infty)$, $\mathbf{l} = (\mathbf{l}_0, \mathbf{l}_1, \ldots, \mathbf{l}_{\mathbf{L}}) \in \mathbb{N}^{\mathbf{L}+1}$, $\mathbf{N} \subseteq \{0,1,\ldots,N\}$, $D \subseteq [-b,b]^d$, assume $0 \in \mathbf{N}$, $\mathbf{l}_0 = d$, $\mathbf{l}_{\mathbf{L}} = 1$, and $\mathbf{d} \geq \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1}+1)$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $X_j\colon \Omega \to D$, $j \in \mathbb{N}$, and $Y_j\colon \Omega \to [u,v]$, $j \in \mathbb{N}$, be functions, assume that (X_j, Y_j), $j \in \{1,2,\ldots,M\}$, are i.i.d. random variables, let $\mathcal{E}\colon D \to [u,v]$ be a $\mathcal{B}(D)/\mathcal{B}([u,v])$-measurable function, assume that it holds $\mathbb{P}$-a.s. that $\mathcal{E}(X_1) = \mathbb{E}[Y_1|X_1]$, let $\Theta_{k,n}\colon \Omega \to \mathbb{R}^{\mathbf{d}}$, $k,n \in \mathbb{N}_0$, and $\mathbf{k}\colon \Omega \to (\mathbb{N}_0)^2$ be random variables, assume $\big(\bigcup_{k=1}^\infty \Theta_{k,0}(\Omega)\big) \subseteq [-B,B]^{\mathbf{d}}$, assume that $\Theta_{k,0}$, $k \in \{1,2,\ldots,K\}$, are i.i.d., assume that $\Theta_{1,0}$ is continuous uniformly distributed on $[-c,c]^{\mathbf{d}}$, and let $\mathcal{R}\colon \mathbb{R}^{\mathbf{d}} \times \Omega \to [0,\infty)$ satisfy for all $\theta \in \mathbb{R}^{\mathbf{d}}$, $\omega \in \Omega$ that*

$$\mathcal{R}(\theta,\omega) = \frac{1}{M}\left[\sum_{j=1}^M |\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_j(\omega)) - Y_j(\omega)|^2\right] \qquad \textit{and} \tag{5.148}$$

$$\mathbf{k}(\omega) \in \operatorname{arg\,min}_{(k,n)\in\{1,2,\ldots,K\}\times\mathbf{N},\, \|\Theta_{k,n}(\omega)\|_\infty \leq B} \mathcal{R}(\Theta_{k,n}(\omega),\omega) \tag{5.149}$$

(cf. Definitions 5.8 and 5.9). Then it holds for all $p \in (0,\infty)$ that

$$\begin{aligned}
&\Big(\mathbb{E}\Big[\Big(\textstyle\int_D |\mathscr{N}^{\Theta_{\mathbf{k}},\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x)\Big)^p\Big]\Big)^{1/p} \\
&\le \big[\inf_{\theta\in[-c,c]^{\mathbf{d}}} \sup_{x\in D} |\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\big] + \frac{4(v-u)b\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}}\max\{1,p\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} \\
&\quad + \frac{18\max\{1,(v-u)^2\}\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{p,\ln(3MBb)\}}{\sqrt{M}} \\
&\le \big[\inf_{\theta\in[-c,c]^{\mathbf{d}}} \sup_{x\in D} |\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\big] \\
&\quad + \frac{20\max\{1,(v-u)^2\}b\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}+1}B^{\mathbf{L}}\max\{p,\ln(3M)\}}{\min\{\sqrt{M},K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}\}}
\end{aligned} \tag{5.150}$$

(cf. (iii) in Lemma 5.38).

Proof of Proposition 5.39. Throughout this proof let $\mathbf{R}\colon \mathbb{R}^{\mathbf{d}} \to [0,\infty)$ satisfy for all $\theta \in \mathbb{R}^{\mathbf{d}}$ that $\mathbf{R}(\theta) = \mathbb{E}[|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_1) - Y_1|^2]$. First of all, observe that the assumption that $\big(\bigcup_{k=1}^\infty \Theta_{k,0}(\Omega)\big) \subseteq [-B,B]^{\mathbf{d}}$, the assumption that $0 \in \mathbf{N}$, and Proposition 5.37 show for all $\vartheta \in [-B,B]^{\mathbf{d}}$ that

$$\begin{aligned}
&\int_D |\mathscr{N}^{\Theta_{\mathbf{k}},\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x) \\
&\le \big[\sup_{x\in D} |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\big] + 2\big[\sup_{\theta\in[-B,B]^{\mathbf{d}}} |\mathcal{R}(\theta) - \mathbf{R}(\theta)|\big] \\
&\quad + \min_{(k,n)\in\{1,2,\dots,K\}\times\mathbf{N},\, \|\Theta_{k,n}\|_\infty \le B} |\mathcal{R}(\Theta_{k,n}) - \mathcal{R}(\vartheta)| \\
&\le \big[\sup_{x\in D} |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\big] + 2\big[\sup_{\theta\in[-B,B]^{\mathbf{d}}} |\mathcal{R}(\theta) - \mathbf{R}(\theta)|\big] \\
&\quad + \min_{k\in\{1,2,\dots,K\},\, \|\Theta_{k,0}\|_\infty \le B} |\mathcal{R}(\Theta_{k,0}) - \mathcal{R}(\vartheta)| \\
&= \big[\sup_{x\in D} |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\big] + 2\big[\sup_{\theta\in[-B,B]^{\mathbf{d}}} |\mathcal{R}(\theta) - \mathbf{R}(\theta)|\big] \\
&\quad + \min_{k\in\{1,2,\dots,K\}} |\mathcal{R}(\Theta_{k,0}) - \mathcal{R}(\vartheta)|.
\end{aligned} \tag{5.151}$$

Minkowski's inequality hence establishes for all $p \in [1,\infty)$, $\vartheta \in [-c,c]^{\mathbf{d}} \subseteq [-B,B]^{\mathbf{d}}$ that

$$\begin{aligned}
&\Big(\mathbb{E}\Big[\Big(\textstyle\int_D |\mathscr{N}^{\Theta_{\mathbf{k}},\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x)\Big)^p\Big]\Big)^{1/p} \\
&\le \big(\mathbb{E}\big[\sup_{x\in D} |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^{2p}\big]\big)^{1/p} + 2\big(\mathbb{E}\big[\sup_{\theta\in[-B,B]^{\mathbf{d}}} |\mathcal{R}(\theta) - \mathbf{R}(\theta)|^p\big]\big)^{1/p} \\
&\quad + \big(\mathbb{E}\big[\min_{k\in\{1,2,\dots,K\}} |\mathcal{R}(\Theta_{k,0}) - \mathcal{R}(\vartheta)|^p\big]\big)^{1/p} \\
&\le \big[\sup_{x\in D} |\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\big] + 2\big(\mathbb{E}\big[\sup_{\theta\in[-B,B]^{\mathbf{d}}} |\mathcal{R}(\theta) - \mathbf{R}(\theta)|^p\big]\big)^{1/p} \\
&\quad + \sup_{\theta\in[-c,c]^{\mathbf{d}}} \big(\mathbb{E}\big[\min_{k\in\{1,2,\dots,K\}} |\mathcal{R}(\Theta_{k,0}) - \mathcal{R}(\theta)|^p\big]\big)^{1/p}
\end{aligned} \tag{5.152}$$

(cf. (i) in Corollary 5.28 and (i) in Corollary 5.36). Next note that Corollary 5.28 (with $v \leftarrow \max\{u+1,v\}$, $\mathbf{R} \leftarrow \mathbf{R}|_{[-B,B]^{\mathbf{d}}}$, $\mathcal{R} \leftarrow \mathcal{R}|_{[-B,B]^{\mathbf{d}}\times\Omega}$ in the notation of Corollary 5.28) proves for all $p \in (0,\infty)$ that

$$\begin{aligned}
&\big(\mathbb{E}\big[\sup_{\theta\in[-B,B]^{\mathbf{d}}} |\mathcal{R}(\theta) - \mathbf{R}(\theta)|^p\big]\big)^{1/p} \\
&\le \frac{9(\max\{u+1,v\}-u)^2\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{p,\ln(3MBb)\}}{\sqrt{M}} \\
&= \frac{9\max\{1,(v-u)^2\}\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{p,\ln(3MBb)\}}{\sqrt{M}}.
\end{aligned} \tag{5.153}$$

In addition, observe that Corollary 5.36 (with $\mathfrak{d} \leftarrow \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1}+1)$, $B \leftarrow c$, $(\Theta_k)_{k\in\{1,2,\dots,K\}} \leftarrow (\Omega \ni \omega \mapsto \mathbb{1}_{\{\Theta_{k,0}\in[-c,c]^{\mathbf{d}}\}}(\omega)\Theta_{k,0}(\omega) \in [-c,c]^{\mathbf{d}})_{k\in\{1,2,\dots,K\}}$, $\mathcal{R} \leftarrow \mathcal{R}|_{[-c,c]^{\mathbf{d}}\times\Omega}$ in the notation of Corollary 5.36) implies for all $p \in (0,\infty)$ that

$$\begin{aligned}
&\sup\nolimits_{\theta\in[-c,c]^{\mathbf{d}}}\big(\mathbb{E}\big[\min\nolimits_{k\in\{1,2,\dots,K\}}|\mathcal{R}(\Theta_{k,0}) - \mathcal{R}(\theta)|^p\big]\big)^{1/p} \\
&= \sup\nolimits_{\theta\in[-c,c]^{\mathbf{d}}}\big(\mathbb{E}\big[\min\nolimits_{k\in\{1,2,\dots,K\}}|\mathcal{R}(\mathbb{1}_{\{\Theta_{k,0}\in[-c,c]^{\mathbf{d}}\}}\Theta_{k,0}) - \mathcal{R}(\theta)|^p\big]\big)^{1/p} \\
&\leq \frac{4(v-u)b\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}}\max\{1,p\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}}.
\end{aligned} \tag{5.154}$$

Combining this, (5.152), (5.153), and the fact that $\ln(3MBb) \geq 1$ with Jensen's inequality demonstrates for all $p \in (0,\infty)$ that

$$\begin{aligned}
&\Big(\mathbb{E}\Big[\Big(\int_D|\mathscr{N}_{u,v}^{\Theta_{\mathbf{k}},\mathbf{l}}(x) - \mathcal{E}(x)|^2\,\mathbb{P}_{X_1}(\mathrm{d}x)\Big)^p\Big]\Big)^{1/p} \\
&\leq \left(\mathbb{E}\left[\Big(\int_D|\mathscr{N}_{u,v}^{\Theta_{\mathbf{k}},\mathbf{l}}(x) - \mathcal{E}(x)|^2\,\mathbb{P}_{X_1}(\mathrm{d}x)\Big)^{\max\{1,p\}}\right]\right)^{\frac{1}{\max\{1,p\}}} \\
&\leq \big[\inf\nolimits_{\theta\in[-c,c]^{\mathbf{d}}}\sup\nolimits_{x\in D}|\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(x) - \mathcal{E}(x)|^2\big] \\
&\quad + \sup\nolimits_{\theta\in[-c,c]^{\mathbf{d}}}\big(\mathbb{E}\big[\min\nolimits_{k\in\{1,2,\dots,K\}}|\mathcal{R}(\Theta_{k,0}) - \mathcal{R}(\theta)|^{\max\{1,p\}}\big]\big)^{\frac{1}{\max\{1,p\}}} \\
&\quad + 2\big(\mathbb{E}\big[\sup\nolimits_{\theta\in[-B,B]^{\mathbf{d}}}|\mathcal{R}(\theta) - \mathbf{R}(\theta)|^{\max\{1,p\}}\big]\big)^{\frac{1}{\max\{1,p\}}} \\
&\leq \big[\inf\nolimits_{\theta\in[-c,c]^{\mathbf{d}}}\sup\nolimits_{x\in D}|\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(x) - \mathcal{E}(x)|^2\big] + \frac{4(v-u)b\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}}\max\{1,p\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} \\
&\quad + \frac{18\max\{1,(v-u)^2\}\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{p,\ln(3MBb)\}}{\sqrt{M}}.
\end{aligned} \tag{5.155}$$

Moreover, note that the fact that $\forall\, x \in [0,\infty)\colon x+1 \leq e^x \leq 3^x$ and the facts that $Bb \geq 1$ and $M \geq 1$ ensure that

$$\ln(3MBb) \leq \ln(3M3^{Bb-1}) = \ln(3^{Bb}M) = Bb\ln([3^{Bb}M]^{1/(Bb)}) \leq Bb\ln(3M). \tag{5.156}$$

The facts that $\|\mathbf{l}\|_\infty + 1 \geq 2$, $B \geq c \geq 1$, $\ln(3M) \geq 1$, $b \geq 1$, and $\mathbf{L} \geq 1$ hence show for all $p \in (0,\infty)$ that

$$\begin{aligned}
&\frac{4(v-u)b\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}}\max\{1,p\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} \\
&\quad + \frac{18\max\{1,(v-u)^2\}\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{p,\ln(3MBb)\}}{\sqrt{M}} \\
&\leq \frac{2(\|\mathbf{l}\|_\infty+1)\max\{1,(v-u)^2\}b\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}B^{\mathbf{L}}\max\{p,\ln(3M)\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} \\
&\quad + \frac{18\max\{1,(v-u)^2\}b\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2B\max\{p,\ln(3M)\}}{\sqrt{M}} \\
&\leq \frac{20\max\{1,(v-u)^2\}b\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}+1}B^{\mathbf{L}}\max\{p,\ln(3M)\}}{\min\{\sqrt{M},K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}\}}.
\end{aligned} \tag{5.157}$$

This and (5.155) complete the proof of Proposition 5.39. □

Lemma 5.40. *Let $a,x,p \in (0,\infty)$, $M,c \in [1,\infty)$, $B \in [c,\infty)$. Then* *(i) it holds that* $ax^p \leq \exp\big(a^{1/p}\frac{px}{e}\big)$ *and*

(ii) it holds that $\ln(3MBc) \leq \frac{23B}{18}\ln(eM)$.

Proof of Lemma 5.40. First, note that the fact that $\forall\, y \in \mathbb{R}\colon y + 1 \leq e^y$ demonstrates that

$$ax^p = (a^{1/p}x)^p = \big[e\big(a^{1/p}\tfrac{x}{e} - 1 + 1\big)\big]^p \leq \big[e\exp\big(a^{1/p}\tfrac{x}{e} - 1\big)\big]^p = \exp\big(a^{1/p}\tfrac{px}{e}\big). \tag{5.158}$$

This proves (i).

Second, observe that (i) and the fact that $2\sqrt{3}/e \leq 23/18$ ensure that

$$3B^2 \leq \exp\big(\sqrt{3}\tfrac{2B}{e}\big) = \exp\big(\tfrac{2\sqrt{3}B}{e}\big) \leq \exp\big(\tfrac{23B}{18}\big). \tag{5.159}$$

The facts that $B \geq c \geq 1$ and $M \geq 1$ hence imply that

$$\ln(3MBc) \leq \ln(3B^2M) \leq \ln([eM]^{23B/18}) = \tfrac{23B}{18}\ln(eM). \tag{5.160}$$

This establishes (ii). The proof of Lemma 5.40 is thus complete. □

Theorem 5.41. *Let* $d, \mathbf{d}, \mathbf{L}, M, K, N \in \mathbb{N}$, $A \in (0,\infty)$, $L, a, u \in \mathbb{R}$, $b \in (a,\infty)$, $v \in (u,\infty)$, $c \in [\max\{1, L, |a|, |b|, 2|u|, 2|v|\}, \infty)$, $B \in [c,\infty)$, $\mathbf{l} = (\mathbf{l}_0, \mathbf{l}_1, \ldots, \mathbf{l}_\mathbf{L}) \in \mathbb{N}^{\mathbf{L}+1}$, $\mathbf{N} \subseteq \{0, 1, \ldots, N\}$, *assume* $0 \in \mathbf{N}$, $\mathbf{L} \geq A\mathbb{1}_{(6^d,\infty)}(A)/(2d) + 1$, $\mathbf{l}_0 = d$, $\mathbf{l}_1 \geq A\mathbb{1}_{(6^d,\infty)}(A)$, $\mathbf{l}_\mathbf{L} = 1$, *and* $\mathbf{d} \geq \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1} + 1)$, *assume for all* $i \in \{2, 3, \ldots\} \cap [0, \mathbf{L})$ *that* $\mathbf{l}_i \geq \mathbb{1}_{(6^d,\infty)}(A)\max\{A/d - 2i + 3, 2\}$, *let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, let* $X_j\colon \Omega \to [a,b]^d$, $j \in \mathbb{N}$, *and* $Y_j\colon \Omega \to [u,v]$, $j \in \mathbb{N}$, *be functions, assume that* (X_j, Y_j), $j \in \{1, 2, \ldots, M\}$, *are i.i.d. random variables, let* $\mathcal{E}\colon [a,b]^d \to [u,v]$ *satisfy* $\mathbb{P}$*-a.s. that* $\mathcal{E}(X_1) = \mathbb{E}[Y_1|X_1]$, *assume for all* $x, y \in [a,b]^d$ *that* $|\mathcal{E}(x) - \mathcal{E}(y)| \leq L\|x - y\|_1$, *let* $\Theta_{k,n}\colon \Omega \to \mathbb{R}^\mathbf{d}$, $k, n \in \mathbb{N}_0$, *and* $\mathbf{k}\colon \Omega \to (\mathbb{N}_0)^2$ *be random variables, assume* $\big(\bigcup_{k=1}^\infty \Theta_{k,0}(\Omega)\big) \subseteq [-B,B]^\mathbf{d}$, *assume that* $\Theta_{k,0}$, $k \in \{1, 2, \ldots, K\}$, *are i.i.d., assume that* $\Theta_{1,0}$ *is continuous uniformly distributed on* $[-c,c]^\mathbf{d}$, *and let* $\mathcal{R}\colon \mathbb{R}^\mathbf{d} \times \Omega \to [0,\infty)$ *satisfy for all* $\theta \in \mathbb{R}^\mathbf{d}$, $\omega \in \Omega$ *that*

$$\mathcal{R}(\theta,\omega) = \frac{1}{M}\left[\sum_{j=1}^{M} |\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(X_j(\omega)) - Y_j(\omega)|^2\right] \qquad \textit{and} \tag{5.161}$$

$$\mathbf{k}(\omega) \in \arg\min\nolimits_{(k,n)\in\{1,2,\ldots,K\}\times\mathbf{N},\, \|\Theta_{k,n}(\omega)\|_\infty \leq B} \mathcal{R}(\Theta_{k,n}(\omega),\omega) \tag{5.162}$$

(cf. Definitions 5.8 and 5.9). Then it holds for all $p \in (0,\infty)$ *that*

$$\begin{aligned}
&\left(\mathbb{E}\left[\left(\int_{[a,b]^d} |\mathscr{N}_{u,v}^{\Theta_\mathbf{k},\mathbf{l}}(x) - \mathcal{E}(x)|^2\, \mathbb{P}_{X_1}(\mathrm{d}x)\right)^p\right]\right)^{1/p} \\
&\leq \frac{9d^2L^2(b-a)^2}{A^{2/d}} + \frac{4(v-u)\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^\mathbf{L} c^{\mathbf{L}+1}\max\{1,p\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} \\
&\quad + \frac{18\max\{1,(v-u)^2\}\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^2\max\{p, \ln(3MBc)\}}{\sqrt{M}} \\
&\leq \frac{36d^2c^4}{A^{2/d}} + \frac{4\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^\mathbf{L} c^{\mathbf{L}+2}\max\{1,p\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} + \frac{23B^3\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^2\max\{p,\ln(eM)\}}{\sqrt{M}}
\end{aligned} \tag{5.163}$$

(cf. (iii) in Lemma 5.38).

