Numerical approximation methods for semilinear partial differential equations with gradient-dependent nonlinearities

Von der Fakultät für Mathematik der Universität
Duisburg-Essen zur
Erlangung des akademischen Grades eines Doktors der
Naturwissenschaften
(Dr. rer. nat.)
genehmigte Dissertation

von

Katharina Pohl

Datum der mündlichen Prüfung:
2. Mai 2024

Erstgutachter: Prof. Dr. Martin Hutzenthaler
Zweitgutachter: Prof. Dr. Thomas Kruse

DOI: 10.17185/duepublico/81943
URN: urn:nbn:de:hbz:465-20240524-073450-1

Zusammenfassung

In dieser Doktorarbeit untersuchen wir zwei numerische Approximationsmethoden zur Bestimmung von Lösungen für eine Klasse semilinearer partieller Differentialgleichung (PDE) mit Gradienten-abhängigen Nichtlinearitäten. Das Ziel ist in beiden Fällen den sogenannten Fluch der Dimensionen zu brechen, welcher besagt, dass der Berechnungsaufwand einer Lösung exponentiell mit der Dimension der zu lösenden PDE wächst. Das erste betrachtete Verfahren ist ein Multilevel Picard (MLP) Approximationsschema, dessen theoretische Grundlage aus der Formulierung der PDE als stochastische Fixpunktgleichung (SFPE) folgt. Dafür erarbeiten wir eine angepasste Bismut-Elworthy-Li Formel, die den Transfer von PDEs zu korrespondierenden SFPEs ermöglicht. Diese SFPEs studieren wir zunächst in einer abstrakten Umgebung und zeigen dort mittels eines Banachschen Fixpunkt Arguments die Existenz und Eindeutigkeit von Lösungen. Den erfolgreichen Rücktransfer beweisen wir, indem wir zeigen, dass die SFPE Lösung - unter geeigneten Annahmen - die eindeutige Viskositätslösung der betrachteten PDE ist. Damit können wir die Lösungen von PDEs als SFPE Lösungen schreiben und begründen die Konstruktion eines Multilevel-Picard (MLP) Approximationsschemas. Wir definieren und untersuchen dieses MLP Approximationsschema und bestimmen eine obere Schranke an den Approximationsfehler. Im Fall einer glatten Lösung zeigen wir - unter geeigneten Annahmen - außerdem, dass das MLP Approximationsschema nicht unter dem "Fluch der Dimensionen" leidet. Als zweiten Ansatz zur numerischen Approximation von PDE Lösungen betrachten wir stochastische Gradientenverfahren (SGD) im Training tiefer neuronaler Netze mit Rectified Linear Unit Aktivierungsfunktion. Wir beweisen unter der Annahme einer konstanten Zielfunktion und ausreichend kleinen, aber nicht L^1-summierbaren Schrittgrößen, dass der Erwartungswert der Risikofunktionen des betrachteten SGD Prozessen im Training der neuronalen Netze gegen Null konvergiert, wenn die Anzahl der SGD Schritte gegen unendlich konvergiert.

Abstract

In this thesis we study two numerical approximation methods for the estimation of solutions of a class of semilinear partial differential equations (PDEs) with gradient-dependent nonlinearities. In both cases the main goal is to overcome the so called curse of dimensionality which means that the computational effort for the solution grows exponentially in the dimension of the PDE to be solved. The first considered method is a Multilevel Picard (MLP) approximation scheme which is based on the reformulation of the PDE as stochastic fixed point equation (SFPE). For this transfer from PDEs to the corresponding SFPEs we develop an adjusted Bismut-Elworthy-Li formula. Then we analyse SFPEs in an abstract setting and prove existence and uniqueness of solutions by using a Banach fixed point argument. We successfully re-transfer the SFPE to the considered PDE by proving that the SFPE solution is the unique viscosity solution of the PDE. Thus, we can write the PDE solutions as SFPE solutions and therefore justify the constructions of a MLP approximation scheme. We define and study this MLP approximation scheme and establish an upper bound on the approximation error. Moreover, in the setting of a smooth solution we prove - under certain assumptions - that the MLP approximation scheme does not suffer from the curse of dimensionality. As a second approach to numerically approximate PDE solutions, we consider stochastic gradient descent (SGD) type optimization methods in the training of deep neural networks (DNNs) with Rectified Linear Unit activation function. We prove that under the assumption of a constant target function and sufficiently small but not L^1-summable SGD step sizes, the expected value of the risks of the considered SGD process in the training of DNNs converges to zero if the number of SGD steps goes to infinity.