Proof of Theorem 5.41. First of all, note that the assumption that $\forall\, x, y \in [a,b]^d \colon |\mathcal{E}(x) - \mathcal{E}(y)| \leq L\|x-y\|_1$ ensures that $\mathcal{E} \colon [a,b]^d \to [u,v]$ is a $\mathcal{B}([a,b]^d)/\mathcal{B}([u,v])$-measurable function. The fact that $\max\{1,|a|,|b|\} \leq c$ and Proposition 5.39 (with $b \leftarrow \max\{1,|a|,|b|\}$, $D \leftarrow [a,b]^d$ in the notation of Proposition 5.39) hence show for all $p \in (0,\infty)$ that

$$\begin{aligned}
&\Big(\mathbb{E}\Big[\Big(\textstyle\int_{[a,b]^d}|\mathscr{N}^{\Theta_{\mathbf{k}},\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\,\mathbb{P}_{X_1}(\mathrm{d}x)\Big)^p\Big]\Big)^{1/p}\\
&\leq \big[\inf\nolimits_{\theta\in[-c,c]^{\mathbf{d}}}\sup\nolimits_{x\in[a,b]^d}|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\big]\\
&\quad+\frac{4(v-u)\max\{1,|a|,|b|\}\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}}\max\{1,p\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}}\\
&\quad+\frac{18\max\{1,(v-u)^2\}\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{p,\ln(3MB\max\{1,|a|,|b|\})\}}{\sqrt{M}}\\
&\leq \big[\inf\nolimits_{\theta\in[-c,c]^{\mathbf{d}}}\sup\nolimits_{x\in[a,b]^d}|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\big] + \frac{4(v-u)\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}+1}\max\{1,p\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}}\\
&\quad+\frac{18\max\{1,(v-u)^2\}\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{p,\ln(3MBc)\}}{\sqrt{M}}.
\end{aligned}\tag{5.164}$$

Furthermore, observe that Proposition 5.13 (with $f \leftarrow \mathcal{E}$ in the notation of Proposition 5.13) proves that there exists $\vartheta \in \mathbb{R}^{\mathbf{d}}$ such that $\|\vartheta\|_\infty \leq \max\{1, L, |a|, |b|, 2[\sup_{x\in[a,b]^d}|\mathcal{E}(x)|]\}$ and

$$\sup\nolimits_{x\in[a,b]^d}|\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)| \leq \frac{3dL(b-a)}{A^{1/d}}.\tag{5.165}$$

The fact that $\forall\, x \in [a,b]^d \colon \mathcal{E}(x) \in [u,v]$ hence implies that

$$\|\vartheta\|_\infty \leq \max\{1, L, |a|, |b|, 2|u|, 2|v|\} \leq c.\tag{5.166}$$

This and (5.165) demonstrate that

$$\begin{aligned}
&\inf\nolimits_{\theta\in[-c,c]^{\mathbf{d}}}\sup\nolimits_{x\in[a,b]^d}|\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\\
&\leq \sup\nolimits_{x\in[a,b]^d}|\mathscr{N}^{\vartheta,\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\\
&\leq \left[\frac{3dL(b-a)}{A^{1/d}}\right]^2 = \frac{9d^2L^2(b-a)^2}{A^{2/d}}.
\end{aligned}\tag{5.167}$$

Combining this with (5.164) establishes for all $p \in (0,\infty)$ that

$$\begin{aligned}
&\Big(\mathbb{E}\Big[\Big(\textstyle\int_{[a,b]^d}|\mathscr{N}^{\Theta_{\mathbf{k}},\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\,\mathbb{P}_{X_1}(\mathrm{d}x)\Big)^p\Big]\Big)^{1/p}\\
&\leq \frac{9d^2L^2(b-a)^2}{A^{2/d}} + \frac{4(v-u)\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}+1}\max\{1,p\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}}\\
&\quad+\frac{18\max\{1,(v-u)^2\}\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{p,\ln(3MBc)\}}{\sqrt{M}}.
\end{aligned}\tag{5.168}$$

Moreover, note that the facts that $\max\{1, L, |a|, |b|\} \leq c$ and $(b-a)^2 \leq (|a|+|b|)^2 \leq 2(a^2+b^2)$ yield that

$$9L^2(b-a)^2 \leq 18c^2(a^2+b^2) \leq 18c^2(c^2+c^2) = 36c^4.\tag{5.169}$$

In addition, the fact that $B \geq c \geq 1$, the fact that $M \geq 1$, and (ii) in Lemma 5.40 ensure that $\ln(3MBc) \leq \frac{23B}{18}\ln(eM)$. This, (5.169), the fact that $(v-u) \leq 2\max\{|u|,|v|\} = \max\{2|u|, 2|v|\} \leq c \leq B$, and the fact that $B \geq 1$ prove for all $p \in (0,\infty)$ that

$$\begin{aligned}
&\frac{9d^2L^2(b-a)^2}{A^{2/d}} + \frac{4(v-u)\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}+1}\max\{1,p\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}}\\
&+\frac{18\max\{1,(v-u)^2\}\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{p,\ln(3MBc)\}}{\sqrt{M}}\\
&\leq \frac{36d^2c^4}{A^{2/d}} + \frac{4\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}+2}\max\{1,p\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} + \frac{23B^3\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{p,\ln(eM)\}}{\sqrt{M}}.
\end{aligned} \tag{5.170}$$

Combining this with (5.168) shows (5.163). The proof of Theorem 5.41 is thus complete. □

Corollary 5.42. *Let $d,\mathbf{d},\mathbf{L},M,K,N \in \mathbb{N}$, $L,a,u \in \mathbb{R}$, $b \in (a,\infty)$, $v \in (u,\infty)$, $c \in [\max\{1,L,|a|,|b|,2|u|,2|v|\},\infty)$, $B \in [c,\infty)$, $\mathbf{l} = (\mathbf{l}_0,\mathbf{l}_1,\ldots,\mathbf{l}_\mathbf{L}) \in \mathbb{N}^{\mathbf{L}+1}$, $\mathbf{N} \subseteq \{0,1,\ldots,N\}$, assume $0 \in \mathbf{N}$, $\mathbf{l}_0 = d$, $\mathbf{l}_\mathbf{L} = 1$, and $\mathbf{d} \geq \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1}+1)$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $X_j\colon \Omega \to [a,b]^d$, $j \in \mathbb{N}$, and $Y_j\colon \Omega \to [u,v]$, $j \in \mathbb{N}$, be functions, assume that (X_j,Y_j), $j \in \{1,2,\ldots,M\}$, are i.i.d. random variables, let $\mathcal{E}\colon [a,b]^d \to [u,v]$ satisfy $\mathbb{P}$-a.s. that $\mathcal{E}(X_1) = \mathbb{E}[Y_1|X_1]$, assume for all $x,y \in [a,b]^d$ that $|\mathcal{E}(x)-\mathcal{E}(y)| \leq L\|x-y\|_1$, let $\Theta_{k,n}\colon \Omega \to \mathbb{R}^\mathbf{d}$, $k,n \in \mathbb{N}_0$, and $\mathbf{k}\colon \Omega \to (\mathbb{N}_0)^2$ be random variables, assume $\left(\bigcup_{k=1}^\infty \Theta_{k,0}(\Omega)\right) \subseteq [-B,B]^\mathbf{d}$, assume that $\Theta_{k,0}$, $k \in \{1,2,\ldots,K\}$, are i.i.d., assume that $\Theta_{1,0}$ is continuous uniformly distributed on $[-c,c]^\mathbf{d}$, and let $\mathcal{R}\colon \mathbb{R}^\mathbf{d} \times \Omega \to [0,\infty)$ satisfy for all $\theta \in \mathbb{R}^\mathbf{d}$, $\omega \in \Omega$ that*

$$\mathcal{R}(\theta,\omega) = \frac{1}{M}\left[\sum_{j=1}^M |\mathscr{N}^{\theta,\mathbf{l}}_{u,v}(X_j(\omega)) - Y_j(\omega)|^2\right] \quad \text{and} \tag{5.171}$$

$$\mathbf{k}(\omega) \in \arg\min\nolimits_{(k,n)\in\{1,2,\ldots,K\}\times\mathbf{N},\, \|\Theta_{k,n}(\omega)\|_\infty \leq B} \mathcal{R}(\Theta_{k,n}(\omega),\omega) \tag{5.172}$$

(cf. Definitions 5.8 and 5.9). Then it holds for all $p \in (0,\infty)$ that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left(\int_{[a,b]^d} |\mathscr{N}^{\Theta_\mathbf{k},\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2\,\mathbb{P}_{X_1}(\mathrm{d}x)\right)^{p/2}\right]\right)^{1/p}\\
&\leq \frac{3dL(b-a)}{[\min(\{\mathbf{L}\}\cup\{\mathbf{l}_i\colon i \in \mathbb{N}\cap[0,\mathbf{L})\})]^{1/d}} + \frac{2[(v-u)\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}+1}\max\{1,p/2\}]^{1/2}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}}\\
&\quad+\frac{3\max\{1,v-u\}(\|\mathbf{l}\|_\infty+1)[\mathbf{L}\max\{p,2\ln(3MBc)\}]^{1/2}}{M^{1/4}}\\
&\leq \frac{6dc^2}{[\min(\{\mathbf{L}\}\cup\{\mathbf{l}_i\colon i \in \mathbb{N}\cap[0,\mathbf{L})\})]^{1/d}} + \frac{2\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}+1}\max\{1,p\}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}}\\
&\quad+\frac{5B^2\mathbf{L}(\|\mathbf{l}\|_\infty+1)\max\{p,\ln(eM)\}}{M^{1/4}}
\end{aligned} \tag{5.173}$$

(cf. (iii) in Lemma 5.38).

Proof of Corollary 5.42. Throughout this proof let $A \in (0,\infty)$ be given by

$$A = \min(\{\mathbf{L}\}\cup\{\mathbf{l}_i\colon i \in \mathbb{N}\cap[0,\mathbf{L})\}). \tag{5.174}$$

Note that (5.174) ensures that

$$\begin{aligned}\mathbf{L} &\geq A = A - 1 + 1 \geq (A-1)\mathbb{1}_{[2,\infty)}(A) + 1 \\ &\geq \left(A - \tfrac{A}{2}\right)\mathbb{1}_{[2,\infty)}(A) + 1 = \tfrac{A\mathbb{1}_{[2,\infty)}(A)}{2} + 1 \geq \tfrac{A\mathbb{1}_{(6^d,\infty)}(A)}{2d} + 1.\end{aligned} \tag{5.175}$$

Moreover, the assumption that $\mathbf{l}_\mathbf{L} = 1$ and (5.174) imply that

$$\mathbf{l}_1 = \mathbf{l}_1 \mathbb{1}_{\{1\}}(\mathbf{L}) + \mathbf{l}_1 \mathbb{1}_{[2,\infty)}(\mathbf{L}) \geq \mathbb{1}_{\{1\}}(\mathbf{L}) + A\mathbb{1}_{[2,\infty)}(\mathbf{L}) = A \geq A\mathbb{1}_{(6^d,\infty)}(A). \tag{5.176}$$

Moreover, again (5.174) shows for all $i \in \{2, 3, \dots\} \cap [0, \mathbf{L})$ that

$$\begin{aligned}\mathbf{l}_i &\geq A \geq A\mathbb{1}_{[2,\infty)}(A) \geq \mathbb{1}_{[2,\infty)}(A)\max\{A-1, 2\} = \mathbb{1}_{[2,\infty)}(A)\max\{A - 4 + 3, 2\} \\ &\geq \mathbb{1}_{[2,\infty)}(A)\max\{A - 2i + 3, 2\} \geq \mathbb{1}_{(6^d,\infty)}(A)\max\{{}^{A}\!/\!_{d} - 2i + 3, 2\}.\end{aligned} \tag{5.177}$$

Combining (5.175)–(5.177) and Theorem 5.41 (with $p \leftarrow {}^{p}\!/\!_{2}$ for $p \in (0,\infty)$ in the notation of Theorem 5.41) establishes for all $p \in (0,\infty)$ that

$$\begin{aligned}&\left(\mathbb{E}\left[\left(\int_{[a,b]^d} |\mathscr{N}^{\Theta_\mathbf{k},\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x)\right)^{p/2}\right]\right)^{2/p} \\ &\leq \frac{9d^2L^2(b-a)^2}{A^{2/d}} + \frac{4(v-u)\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} c^{\mathbf{L}+1}\max\{1, {}^{p}\!/\!_{2}\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} \\ &\quad + \frac{18\max\{1,(v-u)^2\}\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{{}^{p}\!/\!_{2}, \ln(3MBc)\}}{\sqrt{M}} \\ &\leq \frac{36d^2c^4}{A^{2/d}} + \frac{4\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}+2}\max\{1,{}^{p}\!/\!_{2}\}}{K^{[\mathbf{L}^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} + \frac{23B^3\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{{}^{p}\!/\!_{2}, \ln(eM)\}}{\sqrt{M}}.\end{aligned} \tag{5.178}$$

This, (5.174), and the facts that $\mathbf{L} \geq 1$, $c \geq 1$, $B \geq 1$, and $\ln(eM) \geq 1$ demonstrate for all $p \in (0,\infty)$ that

$$\begin{aligned}&\left(\mathbb{E}\left[\left(\int_{[a,b]^d} |\mathscr{N}^{\Theta_\mathbf{k},\mathbf{l}}_{u,v}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1}(\mathrm{d}x)\right)^{p/2}\right]\right)^{1/p} \\ &\leq \frac{3dL(b-a)}{[\min(\{\mathbf{L}\} \cup \{\mathbf{l}_i \colon i \in \mathbb{N} \cap [0,\mathbf{L})\})]^{1/d}} + \frac{2[(v-u)\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}+1}\max\{1,{}^{p}\!/\!_{2}\}]^{1/2}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} \\ &\quad + \frac{3\max\{1, v-u\}(\|\mathbf{l}\|_\infty+1)[\mathbf{L}\max\{p, 2\ln(3MBc)\}]^{1/2}}{M^{1/4}} \\ &\leq \frac{6dc^2}{[\min(\{\mathbf{L}\} \cup \{\mathbf{l}_i \colon i \in \mathbb{N} \cap [0,\mathbf{L})\})]^{1/d}} + \frac{2[\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}+2}\max\{1,{}^{p}\!/\!_{2}\}]^{1/2}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} \\ &\quad + \frac{5B^3[\mathbf{L}(\|\mathbf{l}\|_\infty+1)^2\max\{{}^{p}\!/\!_{2}, \ln(eM)\}]^{1/2}}{M^{1/4}} \\ &\leq \frac{6dc^2}{[\min(\{\mathbf{L}\} \cup \{\mathbf{l}_i \colon i \in \mathbb{N} \cap [0,\mathbf{L})\})]^{1/d}} + \frac{2\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}+1}\max\{1,p\}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} \\ &\quad + \frac{5B^2\mathbf{L}(\|\mathbf{l}\|_\infty+1)\max\{p,\ln(eM)\}}{M^{1/4}}.\end{aligned} \tag{5.179}$$

The proof of Corollary 5.42 is thus complete. □

5.5.3 Overall strong error analysis for the training of DNNs with optimisation via SGD with random initialisation

Corollary 5.43. *Let* $d, \mathbf{d}, \mathbf{L}, M, K, N \in \mathbb{N}$, $L, a, u \in \mathbb{R}$, $b \in (a, \infty)$, $v \in (u, \infty)$, $c \in [\max\{1, L, |a|, |b|, 2|u|, 2|v|\}, \infty)$, $B \in [c, \infty)$, $\mathbf{l} = (\mathbf{l}_0, \mathbf{l}_1, \dots, \mathbf{l}_\mathbf{L}) \in \mathbb{N}^{\mathbf{L}+1}$, $\mathbf{N} \subseteq \{0, 1, \dots, N\}$, $(\mathbf{J}_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$, *assume* $0 \in \mathbf{N}$, $\mathbf{l}_0 = d$, $\mathbf{l}_\mathbf{L} = 1$, *and* $\mathbf{d} \geq \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1} + 1)$, *let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, let* $X_j^{k,n} \colon \Omega \to [a, b]^d$, $k, n, j \in \mathbb{N}_0$, *and* $Y_j^{k,n} \colon \Omega \to [u, v]$, $k, n, j \in \mathbb{N}_0$, *be functions, assume that* $(X_j^{0,0}, Y_j^{0,0})$, $j \in \{1, 2, \dots, M\}$, *are i.i.d. random variables, let* $\mathcal{E} \colon [a, b]^d \to [u, v]$ *satisfy* $\mathbb{P}$*-a.s. that* $\mathcal{E}(X_1^{0,0}) = \mathbb{E}[Y_1^{0,0}|X_1^{0,0}]$, *assume for all* $x, y \in [a, b]^d$ *that* $|\mathcal{E}(x) - \mathcal{E}(y)| \leq L\|x - y\|_1$, *let* $\Theta_{k,n} \colon \Omega \to \mathbb{R}^\mathbf{d}$, $k, n \in \mathbb{N}_0$, *and* $\mathbf{k} \colon \Omega \to (\mathbb{N}_0)^2$ *be random variables, assume* $\left(\bigcup_{k=1}^{\infty} \Theta_{k,0}(\Omega)\right) \subseteq [-B, B]^\mathbf{d}$, *assume that* $\Theta_{k,0}$, $k \in \{1, 2, \dots, K\}$, *are i.i.d., assume that* $\Theta_{1,0}$ *is continuous uniformly distributed on* $[-c, c]^\mathbf{d}$, *let* $\mathcal{R}_J^{k,n} \colon \mathbb{R}^\mathbf{d} \times \Omega \to [0, \infty)$, $k, n, J \in \mathbb{N}_0$, *and* $\mathcal{G}^{k,n} \colon \mathbb{R}^\mathbf{d} \times \Omega \to \mathbb{R}^\mathbf{d}$, $k, n \in \mathbb{N}$, *satisfy for all* $k, n \in \mathbb{N}$, $\omega \in \Omega$, $\theta \in \{\vartheta \in \mathbb{R}^\mathbf{d} \colon (\mathcal{R}_{\mathbf{J}_n}^{k,n}(\cdot, \omega) \colon \mathbb{R}^\mathbf{d} \to [0, \infty)$ *is differentiable at* $\vartheta)\}$ *that* $\mathcal{G}^{k,n}(\theta, \omega) = (\nabla_\theta \mathcal{R}_{\mathbf{J}_n}^{k,n})(\theta, \omega)$, *assume for all* $k, n \in \mathbb{N}$ *that* $\Theta_{k,n} = \Theta_{k,n-1} - \gamma_n \mathcal{G}^{k,n}(\Theta_{k,n-1})$, *and assume for all* $k, n \in \mathbb{N}_0$, $J \in \mathbb{N}$, $\theta \in \mathbb{R}^\mathbf{d}$, $\omega \in \Omega$ *that*

$$\mathcal{R}_J^{k,n}(\theta, \omega) = \frac{1}{J}\left[\sum_{j=1}^{J} |\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(X_j^{k,n}(\omega)) - Y_j^{k,n}(\omega)|^2\right] \qquad \text{and} \tag{5.180}$$

$$\mathbf{k}(\omega) \in \arg\min\nolimits_{(l,m)\in\{1,2,\dots,K\}\times\mathbf{N},\, \|\Theta_{l,m}(\omega)\|_\infty \leq B} \mathcal{R}_M^{0,0}(\Theta_{l,m}(\omega), \omega) \tag{5.181}$$

(cf. Definitions 5.8 and 5.9). Then it holds for all $p \in (0, \infty)$ *that*

$$\begin{aligned}
&\left(\mathbb{E}\left[\left(\int_{[a,b]^d} |\mathscr{N}_{u,v}^{\Theta_\mathbf{k},\mathbf{l}}(x) - \mathcal{E}(x)|^2 \, \mathbb{P}_{X_1^{0,0}}(\mathrm{d}x)\right)^{p/2}\right]\right)^{1/p} \\
&\leq \frac{3dL(b-a)}{[\min(\{\mathbf{L}\} \cup \{\mathbf{l}_i \colon i \in \mathbb{N} \cap [0, \mathbf{L})\})]^{1/d}} + \frac{2[(v-u)\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^\mathbf{L} c^{\mathbf{L}+1} \max\{1, p/2\}]^{1/2}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty + 1)^{-2}]}} \\
&\quad + \frac{3\max\{1, v-u\}(\|\mathbf{l}\|_\infty + 1)[\mathbf{L}\max\{p, 2\ln(3MBc)\}]^{1/2}}{M^{1/4}} \\
&\leq \frac{6dc^2}{[\min(\{\mathbf{L}\} \cup \{\mathbf{l}_i \colon i \in \mathbb{N} \cap [0, \mathbf{L})\})]^{1/d}} + \frac{2\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^\mathbf{L} c^{\mathbf{L}+1} \max\{1, p\}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty + 1)^{-2}]}} \\
&\quad + \frac{5B^2\mathbf{L}(\|\mathbf{l}\|_\infty + 1)\max\{p, \ln(eM)\}}{M^{1/4}}
\end{aligned} \tag{5.182}$$

(cf. (iii) in Lemma 5.38).

Proof of Corollary 5.43. Observe that Corollary 5.42 (with $(X_j)_{j\in\mathbb{N}} \leftarrow (X_j^{0,0})_{j\in\mathbb{N}}$, $(Y_j)_{j\in\mathbb{N}} \leftarrow (Y_j^{0,0})_{j\in\mathbb{N}}$, $\mathcal{R} \leftarrow \mathcal{R}_M^{0,0}$ in the notation of Corollary 5.42) shows (5.182). The proof of Corollary 5.43 is thus complete. □

Corollary 5.44. *Let* $d, \mathbf{d}, \mathbf{L}, M, K, N \in \mathbb{N}$, $L, a, u \in \mathbb{R}$, $b \in (a, \infty)$, $v \in (u, \infty)$, $c \in [\max\{1, L, |a|, |b|, 2|u|, 2|v|\}, \infty)$, $B \in [c, \infty)$, $\mathbf{l} = (\mathbf{l}_0, \mathbf{l}_1, \dots, \mathbf{l}_\mathbf{L}) \in \mathbb{N}^{\mathbf{L}+1}$, $\mathbf{N} \subseteq \{0, 1, \dots, N\}$, $(\mathbf{J}_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$, *assume* $0 \in \mathbf{N}$, $\mathbf{l}_0 = d$, $\mathbf{l}_\mathbf{L} = 1$, *and* $\mathbf{d} \geq \sum_{i=1}^{\mathbf{L}} \mathbf{l}_i(\mathbf{l}_{i-1} + 1)$, *let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, let* $X_j^{k,n} \colon \Omega \to [a, b]^d$, $k, n, j \in \mathbb{N}_0$, *and* $Y_j^{k,n} \colon \Omega \to [u, v]$, $k, n, j \in \mathbb{N}_0$, *be functions, assume that* $(X_j^{0,0}, Y_j^{0,0})$, $j \in \{1, 2, \dots, M\}$, *are i.i.d. random variables, let* $\mathcal{E} \colon [a, b]^d \to [u, v]$ *satisfy* $\mathbb{P}$*-a.s. that* $\mathcal{E}(X_1^{0,0}) = \mathbb{E}[Y_1^{0,0}|X_1^{0,0}]$, *assume for all* $x, y \in [a, b]^d$ *that* $|\mathcal{E}(x) - \mathcal{E}(y)| \leq L\|x - y\|_1$, *let* $\Theta_{k,n} \colon \Omega \to \mathbb{R}^\mathbf{d}$, $k, n \in \mathbb{N}_0$, *and* $\mathbf{k} \colon \Omega \to (\mathbb{N}_0)^2$ *be random variables, assume* $\left(\bigcup_{k=1}^{\infty} \Theta_{k,0}(\Omega)\right) \subseteq [-B, B]^\mathbf{d}$, *assume that* $\Theta_{k,0}$, $k \in$

$\{1,2,\ldots,K\}$, are i.i.d., assume that $\Theta_{1,0}$ is continuous uniformly distributed on $[-c,c]^{\mathbf{d}}$, let $\mathcal{R}_J^{k,n}\colon \mathbb{R}^{\mathbf{d}}\times\Omega\to[0,\infty)$, $k,n,J\in\mathbb{N}_0$, and $\mathcal{G}^{k,n}\colon\mathbb{R}^{\mathbf{d}}\times\Omega\to\mathbb{R}^{\mathbf{d}}$, $k,n\in\mathbb{N}$, satisfy for all $k,n\in\mathbb{N}$, $\omega\in\Omega$, $\theta\in\{\vartheta\in\mathbb{R}^{\mathbf{d}}\colon(\mathcal{R}_{\mathbf{J}_n}^{k,n}(\cdot,\omega)\colon\mathbb{R}^{\mathbf{d}}\to[0,\infty)$ is differentiable at $\vartheta)\}$ that $\mathcal{G}^{k,n}(\theta,\omega)=(\nabla_\theta\mathcal{R}_{\mathbf{J}_n}^{k,n})(\theta,\omega)$, assume for all $k,n\in\mathbb{N}$ that $\Theta_{k,n}=\Theta_{k,n-1}-\gamma_n\mathcal{G}^{k,n}(\Theta_{k,n-1})$, and assume for all $k,n\in\mathbb{N}_0$, $J\in\mathbb{N}$, $\theta\in\mathbb{R}^{\mathbf{d}}$, $\omega\in\Omega$ that

$$\mathcal{R}_J^{k,n}(\theta,\omega)=\frac{1}{J}\left[\sum_{j=1}^{J}|\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(X_j^{k,n}(\omega))-Y_j^{k,n}(\omega)|^2\right]\quad\text{and}\tag{5.183}$$

$$\mathbf{k}(\omega)\in\arg\min\nolimits_{(l,m)\in\{1,2,\ldots,K\}\times\mathbf{N},\,\|\Theta_{l,m}(\omega)\|_\infty\le B}\mathcal{R}_M^{0,0}(\Theta_{l,m}(\omega),\omega)\tag{5.184}$$

(cf. Definitions 5.8 and 5.9). Then

$$\begin{aligned}\mathbb{E}\Big[\int_{[a,b]^d}|\mathscr{N}_{u,v}^{\Theta_{\mathbf{k}},\mathbf{l}}(x)-\mathcal{E}(x)|\,\mathbb{P}_{X_1^{0,0}}(dx)\Big]&\le\frac{2[(v-u)\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}+1}]^{1/2}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}}\\&\quad+\frac{3dL(b-a)}{[\min\{\mathbf{L},\mathbf{l}_1,\mathbf{l}_2,\ldots,\mathbf{l}_{\mathbf{L}-1}\}]^{1/d}}+\frac{3\max\{1,v-u\}(\|\mathbf{l}\|_\infty+1)[2\mathbf{L}\ln(3MBc)]^{1/2}}{M^{1/4}}\\&\le\frac{6dc^2}{[\min\{\mathbf{L},\mathbf{l}_1,\mathbf{l}_2,\ldots,\mathbf{l}_{\mathbf{L}-1}\}]^{1/d}}+\frac{5B^2\mathbf{L}(\|\mathbf{l}\|_\infty+1)\ln(eM)}{M^{1/4}}+\frac{2\mathbf{L}(\|\mathbf{l}\|_\infty+1)^{\mathbf{L}}c^{\mathbf{L}+1}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}}\end{aligned}\tag{5.185}$$

(cf. (iii) in Lemma 5.38).

Proof of Corollary 5.44. Note that Jensen's inequality implies that

$$\mathbb{E}\Big[\int_{[a,b]^d}|\mathscr{N}_{u,v}^{\Theta_{\mathbf{k}},\mathbf{l}}(x)-\mathcal{E}(x)|\,\mathbb{P}_{X_1^{0,0}}(dx)\Big]\le\mathbb{E}\Big[\Big(\int_{[a,b]^d}|\mathscr{N}_{u,v}^{\Theta_{\mathbf{k}},\mathbf{l}}(x)-\mathcal{E}(x)|^2\,\mathbb{P}_{X_1^{0,0}}(dx)\Big)^{1/2}\Big].\tag{5.186}$$

This and Corollary 5.43 (with $p\leftarrow 1$ in the notation of Corollary 5.43) complete the proof of Corollary 5.44. ☐

Corollary 5.45. *Let $d,\mathbf{d},\mathbf{L},M,K,N\in\mathbb{N}$, $L\in\mathbb{R}$, $c\in[\max\{2,L\},\infty)$, $B\in[c,\infty)$, $\mathbf{l}=(\mathbf{l}_0,\mathbf{l}_1,\ldots,\mathbf{l}_{\mathbf{L}})\in\mathbb{N}^{\mathbf{L}+1}$, $\mathbf{N}\subseteq\{0,1,\ldots,N\}$, $(\mathbf{J}_n)_{n\in\mathbb{N}}\subseteq\mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}}\subseteq\mathbb{R}$, assume $0\in\mathbf{N}$, $\mathbf{l}_0=d$, $\mathbf{l}_{\mathbf{L}}=1$, and $\mathbf{d}\ge\sum_{i=1}^{\mathbf{L}}\mathbf{l}_i(\mathbf{l}_{i-1}+1)$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $X_j^{k,n}\colon\Omega\to[0,1]^d$, $k,n,j\in\mathbb{N}_0$, and $Y_j^{k,n}\colon\Omega\to[0,1]$, $k,n,j\in\mathbb{N}_0$, be functions, assume that $(X_j^{0,0},Y_j^{0,0})$, $j\in\{1,2,\ldots,M\}$, are i.i.d. random variables, let $\mathcal{E}\colon[0,1]^d\to[0,1]$ satisfy $\mathbb{P}$-a.s. that $\mathcal{E}(X_1^{0,0})=\mathbb{E}[Y_1^{0,0}|X_1^{0,0}]$, assume for all $x,y\in[0,1]^d$ that $|\mathcal{E}(x)-\mathcal{E}(y)|\le L\|x-y\|_1$, let $\Theta_{k,n}\colon\Omega\to\mathbb{R}^{\mathbf{d}}$, $k,n\in\mathbb{N}_0$, and $\mathbf{k}\colon\Omega\to(\mathbb{N}_0)^2$ be random variables, assume $\big(\bigcup_{k=1}^\infty\Theta_{k,0}(\Omega)\big)\subseteq[-B,B]^{\mathbf{d}}$, assume that $\Theta_{k,0}$, $k\in\{1,2,\ldots,K\}$, are i.i.d., assume that $\Theta_{1,0}$ is continuous uniformly distributed on $[-c,c]^{\mathbf{d}}$, let $\mathcal{R}_J^{k,n}\colon\mathbb{R}^{\mathbf{d}}\times\Omega\to[0,\infty)$, $k,n,J\in\mathbb{N}_0$, and $\mathcal{G}^{k,n}\colon\mathbb{R}^{\mathbf{d}}\times\Omega\to\mathbb{R}^{\mathbf{d}}$, $k,n\in\mathbb{N}$, satisfy for all $k,n\in\mathbb{N}$, $\omega\in\Omega$, $\theta\in\{\vartheta\in\mathbb{R}^{\mathbf{d}}\colon(\mathcal{R}_{\mathbf{J}_n}^{k,n}(\cdot,\omega)\colon\mathbb{R}^{\mathbf{d}}\to[0,\infty)$ is differentiable at $\vartheta)\}$ that $\mathcal{G}^{k,n}(\theta,\omega)=(\nabla_\theta\mathcal{R}_{\mathbf{J}_n}^{k,n})(\theta,\omega)$, assume for all $k,n\in\mathbb{N}$ that $\Theta_{k,n}=\Theta_{k,n-1}-\gamma_n\mathcal{G}^{k,n}(\Theta_{k,n-1})$, and assume for all $k,n\in\mathbb{N}_0$, $J\in\mathbb{N}$, $\theta\in\mathbb{R}^{\mathbf{d}}$, $\omega\in\Omega$ that*

$$\mathcal{R}_J^{k,n}(\theta,\omega)=\frac{1}{J}\left[\sum_{j=1}^{J}|\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(X_j^{k,n}(\omega))-Y_j^{k,n}(\omega)|^2\right]\quad\text{and}\tag{5.187}$$

$$\mathbf{k}(\omega)\in\arg\min\nolimits_{(l,m)\in\{1,2,\ldots,K\}\times\mathbf{N},\,\|\Theta_{l,m}(\omega)\|_\infty\le B}\mathcal{R}_M^{0,0}(\Theta_{l,m}(\omega),\omega)\tag{5.188}$$

(cf. Definitions 5.8 and 5.9). Then

$$
\begin{aligned}
&\mathbb{E}\Big[\int_{[0,1]^d} |\mathscr{N}_{u,v}^{\Theta_{\mathbf{k}},\mathbf{l}}(x) - \mathcal{E}(x)|\, \mathbb{P}_{X_1^{0,0}}(\mathrm{d}x)\Big] \\
&\le \frac{3dL}{[\min\{\mathbf{L}, \mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_{\mathbf{L}-1}\}]^{1/d}} + \frac{3(\|\mathbf{l}\|_\infty + 1)[2\mathbf{L}\ln(3MBc)]^{1/2}}{M^{1/4}} + \frac{2[\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} c^{\mathbf{L}+1}]^{1/2}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} \\
&\le \frac{dc^3}{[\min\{\mathbf{L}, \mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_{\mathbf{L}-1}\}]^{1/d}} + \frac{B^3\mathbf{L}(\|\mathbf{l}\|_\infty + 1)\ln(eM)}{M^{1/4}} + \frac{\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} c^{\mathbf{L}+1}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}}
\end{aligned}
\tag{5.189}
$$

(cf. (iii) in Lemma 5.38).

Proof of Corollary 5.45. Observe that Corollary 5.44 (with $a \leftarrow 0$, $u \leftarrow 0$, $b \leftarrow 1$, $v \leftarrow 1$ in the notation of Corollary 5.44), the facts that $B \ge c \ge \max\{2, L\}$ and $M \ge 1$, and (ii) in Lemma 5.40 show that

$$
\begin{aligned}
&\mathbb{E}\Big[\int_{[0,1]^d} |\mathscr{N}_{u,v}^{\Theta_{\mathbf{k}},\mathbf{l}}(x) - \mathcal{E}(x)|\, \mathbb{P}_{X_1^{0,0}}(\mathrm{d}x)\Big] \\
&\le \frac{3dL}{[\min\{\mathbf{L}, \mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_{\mathbf{L}-1}\}]^{1/d}} + \frac{3(\|\mathbf{l}\|_\infty + 1)[2\mathbf{L}\ln(3MBc)]^{1/2}}{M^{1/4}} + \frac{2[\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} c^{\mathbf{L}+1}]^{1/2}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} \\
&\le \frac{dc^3}{[\min\{\mathbf{L}, \mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_{\mathbf{L}-1}\}]^{1/d}} + \frac{(\|\mathbf{l}\|_\infty + 1)[23B\mathbf{L}\ln(eM)]^{1/2}}{M^{1/4}} + \frac{[\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} c^{2\mathbf{L}+2}]^{1/2}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}} \\
&\le \frac{dc^3}{[\min\{\mathbf{L}, \mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_{\mathbf{L}-1}\}]^{1/d}} + \frac{B^3\mathbf{L}(\|\mathbf{l}\|_\infty + 1)\ln(eM)}{M^{1/4}} + \frac{\mathbf{L}(\|\mathbf{l}\|_\infty + 1)^{\mathbf{L}} c^{\mathbf{L}+1}}{K^{[(2\mathbf{L})^{-1}(\|\mathbf{l}\|_\infty+1)^{-2}]}}.
\end{aligned}
\tag{5.190}
$$

The proof of Corollary 5.45 is thus complete. □

Chapter 6

Conclusion and Outlook

In this thesis and in the preprints Jacobe de Naurois, Jentzen, & Welti [185], Giles, Jentzen, & Welti [134], Becker et al. [30], and Jentzen & Welti [196], which make up this thesis, we have studied stochastic numerical approximation algorithms for tackling four possibly high-dimensional approximation problems. On the one hand, for three of these problems we have carried out a mathematically rigorous analysis of existing stochastic numerical approximation algorithms from the scientific literature and derived corresponding convergence rates (cf. Chapters 2, 3, and 5). Thereby we have been able to gain a better understanding of the way the approximation error arising from employing these algorithms behaves with respect to the computational effort invested into running the algorithm. On the other hand, for one of the considered problems we have designed a stochastic numerical approximation algorithm and presented the results of suitable numerical experiments (cf. Chapter 4). These computational results suggest that the algorithm lives up to the expectations which have led to its design and that the approximation error decays quickly with increasing computational effort. The findings of this thesis and the preprints incorporated into this thesis give rise to a multitude of new research questions which may be the subjects of future research endeavours. In the following we reflect on these findings and mention a number of such emerging questions.

6.1 Stochastic wave equations

In Chapter 2 we have proved essentially sharp rates of convergence in the probabilistically weak sense for spatial spectral Galerkin approximations of semi-linear stochastic wave equations with multiplicative noise. In particular, we have established that spatial spectral Galerkin approximations for the continuous version of the hyperbolic Anderson model converge with weak rate 1- to the true solution (cf. Corollary 2.18 in Subsection 2.2.3). Note that the considered approximations cannot be implemented directly on a computer since they are discretised only in space but remain continuous in time and driven by infinite-dimensional noise.

In the more recent work Cox, Jentzen, & Lindner [86] weak convergence rates for temporal numerical approximations of semi-linear stochastic wave equations with multiplicative noise are derived. More specifically, [86, Theorem 1.1] shows that exponential Euler approximations for the continuous version of the hyperbolic Anderson model converge with weak rate 1- to the true solution. These approximations are discretised only in time, while space and noise remain infinite-dimensional, and are thus not directly implementable either. In view of the results of Chapter 2 and [86] it is a natural next step to consider weak convergence rates of a fully discrete approximation scheme for semi-linear stochastic wave equations with multiplicative noise that combines a noise discretisation (cf., e.g., Harms & Müller [162, Subsection 3.1]) with spectral Galerkin discretisation of space and exponential Euler discretisation of time. In the case of the continuous version of the hyperbolic Anderson model the resulting approximations would be implementable on a computer and, as a consequence, a theoretically proved weak convergence rate could be complemented by numerical experiments.