Danksagung

An dieser Stelle möchte ich mich bei all jenen bedanken, die mich auf meinem Weg zur erfolgreichen Disputation begleitet und unterstützt haben.

Zuallererst danke ich meinem Betreuer Martin Hutzenthaler, der mich bereits im ersten Masterjahr als potenzielle Doktorandin gesehen hat. Ich danke ihm für die fachlichen Impulse, sein Vertrauen in mich und die erfolgreiche Zusammenarbeit. Ein herzlicher Dank geht auch an die weiteren Mitglieder meiner Prüfungskommission für ihr tiefes Interesse und konstruktives Feedback. Zusätzlich möchte ich Arnulf Jentzen für das gemeinsame Projekt danken, das meine Begeisterung für neuronale Netze entfacht hat. Danke auch an Robert und Kai sowie den anderen Mitarbeitenden in der Stochastik für zahlreiche gemeinsame Mittagessen und die nette Arbeitsatmosphäre.

Meiner Familie und insbesondere meinen Eltern möchte ich für ihre Liebe, Unterstützung und unerschütterliche Zuversicht in mich danken - auch wenn sie nicht verstehen, was ich hier mache. Meiner Schwester danke ich außerdem für die wöchentlichen Treffen, bei denen ich abschalten konnte. Ein besonderer Dank gilt auch meinen Freund*innen, die meine Höhen und Tiefen miterlebt haben und immer ein offenes Ohr für meine Probleme hatten. Ihre moralische Unterstützung hat mir die Kraft gegeben, auch in schwierigen Zeiten weiterzumachen. Ein herzliches Dankeschön geht auch an meine Mannschaft, mit der ich gemeinsam immer einen Ausgleich zur Schreibtischarbeit finden konnte.

Und schließlich möchte ich meinem Max für seine bedingungslose Liebe, seine Unterstützung in allen Situationen und seine Ermutigung in schweren Zeiten danken. Du warst während der ganzen Zeit mein Ruhepol.

Ich bin zutiefst dankbar, von so vielen wunderbaren Menschen umgeben zu sein, die mich auf diesem Weg begleitet haben.

Contents

Introduction

In many applications in the world of natural sciences, engineering, and mathematical finance partial differential equations (PDEs) play a significant role in modelling the appearing processes. The aim to understand these processes more thoroughly leads to a strong demand in solving and understanding different kinds of PDEs. Unfortunately, there often do not exist explicit solutions to given PDEs. That is why algorithms which are able to solve PDEs numerically got more and more into focus over the last couple of years. Nevertheless, developing *efficient* algorithms to solve PDEs has been an open problem for a long time. First results were developed for PDEs appearing in applications of physics and engineering which often have dimension equal or smaller than four. In this case the PDEs can be numerically approximated by means of finite element or finite difference methods (see, e.g., [7, 44]). However, PDEs which model processes in finance are often high-dimensional. In this case finite difference and finite element methods lead to a computational effort to achieve a specific approximation accuracy which grows exponentially with the dimension (cf., e.g., [17, 83, 84, 85]). This difficulty is called the "curse of dimensionality". During the last years, there have been different approaches to create effective approximation methods which overcome this "curse of dimensionality". In the case of linear elliptic and parabolic PDEs of second order solutions can be numerically approximated by Monte Carlo sampling without suffering from the curse of dimensionality. The reason for this lies within the connection between linear elliptic and parabolic PDEs and stochastic differential equations (SDEs) given by the Itô formula (see [61, 72]) or the Feynman-Kac formula (see [42, 67]). However, many problems in the world of financial engineering lead to nonlinear PDEs. We refer, e.g., to [8, 27, 39, 43, 79, 98] for nonlinear PDEs arising from pricing models. In this thesis we focus on possible solutions to solve semilinear PDEs with gradient-dependent nonlinearities.

In the field of solving semilinear PDEs there have been different approximation methods proposed over the last few years. Three main approaches can be summarized in the following categories: approximation methods that use stochastic representation of PDE solutions by means of branching diffusion processes (see, e.g. [19, 49, 48]), full-history recursive multilevel Picard (MLP) approximation methods (see, e.g. [16, 36, 37, 52, 53, 54, 56, 58]), and the reformulation of the PDEs as stochastic learning problems which then are solved by using deep artificial neural networks (DNN) (see, e.g., [12, 9, 15, 21, 33, 35, 46, 47, 51, 75, 78, 88, 97]). The main disadvantage of the branching diffusion method is that it is only applicable for PDEs with sufficiently small terminal conditions (see, e.g., [37, Section 4.7]). For this reason, we focus in this thesis on the latter two approaches to establish stochastic approximation schemes which make it possible to efficiently approximate the solution of high-dimensional PDEs and which do not suffer from the curse of dimensionality.