Moreover, so far we have only studied stochastic wave equations with one spatial dimension. Another possible subject of future research efforts is to examine weak convergence properties of numerical approximations for two- and three-dimensional stochastic wave equations with suitably coloured multiplicative noise. In this context, we also note that spatial spectral Galerkin discretisations of space are in practice only feasible to compute in the one-dimensional case or for simple two- and three-dimensional domains, such as rectangles and rectangular cuboids. In order to approximate stochastic wave equations on more complicated domains, we need to discretise space using more sophisticated algorithms, such as, for example, finite element methods. Deriving weak convergence rates for this case does not seem to be possible to achieve using the methodology from Chapter 2. It is thus a topic for further research to establish essentially sharp weak convergence rates for such more sophisticated numerical approximations of stochastic wave equations with multiplicative noise. Furthermore, a variety of relevant research questions about weak convergence rates for numerical approximations of other important SPDEs remain open. An example is the question of how to prove essentially sharp weak convergence rates for spatial spectral Galerkin approximations of the stochastic Burgers equation with additive space-time white noise.

6.2 Generalised multilevel Picard approximations

The main contribution of Chapter 3 has been the development of a mathematical framework in which in essence slight generalisations of the MLP approximations introduced in Hutzenthaler et al. [181] are viewed as random variables taking values in a Banach space. On this level of abstraction we have derived a complete error analysis, cost analysis, and complexity analysis of generalised MLP approximations (cf. Corollary 3.15 in Subsection 3.1.6). Thereafter, we have shown that the framework, when applied to semi-linear heat equations with gradient-independent and globally Lipschitz continuous nonlinearities, allows us to recover a complexity result similar to [181, Theorem 1.1], stating that MLP approximations overcome the curse of dimensionality (cf. Theorem 3.33 in Subsection 3.2.3).

In [181] MLP methods were introduced based on approximating time integrals for the fist time using the Monte Carlo method instead of fixed-grid quadrature rules (cf. [111, 112, 183]). In the meantime, beside the preprint [134], containing the content of Chapter 3, other works have proposed several variants of MLP approximations. More specifically, in Hutzenthaler, Jentzen, & von Wurstemberger [182] the MLP algorithm from [181] for semi-linear heat equations is generalised to approximate a larger class of semi-linear

Kolmogorov PDEs without the curse of dimensionality (cf. [182, Theorem 3.20]). The MLP approximations from Beck et al. [26] are truncated variants of the ones in [181] and are shown to beat the curse of dimensionality in the numerical approximation of reaction–diffusion-type PDEs with a locally Lipschitz continuous coercive non-linearity (cf. [26, Theorem 4.5]). Moreover, Hutzenthaler, Jentzen, & Kruse [179] proposes a new MLP algorithm that solves, provably without suffering from the curse of dimensionality, non-linear heat equations with gradient-dependent non-linearities (cf. [179, Theorem 5.2]). Finally, in Beck, Gonon, & Jentzen [25] a new MLP algorithm is introduced that is proven to approximate certain semi-linear elliptic PDEs and to overcome the curse of dimensionality in doing so (cf. [25, Theorem 3.16]).

A mathematical framework through which all or most of the different MLP approximations from [25, 26, 179, 181, 182] could be interpreted in a unified way would significantly improve the general understanding of MLP algorithms and their ability to beat the curse of dimensionality in several PDE approximation problems. The framework from Chapter 3 appears to be a first step into the right direction. Presumably, it essentially covers, apart from the MLP approximations in [181], also the MLP approximations from [182] and can, presumably, be applied to prove a variant of [182, Theorem 3.20]. However, it does not seem to be possible to employ the framework from Chapter 3 to recover suitable complexity results for the MLP approximations from [25, 26, 179]. Furthermore, such a unifying framework for MLP algorithms would be particularly insightful if it also comprised classical multilevel Monte Carlo methods (cf. Heinrich [167], Giles [133], and, e.g., Cox et al. [85, Subsection 5.3]). In this case it would allow contrasting the latter with MLP algorithms in a concise way and working out key similarities and differences in their respective modes of action, of which the further development of multilevel Monte Carlo-type algorithms could benefit.

The subject of a possible future research article is also to improve the framework from Chapter 3 by reformulating it in such a way that some of the measurability and integrability assumptions can be weakened. Doing this could simplify the verification of said assumptions significantly when the framework is specialised to the context of concrete PDEs. Moreover, we recall that one of the innovations of Chapter 3 and the preprint [134], respectively, has been employing the sequence of Monte Carlo numbers $(M_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ in the definition of generalised MLP approximations (cf. (1.6), (1.13), and the last paragraph in Section 1.2). It would be interesting to conduct a number of numerical experiments to see the possible impact of the freedom in choosing this sequence in practice. Last but not least, it would be fascinating to come up with MLP algorithms exhibiting even faster convergence speeds for high-dimensional PDEs with suitably regular non-linearities as well as MLP algorithms capable of solving high-dimensional PDEs with boundary conditions instead of just PDEs defined on the whole Euclidean space.

6.3 Optimal stopping problems

Chapter 4 has been devoted to a deep learning based algorithm for solving high-dimensional optimal stopping problems, which computes, in the context of early exercise option pricing, both approximations for an optimal exercise strategy and the price (cf. Subsection 4.1.7). While we have presented many numerical experiments which strongly suggest that the algorithm yields accurate and reliable approximations of the price (cf.

Section 4.3), of course there is now a natural desire for more mathematical theory that is capable of rigorously justifying under which precise assumptions this is indeed the case.

In Becker, Cheridito, & Jentzen [28] a few steps into this direction have already been taken. More specifically, [28, Theorem 1 and Remark 2] essentially demonstrate that, in the Markovian case, the stopping time factors from Lemma 4.2 in Subsection 4.1.3 do not need to take past but only current values of the process to be stopped into account (cf. Subsection 4.1.4). In addition, [28, Corollary 5] in essence shows that approximate stopping decisions based on appropriate artificial neural networks with fixed depth and at least one hidden layer have the flexibility to yield stopping times with expected pay-offs that are arbitrarily close to the optimal expected pay-off. Since the approximations for optimal stopping times in Subsection 4.1.7 are based on a single learning procedure for all approximate stopping time factors simultaneously and not on recursively learned approximate stopping decisions (cf. the last paragraph in Section 1.3), [28, Corollary 5] does, however, not apply directly to the algorithm from the preprint [30] and Chapter 4, respectively. It is a possible subject of future work to prove a similar result that also covers approximate optimal stopping times such as the ones delivered by this algorithm.

Nevertheless, [28, Corollary 5] is a mathematical existence result that leaves many for applications highly important questions unanswered. For example, significantly more research efforts need to be invested into understanding DNN architectures best suited for approximatively solving optimal stopping problems, in particular, in view of the fact that all computational results in [28, 30] have been obtained using artificial neural networks with precisely two hidden layers. In addition, an explanation is required of the reasons why stochastic gradient ascent and more sophisticated optimisation algorithms such as the Adam optimiser are able to find parameter vectors which yield sufficiently accurate results for objective functions as complicated as (4.39) in Subsection 4.1.5. Moreover, it is a central aim to derive convergence rates for the approximations of the optimal expected pay-off in terms of various algorithm and model parameters. This may eventually allow to prove that the algorithms from [28, 30] overcome the curse of dimensionality for a large class of optimal stopping problems.

Another direction for a future research article is to use deep learning based algorithms to tackle further relevant optimal stopping problems. One non-Markovian example could be Robbins' problem, an optimal stopping problem which is also referred to as the expected rank problem under full information (cf., e.g., Bruss & Ferguson [62] and Meier & Sögner [242]).

6.4 Empirical risk minimisation

In Chapter 5 we have established a strong convergence analysis of the overall error which emerges in deep learning based empirical risk minimisation in the case that the loss function is quadratic. In particular, we have decomposed the overall error into the approximation error, the generalisation error, and the optimisation error and derived strong convergence rates for each of these three error sources separately. The achieved results are instructive in many ways, especially because the dependence of the obtained error bounds on all algorithm and model parameters, such as depth and width of the employed DNNs and the dimension of the training sample space, is explicit (cf. Theorem 1.4 in Section 1.4). However, the convergence speeds in our results for both the approximation error and the optimisation error suffer under the curse of dimensionality (cf. (1.18) in Theorem 1.4).

A natural question to study as a next step is by how much the obtained convergence rates for the individual error sources can be improved in the generality considered. While we expect that the error term corresponding to the generalisation error in (1.18) in Theorem 1.4 can hardly be improved, it would be interesting to prove lower bounds for the error term corresponding to the approximation error in (1.18). This would clarify to which extent a different set of assumptions on the target function would be required in order to achieve a faster rate of convergence.

Furthermore, although our framework includes the case of training via SGD with random initialisation, our analysis of the optimisation error in Section 5.4 yields a very slow convergence speed since it relies on the convergence of the Minimum Monte Carlo method. It is clear that future research efforts need to exploit the dynamics of SGD and more advanced optimisers in order to prove strong convergence rates which possess more explanatory power for the successful performance of DNN based supervised learning algorithms in applications.

To summarise, it remains an open problem how to develop a convergence analysis capable of rigorously explaining the success of deep learning based empirical risk minimisation witnessed in practice. The same also applies to the algorithm for solving optimal stopping problems from Chapter 4 (cf. Section 6.3) and to deep learning based algorithms in general. We thus live in exciting times for developing more mathematical theory for deep learning.

References

[1] AGARWAL, A., AND CLAISSE, J. Branching diffusion representation of semi-linear elliptic PDEs and estimation using Monte Carlo method. *ArXiv e-prints* (2017), 24 pages. arXiv: 1704.00328 [math.PR].

[2] AGARWAL, R. P. *Difference Equations and Inequalities*. Second. Vol. 228. Monographs and Textbooks in Pure and Applied Mathematics. Theory, Methods, and Applications. Marcel Dekker, Inc., New York, 2000, xvi+971. ISBN: 0-8247-9007-3.

[3] AITSAHLIA, F., AND CARR, P. American Options: A Comparison of Numerical Methods. In: *Numerical Methods in Finance*. Ed. by ROGERS, L. C. G., AND TALAY, D. Vol. 13. Publ. Newton Inst. Cambridge Univ. Press, Cambridge, 1997, 67–87. URL: https://doi.org/10.1017/CBO9781139173056.005.

[4] ALIPRANTIS, C. D., AND BORDER, K. C. *Infinite Dimensional Analysis*. Third. A Hitchhiker's Guide. Springer, Berlin, 2006, xxii+703. ISBN: 978-3-540-32696-0. URL: https://doi.org/10.1007/3-540-29587-9.

[5] ANDERSEN, L. A simple approach to the pricing of Bermudan swaptions in the multi-factor LIBOR market model. *J. Comput. Finance* 3, 2 (2000), 5–32. ISSN: 1460-1559. URL: https://doi.org/10.21314/JCF.1999.041.

[6] ANDERSEN, L., AND BROADIE, M. Primal-Dual Simulation Algorithm for Pricing Multidimensional American Options. *Management Sci.* 50, 9 (2004), 1222–1234. ISSN: 0025-1909. URL: https://doi.org/10.1287/mnsc.1040.0258.

[7] ANDERSSON, A., KOVÁCS, M., AND LARSSON, S. Weak error analysis for semi-linear stochastic Volterra equations with additive noise. *J. Math. Anal. Appl.* 437, 2 (2016), 1283–1304. ISSN: 0022-247X. URL: https://doi.org/10.1016/j.jmaa.2015.09.016.

[8] ANDERSSON, A., AND LARSSON, S. Weak convergence for a spatial approximation of the nonlinear stochastic heat equation. *Math. Comp.* 85, 299 (2016), 1335–1358. ISSN: 0025-5718. URL: https://doi.org/10.1090/mcom/3016.

[9] ANTON, R., COHEN, D., LARSSON, S., AND WANG, X. Full discretization of semi-linear stochastic wave equations driven by multiplicative noise. *SIAM J. Numer. Anal.* 54, 2 (2016), 1093–1119. ISSN: 0036-1429. URL: https://doi.org/10.1137/15M101049X.

[10] Arora, S., Du, S. S., Hu, W., Li, Z., and Wang, R. Fine-Grained Analysis of Optimization and Generalization for Overparameterized Two-Layer Neural Networks. In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Chaudhuri, K., and Salakhutdinov, R. Vol. 97. Proceedings of Machine Learning Research. PMLR, Long Beach, California, USA, June 2019, 322–332. URL: http://proceedings.mlr.press/v97/arora19a.html.

[11] Bach, F. Breaking the curse of dimensionality with convex neural networks. *J. Mach. Learn. Res.* 18 (2017), Paper No. 19, 1–53. ISSN: 1532-4435. URL: http://jmlr.org/papers/v18/14-546.html.

[12] Bach, F., and Moulines, E. Non-strongly-convex smooth stochastic approximation with convergence rate $O(1/n)$. In: *Advances in Neural Information Processing Systems 26*. Ed. by Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q. Curran Associates, Inc., 2013, 773–781. URL: http://papers.nips.cc/paper/4900-non-strongly-convex-smooth-stochastic-approximation-with-convergence-rate-o1n.pdf.

[13] Bally, V., and Pagès, G. A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems. *Bernoulli* 9, 6 (2003), 1003–1049. ISSN: 1350-7265. URL: https://doi.org/10.3150/bj/1072215199.

[14] Bally, V., and Pagès, G. Error analysis of the optimal quantization algorithm for obstacle problems. *Stochastic Process. Appl.* 106, 1 (2003), 1–40. ISSN: 0304-4149. URL: https://doi.org/10.1016/S0304-4149(03)00026-7.

[15] Barraquand, J., and Martineau, D. Numerical Valuation of High Dimensional Multivariate American Securities. *The Journal of Financial and Quantitative Analysis* 30, 3 (1995), 383–405. ISSN: 0022-1090. URL: https://doi.org/10.2307/2331347.

[16] Barron, A. R. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inform. Theory* 39, 3 (1993), 930–945. ISSN: 0018-9448. URL: https://doi.org/10.1109/18.256500.

[17] Barron, A. R. Approximation and estimation bounds for artificial neural networks. *Mach. Learn.* 14 (1994), 115–133. ISSN: 0885-6125. URL: https://doi.org/10.1007/BF00993164.

[18] Bartlett, P. L., Bousquet, O., and Mendelson, S. Local Rademacher complexities. *Ann. Statist.* 33, 4 (2005), 1497–1537. ISSN: 0090-5364. URL: https://doi.org/10.1214/009053605000000282.

[19] Bayer, C., Häppölä, J., and Tempone, R. Implied stopping rules for American basket options from Markovian projection. *Quant. Finance* 19, 3 (2019), 371–390. ISSN: 1469-7688. URL: https://doi.org/10.1080/14697688.2018.1481290.

[20] Bayer, C., Tempone, R., and Wolfers, S. Pricing American Options by Exercise Rate Optimization. *ArXiv e-prints* (2018), 17 pages. arXiv: 1809.07300 [q-fin.CP].

[21] Beck, C., Becker, S., Cheridito, P., Jentzen, A., and Neufeld, A. Deep splitting method for parabolic PDEs. *ArXiv e-prints* (2019), 40 pages. arXiv: 1907.03452 [math.NA].

[22] Beck, C., Becker, S., Grohs, P., Jaafari, N., and Jentzen, A. Solving stochastic differential equations and Kolmogorov equations by means of deep learning. *ArXiv e-prints* (2018), 56 pages. arXiv: 1806.00421 [math.NA].

[23] Beck, C., E, W., and Jentzen, A. Machine Learning Approximation Algorithms for High-Dimensional Fully Nonlinear Partial Differential Equations and Second-order Backward Stochastic Differential Equations. *J. Nonlinear Sci.* 29, 4 (2019), 1563–1619. ISSN: 0938-8974. URL: https://doi.org/10.1007/s00332-018-9525-3.

[24] Beck, C., Gonon, L., Hutzenthaler, M., and Jentzen, A. On existence and uniqueness properties for solutions of stochastic fixed point equations. *ArXiv e-prints* (2019), 33 pages. arXiv: 1908.03382 [math.PR].

[25] Beck, C., Gonon, L., and Jentzen, A. Overcoming the curse of dimensionality in the numerical approximation of high-dimensional semilinear elliptic partial differential equations. *ArXiv e-prints* (2020), 50 pages. arXiv: 2003.00596 [math.PR].

[26] Beck, C., Hornung, F., Hutzenthaler, M., Jentzen, A., and Kruse, T. Overcoming the curse of dimensionality in the numerical approximation of Allen-Cahn partial differential equations via truncated full-history recursive multilevel Picard approximations. *ArXiv e-prints* (2019), 31 pages. arXiv: 1907.06729 [math.NA].

[27] Beck, C., Jentzen, A., and Kuckuck, B. Full error analysis for the training of deep neural networks. *ArXiv e-prints* (2019), 53 pages. arXiv: 1910.00121 [math.NA].

[28] Becker, S., Cheridito, P., and Jentzen, A. Deep optimal stopping. *J. Mach. Learn. Res.* 20 (2019), Paper No. 74, 1–25. ISSN: 1532-4435. URL: http://jmlr.org/papers/v20/18-232.html.

[29] Becker, S., Cheridito, P., and Jentzen, A. Pricing and hedging American-style options with deep learning. *ArXiv e-prints* (2019), 12 pages. arXiv: 1912.11060 [q-fin.CP].

[30] Becker, S., Cheridito, P., Jentzen, A., and Welti, T. Solving high-dimensional optimal stopping problems using deep learning. *ArXiv e-prints* (2019), 42 pages. arXiv: 1908.01602 [cs.CE]. Revision requested from *European J. Appl. Math.*

[31] Belak, C., Hoffmann, D., and Seifried, F. T. Branching Diffusions with Jumps and Valuation with Systemic Counterparties (2020), 29 pages. Available at SSRN: https://doi.org/10.2139/ssrn.3451280.

[32] Bellman, R. *Dynamic Programming.* Princeton Univ. Press, Princeton, N. J., 1957, xxv+342.

[33] Belomestny, D. On the rates of convergence of simulation-based optimization algorithms for optimal stopping problems. *Ann. Appl. Probab.* 21, 1 (2011), 215–239. ISSN: 1050-5164. URL: https://doi.org/10.1214/10-AAP692.

[34] Belomestny, D. Pricing Bermudan options by nonparametric regression: optimal rates of convergence for lower estimates. *Finance Stoch.* 15, 4 (2011), 655–683. issn: 0949-2984. url: https://doi.org/10.1007/s00780-010-0132-x.

[35] Belomestny, D. Solving optimal stopping problems via empirical dual optimization. *Ann. Appl. Probab.* 23, 5 (2013), 1988–2019. issn: 1050-5164. url: https://doi.org/10.1214/12-AAP892.

[36] Belomestny, D., Bender, C., and Schoenmakers, J. True upper bounds for Bermudan products via non-nested Monte Carlo. *Math. Finance* 19, 1 (2009), 53–71. issn: 0960-1627. url: https://doi.org/10.1111/j.1467-9965.2008.00357.x.

[37] Belomestny, D., Dickmann, F., and Nagapetyan, T. Pricing Bermudan options via multilevel approximation methods. *SIAM J. Financial Math.* 6, 1 (2015), 448–466. issn: 1945-497X. url: https://doi.org/10.1137/130912426.

[38] Belomestny, D., Ladkau, M., and Schoenmakers, J. Multilevel simulation based policy iteration for optimal stopping—convergence and complexity. *SIAM/ASA J. Uncertain. Quantif.* 3, 1 (2015), 460–483. issn: 2166-2525. url: https://doi.org/10.1137/140958463.

[39] Belomestny, D., Schoenmakers, J., and Dickmann, F. Multilevel dual approach for pricing American style derivatives. *Finance Stoch.* 17, 4 (2013), 717–742. issn: 0949-2984. url: https://doi.org/10.1007/s00780-013-0208-5.

[40] Bender, C., and Denk, R. A forward scheme for backward SDEs. *Stochastic Process. Appl.* 117, 12 (2007), 1793–1812. issn: 0304-4149. url: https://doi.org/10.1016/j.spa.2007.03.005.

[41] Bender, C., Kolodko, A., and Schoenmakers, J. Policy iteration for American options: overview. *Monte Carlo Methods Appl.* 12, 5-6 (2006), 347–362. issn: 0929-9629. url: https://doi.org/10.1163/156939606779329053.

[42] Bender, C., Kolodko, A., and Schoenmakers, J. Enhanced policy iteration for American options via scenario selection. *Quant. Finance* 8, 2 (2008), 135–146. issn: 1469-7688. url: https://doi.org/10.1080/14697680701253013.

[43] Bender, C., Schweizer, N., and Zhuo, J. A primal-dual algorithm for BSDEs. *Math. Finance* 27, 3 (2017), 866–901. issn: 0960-1627. url: https://doi.org/10.1111/mafi.12100.

[44] Bercu, B., and Fort, J.-C. Generic Stochastic Gradient Methods. In: *Wiley Encyclopedia of Operations Research and Management Science*. American Cancer Society, 2013, 1–8. isbn: 978-0-470-40053-1. url: https://doi.org/10.1002/9780470400531.eorms1068.

[45] Berg, J., and Nyström, K. A unified deep artificial neural network approach to partial differential equations in complex geometries. *Neurocomputing* 317 (2018), 28–41. issn: 0925-2312. url: https://doi.org/10.1016/j.neucom.2018.06.056.

[46] Bernal, F., dos Reis, G., and Smith, G. Hybrid PDE solver for data-driven problems and modern branching. *European J. Appl. Math.* 28, 6 (2017), 949–972. issn: 0956-7925. url: https://doi.org/10.1017/S0956792517000109.

[47] Berner, J., Grohs, P., and Jentzen, A. Analysis of the generalization error: Empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *ArXiv e-prints* (2018), 30 pages. arXiv: 1809.03062 [cs.LG].

[48] Berridge, S. J., and Schumacher, J. M. An irregular grid approach for pricing high-dimensional American options. *J. Comput. Appl. Math.* 222, 1 (2008), 94–111. issn: 0377-0427. url: https://doi.org/10.1016/j.cam.2007.10.045.

[49] Blum, E. K., and Li, L. K. Approximation theory and feedforward networks. *Neural Networks* 4, 4 (1991), 511–515. issn: 0893-6080. url: https://doi.org/10.1016/0893-6080(91)90047-9.

[50] Bölcskei, H., Grohs, P., Kutyniok, G., and Petersen, P. Optimal approximation with sparsely connected deep neural networks. *SIAM J. Math. Data Sci.* 1, 1 (2019), 8–45. issn: 2577-0187. url: https://doi.org/10.1137/18M118709X.

[51] Bouchard, B., Tan, X., and Warin, X. Numerical approximation of general Lipschitz BSDEs with branching processes. In: *CEMRACS 2017—Numerical methods for stochastic models: control, uncertainty quantification, mean-field*. Vol. 65. ESAIM Proc. Surveys. EDP Sci., Les Ulis, 2019, 309–329. url: https://doi.org/10.1051/proc/201965309.

[52] Bouchard, B., Tan, X., Warin, X., and Zou, Y. Numerical approximation of BSDEs using local polynomial drivers and branching processes. *Monte Carlo Methods Appl.* 23, 4 (2017), 241–263. issn: 0929-9629. url: https://doi.org/10.1515/mcma-2017-0116.

[53] Bouchard, B., and Touzi, N. Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations. *Stochastic Process. Appl.* 111, 2 (2004), 175–206. issn: 0304-4149. url: https://doi.org/10.1016/j.spa.2004.01.001.

[54] Bréhier, C.-E. Strong and weak orders in averaging for SPDEs. *Stochastic Process. Appl.* 122, 7 (2012), 2553–2593. issn: 0304-4149. url: https://doi.org/10.1016/j.spa.2012.04.007.

[55] Bréhier, C.-E. Approximation of the invariant measure with an Euler scheme for stochastic PDEs driven by space-time white noise. *Potential Anal.* 40, 1 (2014), 1–40. issn: 0926-2601. url: https://doi.org/10.1007/s11118-013-9338-9.

[56] Bréhier, C.-E., and Kopec, M. Approximation of the invariant law of SPDEs: error analysis using a Poisson equation for a full-discretization scheme. *IMA J. Numer. Anal.* 37, 3 (2017), 1375–1410. issn: 0272-4979. url: https://doi.org/10.1093/imanum/drw030.

[57] Briand, P., and Labart, C. Simulation of BSDEs by Wiener chaos expansion. *Ann. Appl. Probab.* 24, 3 (2014), 1129–1171. issn: 1050-5164. url: https://doi.org/10.1214/13-AAP943.

[58] Broadie, M., and Cao, M. Improved lower and upper bound algorithms for pricing American options by simulation. *Quant. Finance* 8, 8 (2008), 845–861. issn: 1469-7688. url: https://doi.org/10.1080/14697680701763086.

[59] Broadie, M., and Glasserman, P. Pricing American-style securities using simulation. *J. Econom. Dynam. Control* 21, 8-9 (1997). Computational financial modelling, 1323–1352. issn: 0165-1889. url: https://doi.org/10.1016/S0165-1889(97)00029-8.

[60] Broadie, M., and Glasserman, P. A stochastic mesh method for pricing high-dimensional American options. *J. Comput. Finance* 7, 4 (2004), 35–72. issn: 1460-1559. url: https://doi.org/10.21314/JCF.2004.117.

[61] Broadie, M., Glasserman, P., and Ha, Z. Pricing American options by simulation using a stochastic mesh with optimized weights. In: *Probabilistic constrained optimization*. Vol. 49. Nonconvex Optim. Appl. Kluwer Acad. Publ., Dordrecht, 2000, 26–44. url: https://doi.org/10.1007/978-1-4757-3150-7_2.

[62] Bruss, F. T., and Ferguson, T. S. Minimizing the expected rank with full information. *J. Appl. Probab.* 30, 3 (1993), 616–626. issn: 0021-9002. url: https://doi.org/10.2307/3214770.

[63] Burger, M., and Neubauer, A. Error bounds for approximation with neural networks. *J. Approx. Theory* 112, 2 (2001), 235–250. issn: 0021-9045. url: https://doi.org/10.1006/jath.2001.3613.

[64] Candès, E. J. Ridgelets: Theory and applications. PhD thesis. Stanford University, 1998.

[65] Carriere, J. F. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance Math. Econom.* 19, 1 (1996), 19–30. issn: 0167-6687. url: https://doi.org/10.1016/S0167-6687(96)00004-2.

[66] Chan-Wai-Nam, Q., Mikael, J., and Warin, X. Machine learning for semi linear PDEs. *J. Sci. Comput.* 79, 3 (2019), 1667–1712. issn: 0885-7474. url: https://doi.org/10.1007/s10915-019-00908-3.

[67] Chang, D., Liu, H., and Xiong, J. A branching particle system approximation for a class of FBSDEs. *Probab. Uncertain. Quant. Risk* 1 (2016), Paper No. 9, 1–34. issn: 2367-0126. url: https://doi.org/10.1186/s41546-016-0007-y.

[68] Chassagneux, J.-F. Linear multistep schemes for BSDEs. *SIAM J. Numer. Anal.* 52, 6 (2014), 2815–2836. issn: 0036-1429. url: https://doi.org/10.1137/120902951.

[69] Chassagneux, J.-F., and Crisan, D. Runge-Kutta schemes for backward stochastic differential equations. *Ann. Appl. Probab.* 24, 2 (2014), 679–720. issn: 1050-5164. url: https://doi.org/10.1214/13-AAP933.

[70] Chassagneux, J.-F., and Richou, A. Numerical stability analysis of the Euler scheme for BSDEs. *SIAM J. Numer. Anal.* 53, 2 (2015), 1172–1193. issn: 0036-1429. url: https://doi.org/10.1137/140977047.

[71] Chassagneux, J.-F., and Richou, A. Numerical simulation of quadratic BSDEs. *Ann. Appl. Probab.* 26, 1 (2016), 262–304. issn: 1050-5164. url: https://doi.org/10.1214/14-AAP1090.

[72] Chau, N. H., Moulines, É., Rásonyi, M., Sabanis, S., and Zhang, Y. On stochastic gradient Langevin dynamics with dependent data streams: the fully non-convex case. *ArXiv e-prints* (2019), 27 pages. arXiv: 1905.13142 [math.ST].

[73] Chen, N., and Glasserman, P. Additive and multiplicative duals for American option pricing. *Finance Stoch.* 11, 2 (2007), 153–179. issn: 0949-2984. url: https://doi.org/10.1007/s00780-006-0031-3.

[74] Chen, T., and Chen, H. Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks. *IEEE Trans. Neural Networks* 6, 4 (1995), 904–910. issn: 1941-0093. url: https://doi.org/10.1109/72.392252.

[75] Chen, Y., and Wan, J. W. L. Deep Neural Network Framework Based on Backward Stochastic Differential Equations for Pricing and Hedging American Options in High Dimensions. *ArXiv e-prints* (2019), 35 pages. arXiv: 1909.11532 [q-fin.CP].

[76] Cheridito, P., Jentzen, A., and Rossmannek, F. Efficient approximation of high-dimensional functions with deep neural networks. *ArXiv e-prints* (2019), 19 pages. arXiv: 1912.04310 [math.NA].

[77] Christensen, S. A method for pricing American options using semi-infinite linear programming. *Math. Finance* 24, 1 (2014), 156–172. issn: 0960-1627. url: https://doi.org/10.1111/j.1467-9965.2012.00523.x.

[78] Chui, C. K., Li, X., and Mhaskar, H. N. Neural networks for localized approximation. *Math. Comp.* 63, 208 (1994), 607–623. issn: 0025-5718. url: https://doi.org/10.2307/2153285.

[79] Cohen, D., Larsson, S., and Sigg, M. A trigonometric method for the linear stochastic wave equation. *SIAM J. Numer. Anal.* 51, 1 (2013), 204–222. issn: 0036-1429. url: https://doi.org/10.1137/12087030X.

[80] Cohen, D., and Quer-Sardanyons, L. A fully discrete approximation of the one-dimensional stochastic wave equation. *IMA J. Numer. Anal.* 36, 1 (2016), 400–420. issn: 0272-4979. url: https://doi.org/10.1093/imanum/drv006.

[81] Company, R., Egorova, V., Jódar, L., and Soleymani, F. Computing stable numerical solutions for multidimensional American option pricing problems: a semi-discretization approach. *ArXiv e-prints* (2017), 16 pages. arXiv: 1701.08545 [q-fin.PR].

[82] Conus, D., Jentzen, A., and Kurniawan, R. Weak convergence rates of spectral Galerkin approximations for SPDEs with nonlinear diffusion coefficients. *Ann. Appl. Probab.* 29, 2 (2019), 653–716. issn: 1050-5164. url: https://doi.org/10.1214/17-AAP1352.

[83] Conus, D., Joseph, M., Khoshnevisan, D., and Shiu, S.-Y. Intermittency and Chaos for a Nonlinear Stochastic Wave Equation in Dimension 1. In: *Malliavin Calculus and Stochastic Analysis*. Vol. 34. Springer Proc. Math. Stat. Springer, New York, 2013, 251–279. isbn: 978-1-4614-5906-4. url: https://doi.org/10.1007/978-1-4614-5906-4_11.

[84] Conway, J. B. *A Course in Functional Analysis*. Second. Vol. 96. Graduate Texts in Mathematics. Springer-Verlag, New York, 1990, xvi+399. isbn: 0-387-97245-5.

[85] Cox, S., Hutzenthaler, M., Jentzen, A., van Neerven, J., and Welti, T. Convergence in Hölder norms with applications to Monte Carlo methods in infinite dimensions. *ArXiv e-prints* (2016), 50 pages. arXiv: 1605.00856 [math.NA]. To appear in *IMA J. Numer. Anal.*

[86] Cox, S., Jentzen, A., and Lindner, F. Weak convergence rates for temporal numerical approximations of stochastic wave equations with multiplicative noise. *ArXiv e-prints* (2019), 51 pages. arXiv: 1901.05535 [math.PR].

[87] Crisan, D., and Manolarakis, K. Probabilistic methods for semilinear partial differential equations. Applications to finance. *M2AN Math. Model. Numer. Anal.* 44, 5 (2010), 1107–1133. issn: 0764-583X. url: https://doi.org/10.1051/m2an/2010054.

[88] Crisan, D., and Manolarakis, K. Solving backward stochastic differential equations using the cubature method: application to nonlinear pricing. *SIAM J. Financial Math.* 3, 1 (2012), 534–571. issn: 1945-497X. url: https://doi.org/10.1137/090765766.

[89] Crisan, D., and Manolarakis, K. Second order discretization of backward SDEs and simulation with the cubature method. *Ann. Appl. Probab.* 24, 2 (2014), 652–678. issn: 1050-5164. url: https://doi.org/10.1214/13-AAP932.

[90] Crisan, D., Manolarakis, K., and Touzi, N. On the Monte Carlo simulation of BSDEs: an improvement on the Malliavin weights. *Stochastic Process. Appl.* 120, 7 (2010), 1133–1158. issn: 0304-4149. url: https://doi.org/10.1016/j.spa.2010.03.015.

[91] Cucker, F., and Smale, S. On the mathematical foundations of learning. *Bull. Amer. Math. Soc. (N.S.)* 39, 1 (2002), 1–49. issn: 0273-0979. url: https://doi.org/10.1090/S0273-0979-01-00923-5.

[92] Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems* 2, 4 (1989), 303–314. issn: 0932-4194. url: https://doi.org/10.1007/BF02551274.

[93] Da Prato, G., Jentzen, A., and Röckner, M. A mild Itô formula for SPDEs. *Trans. Amer. Math. Soc.* 372, 6 (2019), 3755–3807. issn: 0002-9947. url: https://doi.org/10.1090/tran/7165.

[94] Da Prato, G., and Zabczyk, J. *Stochastic Equations in Infinite Dimensions*. Vol. 44. Encyclopedia of Mathematics and its Applications. Cambridge Univ. Press, Cambridge, 1992, xviii+454. isbn: 0-521-38529-6. url: https://doi.org/10.1017/CBO9780511666223.

[95] Dalang, R. C. The Stochastic Wave Equation. In: *A Minicourse on Stochastic Partial Differential Equations*. Ed. by Khoshnevisan, D., and Rassoul-Agha, F. Vol. 1962. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2009, 39–71. isbn: 978-3-540-85993-2. url: https://doi.org/10.1007/978-3-540-85994-9_2.

[96] Davie, A. M., and Gaines, J. G. Convergence of numerical schemes for the solution of parabolic stochastic partial differential equations. *Math. Comp.* 70, 233 (2001), 121–134. issn: 0025-5718. url: https://doi.org/10.1090/S0025-5718-00-01224-2.

[97] Davis, M. H. A., and Karatzas, I. A Deterministic Approach to Optimal Stopping. In: *Probability, Statistics and Optimisation*. Wiley Ser. Probab. Math. Statist. Probab. Math. Statist. Wiley, Chichester, 1994, 455–466.

[98] de Bouard, A., and Debussche, A. Weak and strong order of convergence of a semidiscrete scheme for the stochastic nonlinear Schrödinger equation. *Appl. Math. Optim.* 54, 3 (2006), 369–399. issn: 0095-4616. url: https://doi.org/10.1007/s00245-006-0875-0.

[99] Debussche, A. Weak approximation of stochastic partial differential equations: the nonlinear case. *Math. Comp.* 80, 273 (2011), 89–117. issn: 0025-5718. url: https://doi.org/10.1090/S0025-5718-2010-02395-6.

[100] Debussche, A., and Printems, J. Weak order for the discretization of the stochastic heat equation. *Math. Comp.* 78, 266 (2009), 845–863. issn: 0025-5718. url: https://doi.org/10.1090/S0025-5718-08-02184-4.

[101] Delarue, F., and Menozzi, S. A forward-backward stochastic algorithm for quasi-linear PDEs. *Ann. Appl. Probab.* 16, 1 (2006), 140–184. issn: 1050-5164. url: https://doi.org/10.1214/105051605000000674.

[102] Dereich, S., and Kassing, S. Central limit theorems for stochastic gradient descent with averaging for stable manifolds. *ArXiv e-prints* (2019), 42 pages. arXiv: 1912.09187 [math.PR].

[103] Dereich, S., and Müller-Gronbach, T. General multilevel adaptations for stochastic approximation algorithms of Robbins-Monro and Polyak-Ruppert type. *Numer. Math.* 142, 2 (2019), 279–328. issn: 0029-599X. url: https://doi.org/10.1007/s00211-019-01024-y.

[104] Desai, V. V., Farias, V. F., and Moallemi, C. C. Pathwise Optimization for Optimal Stopping Problems. *Management Sci.* 58, 12 (2012), 2292–2308. issn: 0025-1909. url: https://doi.org/10.1287/mnsc.1120.1551.

[105] DeVore, R. A., Oskolkov, K. I., and Petrushev, P. P. Approximation by feed-forward neural networks. In: *The heritage of P. L. Chebyshev: a Festschrift in honor of the 70th birthday of T. J. Rivlin*. Vol. 4. Baltzer Science Publishers BV, Amsterdam, 1997, 261–287.

[106] Dockhorn, T. A Discussion on Solving Partial Differential Equations using Neural Networks. *ArXiv e-prints* (2019), 9 pages. arXiv: 1904.07200 [cs.LG].

[107] Douglas Jr., J., Ma, J., and Protter, P. Numerical methods for forward-backward stochastic differential equations. *Ann. Appl. Probab.* 6, 3 (1996), 940–968. issn: 1050-5164. url: https://doi.org/10.1214/aoap/1034968235.

[108] Du, S. S., Lee, J. D., Li, H., Wang, L., and Zhai, X. Gradient Descent Finds Global Minima of Deep Neural Networks. In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Chaudhuri, K., and Salakhutdinov, R. Vol. 97. Proceedings of Machine Learning Research. PMLR, Long Beach, California, USA, June 2019, 1675–1685. URL: http://proceedings.mlr.press/v97/du19c.html.

[109] Du, S. S., Zhai, X., Poczós, B., and Singh, A. Gradient Descent Provably Optimizes Over-parameterized Neural Networks. *ArXiv e-prints* (2018), 19 pages. arXiv: 1810.02054 [cs.LG]. Published as a conference paper at ICLR 2019.

[110] E, W., Han, J., and Jentzen, A. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Commun. Math. Stat.* 5, 4 (2017), 349–380. ISSN: 2194-6701. URL: https://doi.org/10.1007/s40304-017-0117-6.

[111] E, W., Hutzenthaler, M., Jentzen, A., and Kruse, T. Multilevel Picard iterations for solving smooth semilinear parabolic heat equations. *ArXiv e-prints* (2016), 19 pages. arXiv: 1607.03295 [math.NA].

[112] E, W., Hutzenthaler, M., Jentzen, A., and Kruse, T. On multilevel Picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations. *J. Sci. Comput.* 79, 3 (2019), 1534–1571. ISSN: 0885-7474. URL: https://doi.org/10.1007/s10915-018-00903-0.

[113] E, W., Ma, C., and Wang, Q. A Priori Estimates of the Population Risk for Residual Networks. *ArXiv e-prints* (2019), 19 pages. arXiv: 1903.02154 [cs.LG].

[114] E, W., Ma, C., and Wu, L. A Comparative Analysis of the Optimization and Generalization Property of Two-layer Neural Network and Random Feature Models Under Gradient Descent Dynamics. *ArXiv e-prints* (2019), 30 pages. arXiv: 1904.04326 [cs.LG]. To appear in *Sci. China Math.*

[115] E, W., Ma, C., and Wu, L. A priori estimates of the population risk for two-layer neural networks. *Commun. Math. Sci.* 17, 5 (2019), 1407–1425. ISSN: 1539-6746. URL: https://doi.org/10.4310/CMS.2019.v17.n5.a11.

[116] E, W., and Wang, Q. Exponential convergence of the deep neural network approximation for analytic functions. *Sci. China Math.* 61 (2018), 1733–1740. ISSN: 1674-7283. URL: https://doi.org/10.1007/s11425-018-9387-x.

[117] E, W., and Yu, B. The Deep Ritz Method: A Deep Learning-Based Numerical Algorithm for Solving Variational Problems. *Commun. Math. Stat.* 6, 1 (2018), 1–12. ISSN: 2194-6701. URL: https://doi.org/10.1007/s40304-018-0127-z.

[118] Egloff, D. Monte Carlo algorithms for optimal stopping and statistical learning. *Ann. Appl. Probab.* 15, 2 (2005), 1396–1432. ISSN: 1050-5164. URL: https://doi.org/10.1214/105051605000000043.

[119] Egloff, D., Kohler, M., and Todorovic, N. A dynamic look-ahead Monte Carlo algorithm for pricing Bermudan options. *Ann. Appl. Probab.* 17, 4 (2007), 1138–1171. ISSN: 1050-5164. URL: https://doi.org/10.1214/105051607000000249.

[120] Elbrächter, D., Grohs, P., Jentzen, A., and Schwab, C. DNN Expression Rate Analysis of High-dimensional PDEs: Application to Option Pricing. *ArXiv e-prints* (2018), 50 pages. arXiv: 1809.07669 [math.FA].

[121] Eldan, R., and Shamir, O. The Power of Depth for Feedforward Neural Networks. In: *29th Annual Conference on Learning Theory*. Ed. by Feldman, V., Rakhlin, A., and Shamir, O. Vol. 49. Proceedings of Machine Learning Research. PMLR, Columbia University, New York, New York, USA, June 2016, 907–940. url: http://proceedings.mlr.press/v49/eldan16.html.

[122] Ellacott, S. W. Aspects of the numerical analysis of neural networks. In: *Acta numerica, 1994*. Acta Numer. Cambridge Univ. Press, Cambridge, 1994, 145–202. url: https://doi.org/10.1017/S0962492900002439.

[123] Farahmand, A.-m., Nabi, S., and Nikovski, D. N. Deep reinforcement learning for partial differential equation control. In: *2017 American Control Conference (ACC)*. May 2017, 3120–3127. doi: 10.23919/ACC.2017.7963427.

[124] Fehrman, B., Gess, B., and Jentzen, A. Convergence rates for the stochastic gradient descent method for non-convex objective functions. *ArXiv e-prints* (2019), 59 pages. arXiv: 1904.01517 [math.NA]. Accepted in *J. Mach. Learn. Res.*

[125] Ferguson, T. S. Optimal Stopping and Applications: Chapter 1. Stopping Rule Problems. Mathematics Department, UCLA, available online at https://www.math.ucla.edu/~tom/Stopping/sr1.pdf, last accessed on April 10, 2020.

[126] Firth, N. P. High dimensional American options. PhD thesis. University of Oxford, 2005.

[127] Fujii, M., Takahashi, A., and Takahashi, M. Asymptotic Expansion as Prior Knowledge in Deep Learning Method for High dimensional BSDEs. *Asia-Pac. Financ. Markets* 26 (2019), 391–408. issn: 1387-2834. url: https://doi.org/10.1007/s10690-019-09271-7.

[128] Funahashi, K.-I. On the approximate realization of continuous mappings by neural networks. *Neural Networks* 2, 3 (1989), 183–192. issn: 0893-6080. url: https://doi.org/10.1016/0893-6080(89)90003-8.

[129] García, D. Convergence and biases of Monte Carlo estimates of American option prices using a parametric exercise rule. *J. Econom. Dynam. Control* 27, 10 (2003), 1855–1879. issn: 0165-1889. url: https://doi.org/10.1016/S0165-1889(02)00086-6.

[130] Gautschi, W. Some elementary inequalities relating to the gamma and incomplete gamma function. *J. Math. and Phys.* 38 (1959), 77–81. issn: 0097-1421. url: https://doi.org/10.1002/sapm195938177.

[131] Geiss, C., and Labart, C. Simulation of BSDEs with jumps by Wiener chaos expansion. *Stochastic Process. Appl.* 126, 7 (2016), 2123–2162. issn: 0304-4149. url: https://doi.org/10.1016/j.spa.2016.01.006.

[132] Geissert, M., Kovács, M., and Larsson, S. Rate of weak convergence of the finite element method for the stochastic heat equation with additive noise. *BIT* 49, 2 (2009), 343–356. issn: 0006-3835. url: https://doi.org/10.1007/s10543-009-0227-y.

[133] Giles, M. B. Multilevel Monte Carlo path simulation. *Oper. Res.* 56, 3 (2008), 607–617. issn: 0030-364X. url: https://doi.org/10.1287/opre.1070.0496.

[134] Giles, M. B., Jentzen, A., and Welti, T. Generalised multilevel Picard approximations. *ArXiv e-prints* (2019), 61 pages. arXiv: 1911.03188 [math.NA]. Revision requested from *IMA J. Numer. Anal.*

[135] Glasserman, P. *Monte Carlo Methods in Financial Engineering*. Vol. 53. Applications of Mathematics (New York). Stochastic Modelling and Applied Probability. Springer-Verlag, New York, 2004, xiv+596. isbn: 0-387-00451-3.

[136] Glorot, X., and Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Teh, Y. W., and Titterington, M. Vol. 9. Proceedings of Machine Learning Research. PMLR, Chia Laguna Resort, Sardinia, Italy, May 2010, 249–256. url: http://proceedings.mlr.press/v9/glorot10a.html.

[137] Gobet, E., and Labart, C. Solving BSDE with adaptive control variate. *SIAM J. Numer. Anal.* 48, 1 (2010), 257–277. issn: 0036-1429. url: https://doi.org/10.1137/090755060.

[138] Gobet, E., and Lemor, J.-P. Numerical simulation of BSDEs using empirical regression methods: theory and practice. *ArXiv e-prints* (2008), 17 pages. arXiv: 0806.4447 [math.PR].

[139] Gobet, E., Lemor, J.-P., and Warin, X. A regression-based Monte Carlo method to solve backward stochastic differential equations. *Ann. Appl. Probab.* 15, 3 (2005), 2172–2202. issn: 1050-5164. url: https://doi.org/10.1214/105051605000000412.

[140] Gobet, E., López-Salas, J. G., Turkedjiev, P., and Vázquez, C. Stratified regression Monte-Carlo scheme for semilinear PDEs and BSDEs with large scale parallelization on GPUs. *SIAM J. Sci. Comput.* 38, 6 (2016), C652–C677. issn: 1064-8275. url: https://doi.org/10.1137/16M106371X.

[141] Gobet, E., and Turkedjiev, P. Approximation of backward stochastic differential equations using Malliavin weights and least-squares regression. *Bernoulli* 22, 1 (2016), 530–562. issn: 1350-7265. url: https://doi.org/10.3150/14-BEJ667.

[142] Gobet, E., and Turkedjiev, P. Linear regression MDP scheme for discrete backward stochastic differential equations under general conditions. *Math. Comp.* 85, 299 (2016), 1359–1391. issn: 0025-5718. url: https://doi.org/10.1090/mcom/3013.

[143] Goldberg, D. A., and Chen, Y. Beating the curse of dimensionality in options pricing and optimal stopping. *ArXiv e-prints* (2018), 62 pages. arXiv: 1807.02227 [math.PR].

[144] Gonon, L., Grohs, P., Jentzen, A., Kofler, D., and Šiška, D. Uniform error estimates for artificial neural network approximations for heat equations. *ArXiv e-prints* (2019), 70 pages. arXiv: 1911.09647 [math.NA].

[145] Goodfellow, I., Bengio, Y., and Courville, A. *Deep Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2016, xxii+775. isbn: 978-0-262-03561-3.

[146] Goudenège, L., Molent, A., and Zanette, A. Variance Reduction Applied to Machine Learning for Pricing Bermudan/American Options in High Dimension. *ArXiv e-prints* (2019), 25 pages. arXiv: 1903.11275 [q-fin.CP].

[147] Gribonval, R., Kutyniok, G., Nielsen, M., and Voigtlaender, F. Approximation spaces of deep neural networks. *ArXiv e-prints* (2019), 63 pages. arXiv: 1905.01208 [math.FA].

[148] Grohs, P., Hornung, F., Jentzen, A., and von Wurstemberger, P. A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *ArXiv e-prints* (2018), 124 pages. arXiv: 1809.02362 [math.NA]. To appear in *Mem. Amer. Math. Soc.*

[149] Grohs, P., Hornung, F., Jentzen, A., and Zimmermann, P. Space-time error estimates for deep neural network approximations for differential equations. *ArXiv e-prints* (2019), 86 pages. arXiv: 1908.03833 [math.NA].

[150] Grohs, P., Jentzen, A., and Salimova, D. Deep neural network approximations for Monte Carlo algorithms. *ArXiv e-prints* (2019), 45 pages. arXiv: 1908.10828 [math.NA].

[151] Grohs, P., Perekrestenko, D., Elbrächter, D., and Bölcskei, H. Deep Neural Network Approximation Theory. *ArXiv e-prints* (2019), 60 pages. arXiv: 1901.02220 [cs.LG].

[152] Gühring, I., Kutyniok, G., and Petersen, P. Error bounds for approximations with deep ReLU neural networks in $W^{s,p}$ norms. *ArXiv e-prints* (2019), 42 pages. arXiv: 1902.07896 [math.FA]. To appear in *Anal. Appl. (Singap.)*

[153] Guliyev, N. J., and Ismailov, V. E. Approximation capability of two hidden layer feedforward neural networks with fixed weights. *Neurocomputing* 316 (2018), 262–269. issn: 0925-2312. url: https://doi.org/10.1016/j.neucom.2018.07.075.

[154] Guliyev, N. J., and Ismailov, V. E. On the approximation by single hidden layer feedforward neural networks with fixed weights. *Neural Networks* 98 (2018), 296–304. issn: 0893-6080. url: https://doi.org/10.1016/j.neunet.2017.12.007.

[155] Guyon, J., and Henry-Labordère, P. *Nonlinear Option Pricing*. Chapman & Hall/CRC Financial Mathematics Series. CRC Press, Boca Raton, FL, 2014, xxxviii+445. isbn: 978-1-4665-7033-7.

[156] Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. *A Distribution-Free Theory of Nonparametric Regression*. Springer Series in Statistics. Springer-Verlag, New York, 2002, xvi+647. isbn: 0-387-95441-4. url: https://doi.org/10.1007/b97848.

[157] Hairer, M., Hutzenthaler, M., and Jentzen, A. Loss of regularity for Kolmogorov equations. *Ann. Probab.* 43, 2 (2015), 468–527. ISSN: 0091-1798. URL: https://doi.org/10.1214/13-AOP838.

[158] Han, J., Jentzen, A., and E, W. Solving high-dimensional partial differential equations using deep learning. *Proc. Natl. Acad. Sci. USA* 115, 34 (2018), 8505–8510. ISSN: 0027-8424. URL: https://doi.org/10.1073/pnas.1718942115.

[159] Han, J., and Long, J. Convergence of the Deep BSDE Method for Coupled FBSDEs. *ArXiv e-prints* (2018), 27 pages. arXiv: 1811.01165 [math.PR].

[160] Hanin, B. Universal Function Approximation by Deep Neural Nets with Bounded Width and ReLU Activations. *Mathematics* 7, 10 (2019), 992. ISSN: 2227-7390. URL: https://doi.org/10.3390/math7100992.

[161] Hanin, B., and Rolnick, D. How to Start Training: The Effect of Initialization and Architecture. In: *Advances in Neural Information Processing Systems 31*. Ed. by Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. Curran Associates, Inc., 2018, 571–581. URL: http://papers.nips.cc/paper/7338-how-to-start-training-the-effect-of-initialization-and-architecture.pdf.

[162] Harms, P., and Müller, M. S. Weak convergence rates for stochastic evolution equations and applications to nonlinear stochastic wave, HJMM, stochastic Schrödinger and linearized stochastic Korteweg–de Vries equations. *Z. Angew. Math. Phys.* 70, 1 (2019), Paper No. 16, 28. ISSN: 0044-2275. URL: https://doi.org/10.1007/s00033-018-1060-4.

[163] Hartman, E. J., Keeler, J. D., and Kowalski, J. M. Layered Neural Networks with Gaussian Hidden Units as Universal Approximations. *Neural Comput.* 2, 2 (1990), 210–215. ISSN: 0899-7667. URL: https://doi.org/10.1162/neco.1990.2.2.210.

[164] Haugh, M. B., and Kogan, L. Pricing American options: a duality approach. *Oper. Res.* 52, 2 (2004), 258–270. ISSN: 0030-364X. URL: https://doi.org/10.1287/opre.1030.0070.

[165] Hausenblas, E. Weak approximation for semilinear stochastic evolution equations. In: *Stochastic analysis and related topics VIII*. Vol. 53. Progr. Probab. Birkhäuser, Basel, 2003, 111–128.

[166] Hausenblas, E. Weak approximation of the stochastic wave equation. *J. Comput. Appl. Math.* 235, 1 (2010), 33–58. ISSN: 0377-0427. URL: https://doi.org/10.1016/j.cam.2010.03.026.

[167] Heinrich, S. Monte Carlo complexity of global solution of integral equations. *J. Complexity* 14, 2 (1998), 151–175. ISSN: 0885-064X. URL: https://doi.org/10.1006/jcom.1998.0471.

[168] Henry-Labordère, P. Counterparty Risk Valuation: A Marked Branching Diffusion Approach. *ArXiv e-prints* (2012), 17 pages. arXiv: 1203.2369 [q-fin.CP].

[169] Henry-Labordère, P. Deep Primal-Dual Algorithm for BSDEs: Applications of Machine Learning to CVA and IM (2017), 16 pages. Available at SSRN: https://doi.org/10.2139/ssrn.3071506.

[170] Henry-Labordère, P., Oudjane, N., Tan, X., Touzi, N., and Warin, X. Branching diffusion representation of semilinear PDEs and Monte Carlo approximation. *Ann. Inst. Henri Poincaré Probab. Stat.* 55, 1 (2019), 184–210. ISSN: 0246-0203. URL: https://doi.org/10.1214/17-aihp880.

[171] Henry-Labordère, P., Tan, X., and Touzi, N. A numerical algorithm for a class of BSDEs via the branching process. *Stochastic Process. Appl.* 124, 2 (2014), 1112–1140. ISSN: 0304-4149. URL: https://doi.org/10.1016/j.spa.2013.10.005.

[172] Henry-Labordère, P., and Touzi, N. Branching diffusion representation for nonlinear Cauchyproblems and Monte Carlo approximation. *ArXiv e-prints* (2018), 29 pages. arXiv: 1801.08794 [math.PR].

[173] Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Networks* 4, 2 (1991), 251–257. ISSN: 0893-6080. URL: https://doi.org/10.1016/0893-6080(91)90009-T.

[174] Hornik, K. Some new results on neural network approximation. *Neural Networks* 6, 8 (1993), 1069–1072. ISSN: 0893-6080. URL: https://doi.org/10.1016/S0893-6080(09)80018-X.

[175] Hornik, K., Stinchcombe, M., and White, H. Multilayer feedforward networks are universal approximators. *Neural Networks* 2, 5 (1989), 359–366. ISSN: 0893-6080. URL: https://doi.org/10.1016/0893-6080(89)90020-8.

[176] Hornik, K., Stinchcombe, M., and White, H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks* 3, 5 (1990), 551–560. ISSN: 0893-6080. URL: https://doi.org/10.1016/0893-6080(90)90005-6.

[177] Huijskens, T. P., Ruijter, M. J., and Oosterlee, C. W. Efficient numerical Fourier methods for coupled forward-backward SDEs. *J. Comput. Appl. Math.* 296 (2016), 593–612. ISSN: 0377-0427. URL: https://doi.org/10.1016/j.cam.2015.10.019.

[178] Huré, C., Pham, H., and Warin, X. Some machine learning schemes for high-dimensional nonlinear PDEs. *ArXiv e-prints* (2019), 33 pages. arXiv: 1902.01599 [math.PR].

[179] Hutzenthaler, M., Jentzen, A., and Kruse, T. Overcoming the curse of dimensionality in the numerical approximation of parabolic partial differential equations with gradient-dependent nonlinearities. *ArXiv e-prints* (2019), 33 pages. arXiv: 1912.02571 [math.NA].

[180] Hutzenthaler, M., Jentzen, A., Kruse, T., and Nguyen, T. A. A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations. *SN Partial Differ. Equ. Appl.* 1 (2020), Paper No. 10, 1–34. ISSN: 2662-2963. URL: https://doi.org/10.1007/s42985-019-0006-9.

[181] Hutzenthaler, M., Jentzen, A., Kruse, T., Nguyen, T. A., and von Wurstemberger, P. Overcoming the curse of dimensionality in the numerical approximation of semilinear parabolic partial differential equations. *ArXiv e-prints* (2018), 30 pages. arXiv: 1807.01212 [math.PR].

[182] Hutzenthaler, M., Jentzen, A., and von Wurstemberger, P. Overcoming the curse of dimensionality in the approximative pricing of financial derivatives with default risks. *ArXiv e-prints* (2019), 71 pages. arXiv: 1903.05985 [math.NA]. Accepted in *Electron. J. Probab.*

[183] Hutzenthaler, M., and Kruse, T. Multi-level Picard approximations of high-dimensional semilinear parabolic differential equations with gradient-dependent nonlinearities. *ArXiv e-prints* (2017), 19 pages. arXiv: 1711.01080 [math.NA].

[184] Ioffe, S., and Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by Bach, F., and Blei, D. Vol. 37. Proceedings of Machine Learning Research. PMLR, Lille, France, July 2015, 448–456. url: http://proceedings.mlr.press/v37/ioffe15.html.

[185] Jacobe de Naurois, L., Jentzen, A., and Welti, T. Weak convergence rates for spatial spectral Galerkin approximations of semilinear stochastic wave equations with multiplicative noise. *ArXiv e-prints* (2015), 27 pages. arXiv: 1508.05168 [math.PR].

[186] Jacobe de Naurois, L., Jentzen, A., and Welti, T. Lower Bounds for Weak Approximation Errors for Spatial Spectral Galerkin Approximations of Stochastic Wave Equations. In: *Stochastic Partial Differential Equations and Related Fields*. Ed. by Eberle, A., Grothaus, M., Hoh, W., Kassmann, M., Stannat, W., and Trutnau, G. Springer International Publishing, Cham, 2018, 237–248. isbn: 978-3-319-74929-7. url: https://doi.org/10.1007/978-3-319-74929-7_13.

[187] Jacquier, A., and Oumgari, M. Deep PPDEs for rough local stochastic volatility. *ArXiv e-prints* (2019), 21 pages. arXiv: 1906.02551 [q-fin.PR].

[188] Jain, S., and Oosterlee, C. W. Pricing high-dimensional Bermudan options using the stochastic grid method. *Int. J. Comput. Math.* 89, 9 (2012), 1186–1211. issn: 0020-7160. url: https://doi.org/10.1080/00207160.2012.690035.

[189] Jamshidian, F. The duality of optimal exercise and domineering claims: a Doob-Meyer decomposition approach to the Snell envelope. *Stochastics* 79, 1-2 (2007), 27–60. issn: 1744-2508. url: https://doi.org/10.1080/17442500601051914.

[190] Jentzen, A., Kuckuck, B., Neufeld, A., and von Wurstemberger, P. Strong error analysis for stochastic gradient descent optimization algorithms. *ArXiv e-prints* (2018), 75 pages. arXiv: 1801.09324 [math.NA]. Accepted in *IMA J. Numer. Anal.*

[191] Jentzen, A., and Kurniawan, R. Weak convergence rates for Euler-type approximations of semilinear stochastic evolution equations with nonlinear diffusion coefficients. *ArXiv e-prints* (2015), 51 pages. arXiv: 1501.03539 [math.PR]. Accepted in *Found. Comput. Math.*

[192] Jentzen, A., Mazzonetto, S., and Salimova, D. Existence and uniqueness properties for solutions of a class of Banach space valued evolution equations. *ArXiv e-prints* (2018), 28 pages. arXiv: 1812.06859 [math.CA].

[193] Jentzen, A., Salimova, D., and Welti, T. Strong convergence for explicit space-time discrete numerical approximation methods for stochastic Burgers equations. *ArXiv e-prints* (2017), 60 pages. arXiv: 1710.07123 [math.NA]. Published in *J. Math. Anal. Appl.*

[194] Jentzen, A., Salimova, D., and Welti, T. A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients. *ArXiv e-prints* (2018), 48 pages. arXiv: 1809.07321 [math.NA]. Accepted in *Commun. Math. Sci.*

[195] Jentzen, A., and von Wurstemberger, P. Lower error bounds for the stochastic gradient descent optimization algorithm: Sharp convergence rates for slowly and fast decaying learning rates. *J. Complexity* 57 (2020), 101438. issn: 0885-064X. url: https://doi.org/10.1016/j.jco.2019.101438.

[196] Jentzen, A., and Welti, T. Overall error analysis for the training of deep neural networks via stochastic gradient descent with random initialisation. *ArXiv e-prints* (2020), 51 pages. arXiv: 2003.01291 [math.ST].

[197] Jiang, D. R., and Powell, W. B. An approximate dynamic programming algorithm for monotone value functions. *Oper. Res.* 63, 6 (2015), 1489–1511. issn: 0030-364X. url: https://doi.org/10.1287/opre.2015.1425.

[198] Kallsen, J. Option Pricing. In: *Handbook of Financial Time Series*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, 599–613. isbn: 978-3-540-71297-8. url: https://doi.org/10.1007/978-3-540-71297-8_26.

[199] Karatzas, I., and Shreve, S. E. *Brownian Motion and Stochastic Calculus*. Second. Vol. 113. Graduate Texts in Mathematics. Springer-Verlag, New York, 1991, xxiv+470. isbn: 0-387-97655-8. url: https://doi.org/10.1007/978-1-4612-0949-2.

[200] Karimi, B., Miasojedow, B., Moulines, E., and Wai, H.-T. Non-asymptotic Analysis of Biased Stochastic Approximation Scheme. In: *Proceedings of the Thirty-Second Conference on Learning Theory*. Ed. by Beygelzimer, A., and Hsu, D. Vol. 99. Proceedings of Machine Learning Research. PMLR, Phoenix, USA, June 2019, 1944–1974. url: http://proceedings.mlr.press/v99/karimi19a.html.

[201] Kingma, D. P., and Ba, J. Adam: A Method for Stochastic Optimization. *ArXiv e-prints* (2014), 15 pages. arXiv: 1412.6980 [cs.LG]. Published as a conference paper at ICLR 2015.

[202] Klenke, A. *Probability theory*. Second. Universitext. A comprehensive course. Springer, London, 2014, xii+638. isbn: 978-1-4471-5361-0. url: https://doi.org/10.1007/978-1-4471-5361-0.

[203] Kloeden, P. E., and Platen, E. *Numerical Solution of Stochastic Differential Equations*. Vol. 23. Applications of Mathematics (New York). Springer-Verlag, Berlin, 1992, xxxvi+632. isbn: 3-540-54062-8. url: https://doi.org/10.1007/978-3-662-12616-5.

[204] Kohler, M. A regression-based smoothing spline Monte Carlo algorithm for pricing American options in discrete time. *AStA Adv. Stat. Anal.* 92, 2 (2008), 153–178. issn: 1863-8171. url: https://doi.org/10.1007/s10182-008-0067-0.

[205] Kohler, M. A Review on Regression-based Monte Carlo Methods for Pricing American Options. In: *Recent Developments in Applied Probability and Statistics*. Physica, Heidelberg, 2010, 37–58. url: https://doi.org/10.1007/978-3-7908-2598-5_2.

[206] Kohler, M., and Krzyżak, A. Pricing of American options in discrete time using least squares estimates with complexity penalties. *J. Statist. Plann. Inference* 142, 8 (2012), 2289–2307. issn: 0378-3758. url: https://doi.org/10.1016/j.jspi.2012.02.031.

[207] Kohler, M., Krzyżak, A., and Todorovic, N. Pricing of high-dimensional American options by neural networks. *Math. Finance* 20, 3 (2010), 383–410. issn: 0960-1627. url: https://doi.org/10.1111/j.1467-9965.2010.00404.x.

[208] Kohler, M., Krzyżak, A., and Walk, H. Upper bounds for Bermudan options on Markovian data using nonparametric regression and a reduced number of nested Monte Carlo steps. *Statist. Decisions* 26, 4 (2008), 275–288. issn: 0721-2631. url: https://doi.org/10.1524/stnd.2008.1014.

[209] Kolodko, A., and Schoenmakers, J. Iterative construction of the optimal Bermudan stopping time. *Finance Stoch.* 10, 1 (2006), 27–49. issn: 0949-2984. url: https://doi.org/10.1007/s00780-005-0168-5.

[210] Kopec, M. Quelques contributions à l'analyse numérique d'équations stochastiques. PhD thesis. Ecole normale supérieure de Rennes - ENS Rennes, 2014.

[211] Kovács, M., Larsson, S., and Lindgren, F. Weak convergence of finite element approximations of linear stochastic evolution equations with additive noise. *BIT* 52, 1 (2012), 85–108. issn: 0006-3835. url: https://doi.org/10.1007/s10543-011-0344-2.

[212] Kovács, M., Larsson, S., and Lindgren, F. Weak convergence of finite element approximations of linear stochastic evolution equations with additive noise II. Fully discrete schemes. *BIT* 53, 2 (2013), 497–525. issn: 0006-3835. url: https://doi.org/10.1007/s10543-012-0405-1.

[213] Kovács, M., Larsson, S., and Saedpanah, F. Finite element approximation of the linear stochastic wave equation with additive noise. *SIAM J. Numer. Anal.* 48, 2 (2010), 408–427. issn: 0036-1429. url: https://doi.org/10.1137/090772241.

[214] Kovács, M., Lindner, F., and Schilling, R. L. Weak convergence of finite element approximations of linear stochastic evolution equations with additive Lévy noise. *SIAM/ASA J. Uncertain. Quantif.* 3, 1 (2015), 1159–1199. issn: 2166-2525. url: https://doi.org/10.1137/15M1009792.

[215] Kovács, M., and Printems, J. Weak convergence of a fully discrete approximation of a linear stochastic evolution equation with a positive-type memory term. *J. Math. Anal. Appl.* 413, 2 (2014), 939–952. issn: 0022-247X. url: https://doi.org/10.1016/j.jmaa.2013.12.034.

[216] Krylov, N. V. On Kolmogorov's equations for finite-dimensional diffusions. In: *Stochastic PDE's and Kolmogorov equations in infinite dimensions (Cetraro, 1998)*. Vol. 1715. Lecture Notes in Math. Springer, Berlin, 1999, 1–63. url: https://doi.org/10.1007/BFb0092417.

[217] Kulikov, A. V., and Gusyatnikov, P. P. Stopping Times for Fractional Brownian Motion. In: *Computational Management Science: State of the Art 2014*. Ed. by Fonseca, R. J., Weber, G.-W., and Telhada, J. Vol. 682. Lecture Notes in Economics and Mathematical Systems. Springer International Publishing, Cham, 2016, 195–200. isbn: 978-3-319-20430-7. url: https://doi.org/10.1007/978-3-319-20430-7_25.

[218] Kutyniok, G., Petersen, P., Raslan, M., and Schneider, R. A Theoretical Analysis of Deep Neural Networks and Parametric PDEs. *ArXiv e-prints* (2019), 43 pages. arXiv: 1904.00377 [math.NA].

[219] Labart, C., and Lelong, J. A Parallel Algorithm for solving BSDEs - Application to the pricing and hedging of American options. *ArXiv e-prints* (2011), 25 pages. arXiv: 1102.4666 [math.PR]. Published in *Monte Carlo Methods Appl.*

[220] Lamberton, D., and Lapeyre, B. *Introduction to Stochastic Calculus Applied to Finance*. Second. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2008, 253. isbn: 978-1-58488-626-6.

[221] Lapeyre, B., and Lelong, J. Neural network regression for Bermudan option pricing. *ArXiv e-prints* (2019), 23 pages. arXiv: 1907.06474 [math.PR].

[222] Lei, Y., Hu, T., Li, G., and Tang, K. Stochastic Gradient Descent for Nonconvex Learning without Bounded Gradient Assumptions. *ArXiv e-prints* (2019), 7 pages. arXiv: 1902.00908 [cs.LG]. To appear in *IEEE Trans. Neural Netw. Learn. Syst.*

[223] Lelong, J. Dual pricing of American options by Wiener chaos expansion. *SIAM J. Financial Math.* 9, 2 (2018), 493–519. issn: 1945-497X. url: https://doi.org/10.1137/16M1102161.

[224] Lelong, J. Pricing path-dependent Bermudan options using Wiener chaos expansion: an embarrassingly parallel approach. *ArXiv e-prints* (2019), 23 pages. arXiv: 1901.05672 [q-fin.CP].

[225] Lemor, J.-P., Gobet, E., and Warin, X. Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli* 12, 5 (2006), 889–916. issn: 1350-7265. url: https://doi.org/10.3150/bj/1161614951.

[226] Leshno, M., Lin, V. Y., Pinkus, A., and Schocken, S. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks* 6, 6 (1993), 861–867. issn: 0893-6080. url: https://doi.org/10.1016/S0893-6080(05)80131-5.

[227] Lindgren, F. On weak and strong convergence of numerical approximations of stochastic partial differential equations. PhD thesis. Chalmers University of Technology, 2012. isbn: 978-91-7385-787-1.

[228] Lindner, F., and Schilling, R. L. Weak order for the discretization of the stochastic heat equation driven by impulsive noise. *Potential Anal.* 38, 2 (2013), 345–379. issn: 0926-2601. url: https://doi.org/10.1007/s11118-012-9276-y.

[229] Lionnet, A., dos Reis, G., and Szpruch, L. Time discretization of FBSDE with polynomial growth drivers and reaction-diffusion PDEs. *Ann. Appl. Probab.* 25, 5 (2015), 2563–2625. issn: 1050-5164. url: https://doi.org/10.1214/14-AAP1056.

[230] Liu, W., and Röckner, M. *Stochastic Partial Differential Equations: An Introduction*. Universitext. Springer, Cham, 2015, vi+266. isbn: 978-3-319-22354-4. url: https://doi.org/10.1007/978-3-319-22354-4.

[231] Long, Z., Lu, Y., Ma, X., and Dong, B. PDE-Net: Learning PDEs from Data. In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Dy, J., and Krause, A. Vol. 80. Proceedings of Machine Learning Research. PMLR, July 2018, 3208–3216. url: http://proceedings.mlr.press/v80/long18a.html.

[232] Longstaff, F. A., and Schwartz, E. S. Valuing American Options by Simulation: A Simple Least-Squares Approach. *Rev. Financ. Stud.* 14, 1 (2001), 113–147. issn: 0893-9454. url: https://doi.org/10.1093/rfs/14.1.113.

[233] Lord, R., Fang, F., Bervoets, F., and Oosterlee, C. W. A fast and accurate FFT-based method for pricing early-exercise options under Lévy processes. *SIAM J. Sci. Comput.* 30, 4 (2008), 1678–1705. issn: 1064-8275. url: https://doi.org/10.1137/070683878.

[234] Lye, K. O., Mishra, S., and Ray, D. Deep learning observables in computational fluid dynamics. *J. Comput. Phys.* 410 (2020), 109339. issn: 0021-9991. url: https://doi.org/10.1016/j.jcp.2020.109339.

[235] Ma, J., Protter, P., San Martín, J., and Torres, S. Numerical method for backward stochastic differential equations. *Ann. Appl. Probab.* 12, 1 (2002), 302–316. issn: 1050-5164. url: https://doi.org/10.1214/aoap/1015961165.

[236] Ma, J., Protter, P., and Yong, J. Solving forward-backward stochastic differential equations explicitly—a four step scheme. *Probab. Theory Related Fields* 98, 3 (1994), 339–359. issn: 0178-8051. url: https://doi.org/10.1007/BF01192258.

[237] Ma, J., and Yong, J. *Forward-Backward Stochastic Differential Equations and their Applications*. Vol. 1702. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1999, xiv+270. isbn: 3-540-65960-9.

[238] Magill, M., Qureshi, F., and de Haan, H. Neural Networks Trained to Solve Differential Equations Learn General Representations. In: *Advances in Neural Information Processing Systems 31*. Ed. by Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. Curran Associates, Inc., 2018, 4071–4081. url: http://papers.nips.cc/paper/7662-neural-

networks-trained-to-solve-differential-equations-learn-general-representations.pdf.

[239] Maruyama, G. Continuous Markov processes and stochastic equations. *Rend. Circ. Mat. Palermo (2)* 4 (1955), 48–90. ISSN: 0009-725X. URL: https://doi.org/10.1007/BF02846028.

[240] Massart, P. *Concentration Inequalities and Model Selection.* Vol. 1896. Lecture Notes in Mathematics. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. Springer, Berlin, 2007, xiv+337. ISBN: 978-3-540-48497-4. URL: https://doi.org/10.1007/978-3-540-48503-2.

[241] McKean, H. P. Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov. *Comm. Pure Appl. Math.* 28, 3 (1975), 323–331. ISSN: 0010-3640. URL: https://doi.org/10.1002/cpa.3160280302.

[242] Meier, M., and Sögner, L. A new strategy for Robbins' problem of optimal stopping. *J. Appl. Probab.* 54, 1 (2017), 331–336. ISSN: 0021-9002. URL: https://doi.org/10.1017/jpr.2016.103.

[243] Mhaskar, H. N. Neural Networks for Optimal Approximation of Smooth and Analytic Functions. *Neural Comput.* 8, 1 (1996), 164–177. ISSN: 0899-7667. URL: https://doi.org/10.1162/neco.1996.8.1.164.

[244] Mhaskar, H. N., and Micchelli, C. A. Degree of approximation by neural and translation networks with a single hidden layer. *Adv. in Appl. Math.* 16, 2 (1995), 151–183. ISSN: 0196-8858. URL: https://doi.org/10.1006/aama.1995.1008.

[245] Mhaskar, H. N., and Poggio, T. Deep vs. shallow networks: an approximation theory perspective. *Anal. Appl. (Singap.)* 14, 6 (2016), 829–848. ISSN: 0219-5305. URL: https://doi.org/10.1142/S0219530516400042.

[246] Milstein, G. N. *Numerical Integration of Stochastic Differential Equations.* Vol. 313. Mathematics and its Applications. Translated and revised from the 1988 Russian original. Kluwer Academic Publishers Group, Dordrecht, 1995, viii+169. ISBN: 0-7923-3213-X. URL: https://doi.org/10.1007/978-94-015-8455-5.

[247] Milstein, G. N., and Tretyakov, M. V. Numerical algorithms for forward-backward stochastic differential equations. *SIAM J. Sci. Comput.* 28, 2 (2006), 561–582. ISSN: 1064-8275. URL: https://doi.org/10.1137/040614426.

[248] Milstein, G. N., and Tretyakov, M. V. Discretization of forward-backward stochastic differential equations and related quasi-linear parabolic equations. *IMA J. Numer. Anal.* 27, 1 (2007), 24–44. ISSN: 0272-4979. URL: https://doi.org/10.1093/imanum/drl019.

[249] Montanelli, H., and Du, Q. New error bounds for deep ReLU networks using sparse grids. *SIAM J. Math. Data Sci.* 1, 1 (2019), 78–92. ISSN: 2577-0187. URL: https://doi.org/10.1137/18M1189336.

[250] Müller-Gronbach, T., and Ritter, K. Lower bounds and nonuniform time discretization for approximation of stochastic heat equations. *Found. Comput. Math.* 7, 2 (2007), 135–181. ISSN: 1615-3375. URL: https://doi.org/10.1007/s10208-005-0166-6.

[251] Müller-Gronbach, T., and Ritter, K., and Wagner, T. Optimal pointwise approximation of infinite-dimensional Ornstein-Uhlenbeck processes. *Stoch. Dyn.* 8, 3 (2008), 519–541. ISSN: 0219-4937. URL: https://doi.org/10.1142/S0219493708002433.

[252] Nguyen-Thien, T., and Tran-Cong, T. Approximation of functions and their derivatives: A neural network implementation with applications. *Appl. Math. Model.* 23, 9 (1999), 687–704. ISSN: 0307-904X. URL: https://doi.org/10.1016/S0307-904X(99)00006-2.

[253] Novak, E., and Woźniakowski, H. *Tractability of Multivariate Problems. Vol. 1: Linear Information.* Vol. 6. EMS Tracts in Mathematics. European Mathematical Society (EMS), Zürich, 2008, xii+384. ISBN: 978-3-03719-026-5. URL: https://doi.org/10.4171/026.

[254] Novak, E., and Woźniakowski, H. *Tractability of Multivariate Problems. Volume II: Standard Information for Functionals.* Vol. 12. EMS Tracts in Mathematics. European Mathematical Society (EMS), Zürich, 2010, xviii+657. ISBN: 978-3-03719-084-5. URL: https://doi.org/10.4171/084.

[255] Park, J., and Sandberg, I. W. Universal Approximation Using Radial-Basis-Function Networks. *Neural Comput.* 3, 2 (1991), 246–257. ISSN: 0899-7667. URL: https://doi.org/10.1162/neco.1991.3.2.246.

[256] Perekrestenko, D., Grohs, P., Elbrächter, D., and Bölcskei, H. The universal approximation power of finite-width deep ReLU networks. *ArXiv e-prints* (2018), 16 pages. arXiv: 1806.01528 [cs.LG].

[257] Peskir, G., and Shiryaev, A. *Optimal Stopping and Free-Boundary Problems.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2006, xxii+500. ISBN: 978-3-7643-2419-3.

[258] Petersen, P., Raslan, M., and Voigtlaender, F. Topological properties of the set of functions generated by neural networks of fixed size. *ArXiv e-prints* (2018), 51 pages. arXiv: 1806.08459 [math.GN]. To appear in *Found. Comput. Math.*

[259] Petersen, P., and Voigtlaender, F. Optimal approximation of piecewise smooth functions using deep ReLU neural networks. *Neural Networks* 108 (2018), 296–330. ISSN: 0893-6080. URL: https://doi.org/10.1016/j.neunet.2018.08.019.

[260] Petersen, P., and Voigtlaender, F. Equivalence of approximation by convolutional neural networks and fully-connected networks. *Proc. Amer. Math. Soc.* 148, 4 (2020), 1567–1581. ISSN: 0002-9939. URL: https://doi.org/10.1090/proc/14789.

[261] Pham, H. Feynman-Kac representation of fully nonlinear PDEs and applications. *Acta Math. Vietnam.* 40, 2 (2015), 255–269. ISSN: 0251-4184. URL: https://doi.org/10.1007/s40306-015-0128-x.

[262] Pham, H., and Warin, X. Neural networks-based backward scheme for fully nonlinear PDEs. *ArXiv e-prints* (2019), 15 pages. arXiv: 1908.00412 [math.OC].

[263] Pinkus, A. Approximation theory of the MLP model in neural networks. In: *Acta numerica, 1999*. Vol. 8. Acta Numer. Cambridge Univ. Press, Cambridge, 1999, 143–195. URL: https://doi.org/10.1017/S0962492900002919.

[264] Qi, F. Bounds for the ratio of two gamma functions. *J. Inequal. Appl.* (2010), Art. ID 493058, 84. ISSN: 1025-5834. URL: https://doi.org/10.1155/2010/493058.

[265] Quer-Sardanyons, L., and Sanz-Solé, M. Space semi-discretisations for a stochastic wave equation. *Potential Anal.* 24, 4 (2006), 303–332. ISSN: 0926-2601. URL: https://doi.org/10.1007/s11118-005-9002-0.

[266] Raissi, M. Deep Hidden Physics Models: Deep Learning of Nonlinear Partial Differential Equations. *J. Mach. Learn. Res.* 19 (2018), Paper No. 25, 1–24. ISSN: 1532-4435. URL: http://jmlr.org/papers/v19/18-046.html.

[267] Rasulov, A., Raimova, G., and Mascagni, M. Monte Carlo solution of Cauchy problem for a nonlinear parabolic equation. *Math. Comput. Simulation* 80, 6 (2010), 1118–1123. ISSN: 0378-4754. URL: https://doi.org/10.1016/j.matcom.2009.12.009.

[268] Reisinger, C., and Zhang, Y. Rectified deep neural networks overcome the curse of dimensionality for nonsmooth value functions in zero-sum games of nonlinear stiff systems. *ArXiv e-prints* (2019), 34 pages. arXiv: 1903.06652 [math.NA].

[269] Robbins, H., and Monro, S. A stochastic approximation method. *Ann. Math. Statistics* 22 (1951), 400–407. ISSN: 0003-4851. URL: https://doi.org/10.1214/aoms/1177729586.

[270] Rogers, L. C. G. Monte Carlo valuation of American options. *Math. Finance* 12, 3 (2002), 271–286. ISSN: 0960-1627. URL: https://doi.org/10.1111/1467-9965.02010.

[271] Rogers, L. C. G. Dual valuation and hedging of Bermudan options. *SIAM J. Financial Math.* 1, 1 (2010), 604–608. ISSN: 1945-497X. URL: https://doi.org/10.1137/090772198.

[272] Ruijter, M. J., and Oosterlee, C. W. A Fourier cosine method for an efficient computation of solutions to BSDEs. *SIAM J. Sci. Comput.* 37, 2 (2015), A859–A889. ISSN: 1064-8275. URL: https://doi.org/10.1137/130913183.

[273] Ruijter, M. J., and Oosterlee, C. W. Numerical Fourier method and second-order Taylor scheme for backward SDEs in finance. *Appl. Numer. Math.* 103 (2016), 1–26. ISSN: 0168-9274. URL: https://doi.org/10.1016/j.apnum.2015.12.003.

[274] Ruszczynski, A., and Yao, J. A Dual Method For Backward Stochastic Differential Equations with Application to Risk Valuation. *ArXiv e-prints* (2017), 23 pages. arXiv: 1701.06234 [math.OC].

[275] Schmitt, M. Lower Bounds on the Complexity of Approximating Continuous Functions by Sigmoidal Neural Networks. In: *Advances in Neural Information Processing Systems 12*. Ed. by Solla, S. A., Leen, T. K., and Müller, K. MIT Press, 2000, 328–334. URL: http://papers.nips.cc/paper/1692-lower-bounds-on-the-complexity-of-approximating-continuous-functions-by-sigmoidal-neural-networks.pdf.

[276] Schoenmakers, J., and Zhang, J., and Huang, J. Optimal dual martingales, their analysis, and application to new algorithms for Bermudan products. *SIAM J. Financial Math.* 4, 1 (2013), 86–116. issn: 1945-497X. url: https://doi.org/10.1137/110832513.

[277] Schwab, C., and Zech, J. Deep learning in high dimension: neural network expression rates for generalized polynomial chaos expansions in UQ. *Anal. Appl. (Singap.)* 17, 1 (2019), 19–55. issn: 0219-5305. url: https://doi.org/10.1142/S0219530518500203.

[278] Schweizer, M. On Bermudan options. In: *Advances in Finance and Stochastics.* Springer, Berlin, 2002, 257–270. isbn: 978-3-662-04790-3. url: https://doi.org/10.1007/978-3-662-04790-3_13.

[279] Sell, G. R., and You, Y. *Dynamics of Evolutionary Equations.* Vol. 143. Applied Mathematical Sciences. Springer-Verlag, New York, 2002, xiv+670. isbn: 0-387-98347-3. url: https://doi.org/10.1007/978-1-4757-5037-9.

[280] Shaham, U., Cloninger, A., and Coifman, R. R. Provable approximation properties for deep neural networks. *Appl. Comput. Harmon. Anal.* 44, 3 (2018), 537–557. issn: 1063-5203. url: https://doi.org/10.1016/j.acha.2016.04.003.

[281] Shalev-Shwartz, S., and Ben-David, S. *Understanding Machine Learning: From Theory to Algorithms.* Cambridge Univ. Press, Cambridge, 2014. isbn: 978-1-107-29801-9. url: https://doi.org/10.1017/CBO9781107298019.

[282] Shamir, O. Exponential Convergence Time of Gradient Descent for One-Dimensional Deep Linear Neural Networks. In: *Proceedings of the Thirty-Second Conference on Learning Theory.* Ed. by Beygelzimer, A., and Hsu, D. Vol. 99. Proceedings of Machine Learning Research. PMLR, Phoenix, USA, June 2019, 2691–2713. url: http://proceedings.mlr.press/v99/shamir19a.html.

[283] Shardlow, T. Weak convergence of a numerical method for a stochastic heat equation. *BIT* 43, 1 (2003), 179–193. issn: 0006-3835. url: https://doi.org/10.1023/A:1023661308243.

[284] Shen, Z., Yang, H., and Zhang, S. Deep Network Approximation Characterized by Number of Neurons. *ArXiv e-prints* (2019), 36 pages. arXiv: 1906.05497 [math.NA].

[285] Shen, Z., Yang, H., and Zhang, S. Nonlinear approximation via compositions. *Neural Networks* 119 (2019), 74–84. issn: 0893-6080. url: https://doi.org/10.1016/j.neunet.2019.07.011.

[286] Shreve, S. E. *Stochastic Calculus for Finance II.* Springer Finance. Continuous-Time Models. Springer-Verlag, New York, 2004, xx+550. isbn: 0-387-40101-6.

[287] Sirignano, J., and Spiliopoulos, K. Stochastic gradient descent in continuous time. *SIAM J. Financial Math.* 8, 1 (2017), 933–961. issn: 1945-497X. url: https://doi.org/10.1137/17M1126825.

[288] Sirignano, J., and Spiliopoulos, K. DGM: A deep learning algorithm for solving partial differential equations. *J. Comput. Phys.* 375 (2018), 1339–1364. issn: 0021-9991. url: https://doi.org/10.1016/j.jcp.2018.08.029.

[289] Skorohod, A. V. Branching diffusion processes. *Teor. Verojatnost. i Primenen.* 9 (1964), 492–497. issn: 0040-361x.

[290] Smirnov, M. *Javascript Options and Implied Volatility Calculator*. http://www.math.columbia.edu/~smirnov/options13.html. Columbia University. Last accessed on April 10, 2020.

[291] Solan, E., Tsirelson, B., and Vieille, N. Random Stopping Times in Stopping Problems and Stopping Games. *ArXiv e-prints* (2012), 21 pages. arXiv: 1211.5802 [math.PR].

[292] Thomas, L. E. Persistent energy flow for a stochastic wave equation model in nonequilibrium statistical mechanics. *J. Math. Phys.* 53, 9 (2012), 095208, 10. issn: 0022-2488. url: https://doi.org/10.1063/1.4728986.

[293] Tilley, J. A. Valuing American options in a path simulation model. *Transactions of the Society of Actuaries* 45 (1993), 83–104.

[294] Tsitsiklis, J. N., and Van Roy, B. Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives. *IEEE Trans. Automat. Control* 44, 10 (1999), 1840–1851. issn: 0018-9286. url: https://doi.org/10.1109/9.793723.

[295] Tsitsiklis, J. N., and Van Roy, B. Regression methods for pricing complex American-style options. *IEEE Trans. Neural Networks* 12, 4 (2001), 694–703. issn: 1941-0093. url: https://doi.org/10.1109/72.935083.

[296] Turkedjiev, P. Two algorithms for the discrete time approximation of Markovian backward stochastic differential equations under local conditions. *Electron. J. Probab.* 20 (2015), Paper No. 50, 1–49. issn: 1083-6489. url: https://doi.org/10.1214/EJP.v20-3022.

[297] van de Geer, S. A. *Applications of Empirical Process Theory*. Vol. 6. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge Univ. Press, Cambridge, 2000, xii+286. isbn: 0-521-65002-X.

[298] Voigtlaender, F., and Petersen, P. Approximation in $L^p(\mu)$ with deep ReLU neural networks. In: *13th International conference on Sampling Theory and Applications (SampTA)*. IEEE, Bordeaux, France, July 2019, 1–4. url: https://doi.org/10.1109/SampTA45681.2019.9030992.

[299] Walsh, J. B. An introduction to stochastic partial differential equations. In: *École d'été de probabilités de Saint-Flour, XIV—1984*. Vol. 1180. Lecture Notes in Math. Springer, Berlin, 1986, 265–439. url: https://doi.org/10.1007/BFb0074920.

[300] Walsh, J. B. On numerical solutions of the stochastic wave equation. *Illinois J. Math.* 50, 1-4 (2006), 991–1018. issn: 0019-2082. url: https://doi.org/10.1215/ijm/1258059497.

[301] Wang, H., Chen, H., Sudjianto, A., Liu, R., and Shen, Q. Deep Learning-Based BSDE Solver for Libor Market Model with Application to Bermudan Swaption Pricing and Hedging. *ArXiv e-prints* (2018), 36 pages. arXiv: 1807.06622 [q-fin.CP].

[302] Wang, X. An exponential integrator scheme for time discretization of nonlinear stochastic wave equation. *J. Sci. Comput.* 64, 1 (2015), 234–263. issn: 0885-7474. url: https://doi.org/10.1007/s10915-014-9931-0.

[303] Wang, X. Weak error estimates of the exponential Euler scheme for semi-linear SPDEs without Malliavin calculus. *Discrete Contin. Dyn. Syst.* 36, 1 (2016), 481–497. issn: 1078-0947. url: https://doi.org/10.3934/dcds.2016.36.481.

[304] Wang, X., and Gan, S. Weak convergence analysis of the linear implicit Euler method for semilinear stochastic partial differential equations with additive noise. *J. Math. Anal. Appl.* 398, 1 (2013), 151–169. issn: 0022-247X. url: https://doi.org/10.1016/j.jmaa.2012.08.038.

[305] Wang, X., Gan, S., and Tang, J. Higher order strong approximations of semilinear stochastic wave equation with additive space-time white noise. *SIAM J. Sci. Comput.* 36, 6 (2014), A2611–A2632. issn: 1064-8275. url: https://doi.org/10.1137/130937524.

[306] Warin, X. Variations on branching methods for non linear PDEs. *ArXiv e-prints* (2017), 25 pages. arXiv: 1701.07660 [math.PR].

[307] Watanabe, S. On the branching process for Brownian particles with an absorbing boundary. *J. Math. Kyoto Univ.* 4 (1965), 385–398. issn: 0023-608X. url: https://doi.org/10.1215/kjm/1250524667.

[308] Wendel, J. G. Note on the gamma function. *Amer. Math. Monthly* 55 (1948), 563–564. issn: 0002-9890. url: https://doi.org/10.2307/2304460.

[309] Yarotsky, D. Error bounds for approximations with deep ReLU networks. *Neural Networks* 94 (2017), 103–114. issn: 0893-6080. url: https://doi.org/10.1016/j.neunet.2017.07.002.

[310] Yarotsky, D. Universal approximations of invariant maps by neural networks. *ArXiv e-prints* (2018), 64 pages. arXiv: 1804.10306 [cs.NE].

[311] Zhang, G., Martens, J., and Grosse, R. B. Fast Convergence of Natural Gradient Descent for Over-Parameterized Neural Networks. In: *Advances in Neural Information Processing Systems 32*. Ed. by Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. Curran Associates, Inc., 2019, 8082–8093. url: http://papers.nips.cc/paper/9020-fast-convergence-of-natural-gradient-descent-for-over-parameterized-neural-networks.pdf.

[312] Zhang, J. A numerical scheme for BSDEs. *Ann. Appl. Probab.* 14, 1 (2004), 459–488. issn: 1050-5164. url: https://doi.org/10.1214/aoap/1075828058.

[313] Zou, D., Cao, Y., Zhou, D., and Gu, Q. Gradient descent optimizes over-parameterized deep ReLU networks. *Mach. Learn.* 109, 3 (2020), 467–492. issn: 0885-6125. url: https://doi.org/10.1007/s10994-019-05839-6.

Timo Welti

Seminar for Applied Mathematics, ETH Zürich, Switzerland
twelti@twelti.org • twelti.org • Google Scholar • arXiv

born on October 1, 1993 • citizen of Küsnacht ZH, Switzerland

RESEARCH

PhD candidate; Seminar for Applied Mathematics, ETH Zürich 2015–2020

- Advised by Prof. Dr. Arnulf Jentzen
- 9 publications (see below; abstracts on arXiv)
- Invited research talks given at research conferences in Spain, Australia, Sweden, Austria, France, the USA, and Canada; invited seminar talks given at the University of Oxford, the University of Vienna, the University of Passau, and the University of Klagenfurt
- Head Assistant for lectures on theory of machine learning/linear algebra/numerical analysis/stochastic partial differential equations/stochastic ordinary differential equations

Visiting researcher; Mathematical Institute, University of Oxford 2019

- 6 months; advised by Prof. Dr. Mike Giles

Visiting researcher; Faculty of Mathematics, University of Vienna 2018/19

- 5 months; advised by Prof. Dr. Philipp Grohs

EDUCATION

MSc in Mathematics (with distinction), ETH Zürich 2013–2015

- Semester abroad at the National University of Singapore (6 months) 2013/14
- Teaching Assistant for lectures on numerical analysis

BSc in Mathematics, ETH Zürich 2010–2013

- Teaching Assistant for lectures on numerical analysis/linear algebra

AWARDS & RECOGNITION (SELECTED)

- **GAMM Junior** for the years 2020 to 2022, elected by the *International Association of Applied Mathematics and Mechanics (GAMM)* 2019
- **Doc.Mobility fellowship** (research grant) awarded by the *Swiss National Science Foundation* 2018
- **5 travel awards** for giving research talks at conferences in Germany, Austria, France, and the USA, awarded by the *Society for Industrial and Applied Mathematics*, the *Swiss Study Foundation*, and the journal *Mathematics* 2017–2019
- Admittance to the **Swiss Study Foundation** 2014
- **Teaching Assistant Award**, ETH Zürich 2014

PUBLICATIONS

Preprints or research articles subject to revision

[P9] Jentzen, A., and Welti, T. Overall error analysis for the training of deep neural networks via stochastic gradient descent with random initialisation. *ArXiv e-prints* (2020), 51 pages. arXiv: 2003.01291 [math.ST].

[P8] Giles, M. B., Jentzen, A., and Welti, T. Generalised multilevel Picard approximations. *ArXiv e-prints* (2019), 61 pages. arXiv: 1911.03188 [math.NA]. Revision requested from *IMA J. Numer. Anal.*

[P7] Becker, S., Cheridito, P., Jentzen, A., and Welti, T. Solving high-dimensional optimal stopping problems using deep learning. *ArXiv e-prints* (2019), 42 pages. arXiv: 1908.01602 [cs.CE]. Revision requested from *European J. Appl. Math.*

Peer-reviewed published or accepted research articles

[P6] Jentzen, A., Salimova, D., and Welti, T. A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients. *ArXiv e-prints* (2018), 48 pages. arXiv: 1809.07321 [math.NA]. Accepted in *Commun. Math. Sci.*

[P5] Jentzen, A., Salimova, D., and Welti, T. Strong convergence for explicit space–time discrete numerical approximation methods for stochastic Burgers equations. *J. Math. Anal. Appl.* 469, 2 (2019), 661–704. issn: 0022-247X. url: https://doi.org/10.1016/j.jmaa.2018.09.032.

[P4] Jacobe de Naurois, L., Jentzen, A., and Welti, T. Lower Bounds for Weak Approximation Errors for Spatial Spectral Galerkin Approximations of Stochastic Wave Equations. In: *Stochastic Partial Differential Equations and Related Fields*. Ed. by Eberle, A., Grothaus, M., Hoh, W., Kassmann, M., Stannat, W., and Trutnau, G. Springer International Publishing, Cham, 2018, 237–248. isbn: 978-3-319-74929-7. url: https://doi.org/10.1007/978-3-319-74929-7_13.

[P3] Andersson, A., Jentzen, A., Kurniawan, R., and Welti, T. On the differentiability of solutions of stochastic evolution equations with respect to their initial values. *Nonlinear Anal.* 162 (2017), 128–161. issn: 0362-546X. url: https://doi.org/10.1016/j.na.2017.03.003.

[P2] Cox, S., Hutzenthaler, M., Jentzen, A., van Neerven, J., and Welti, T. Convergence in Hölder norms with applications to Monte Carlo methods in infinite dimensions. *ArXiv e-prints* (2016), 50 pages. arXiv: 1605.00856 [math.NA]. To appear in *IMA J. Numer. Anal.*

[P1] Jacobe de Naurois, L., Jentzen, A., and Welti, T. Weak convergence rates for spatial spectral Galerkin approximations of semilinear stochastic wave equations with multiplicative noise. *ArXiv e-prints* (2015), 27 pages. arXiv: 1508.05168 [math.PR].