In the setting of MLP approximation methods first results for the heat equation with a gradient-dependent nonlinearity have already been accomplished in [52, 54]. We extend these findings to a more general setting of semilinear PDEs of second order with gradient-dependent nonlinearities. The main procedure works similar as the central idea to achieve a connection between a certain class of PDEs and SFPEs stays the same. However, our generalized assumptions allow non-differentiable terminal conditions which, together with the Bismut-Elworthy-Li formula, cause a singularity in the endpoint of the considered SFPEs (cf. (3) and (4)). Handling this singularity requires several new techniques and strategies. We start by generalizing the above-mentioned connection between PDEs and SDEs to the case of semilinear PDEs with gradient-dependent nonlinearities in Chapter 1. For this, we first study an abstract class of SFPEs. In this setting our first main result of Chapter 1 is the general existence and uniqueness result of solutions of SFPEs in Theorem 1.1.11 below which is proved by Banach's fixed point theorem. These findings are applied in Section 1.2 to SFPEs associated with semilinear PDEs to obtain our second main result of this chapter, Theorem 1.2.5 below. In particular, we prove in Theorem 1.2.5 the existence and uniqueness of solutions of SFPEs associated with a certain class of semilinear PDEs with gradient-dependent nonlinearities. To illustrate the findings of Chapter 1 in more detail, we present in Theorem 1 below a special case of Theorem 1.2.5.

Theorem 1. Let $d \in \mathbb{N}$, $c, T \in (0,\infty)$, let $\|\cdot\| \colon \mathbb{R}^d \to [0,\infty)$, $|\!|\!|\cdot|\!|\!| \colon \mathbb{R}^{d+1} \to [0,\infty)$, $\|\cdot\|_F \colon \mathbb{R}^{d\times d} \to [0,\infty)$ be norms, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_s)_{s\in[0,T]})$ be a filtered probability space, let $W \colon [0,T]\times\Omega \to \mathbb{R}^d$ be a standard $(\mathbb{F}_s)_{s\in[0,T]}$-Brownian motion, let $\mu \in C^2(\mathbb{R}^d, \mathbb{R}^d)$, $\sigma \in C^2(\mathbb{R}^d, \mathbb{R}^{d\times d})$ satisfy for all $x, y \in \mathbb{R}^d$ that $y^*\sigma(x)(\sigma(x))^* y \geqslant \frac{1}{c}\|y\|^2$ and

$$\max\left\{(x-y)^*(\mu(x)-\mu(y)), \tfrac{1}{2}\|\sigma(x)-\sigma(y)\|_F^2\right\} \leqslant c\|x-y\|^2, \tag{1}$$

for every $t \in [0,T]$, $x \in \mathbb{R}^d$ let $X_t^x = (X_{t,s}^x)_{s\in[t,T]} \colon [t,T]\times\Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in [t,T]$ it holds a.s. that

$$X_{t,s}^x = x + \int_t^s \mu(X_{t,r}^x)\,\mathrm{d}r + \int_t^s \sigma(X_{t,r}^x)\,\mathrm{d}W_r, \tag{2}$$

assume for all $t \in [0,T]$, $\omega \in \Omega$ that $([t,T]\times\mathbb{R}^d \ni (s,x) \mapsto X_{t,s}^x(\omega) \in \mathbb{R}^d) \in C^{0,1}([t,T]\times\mathbb{R}^d, \mathbb{R}^d)$, for every $t \in [0,T]$, $x \in \mathbb{R}^d$ let $Z_t^x = (Z_{t,s}^x)_{s\in(t,T]} \colon (t,T]\times\Omega \to \mathbb{R}^{d+1}$ be an $(\mathbb{F}_s)_{s\in(t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in (t,T]$ it holds a.s. that

$$Z_{t,s}^x = \begin{pmatrix} 1 \\ \frac{1}{s-t}\int_t^s (\sigma(X_{t,r}^x))^{-1}\left(\frac{\partial}{\partial x}X_{t,r}^x\right)\mathrm{d}W_r \end{pmatrix}, \tag{3}$$

let $f \in C([0,T)\times\mathbb{R}^d\times\mathbb{R}^{d+1}, \mathbb{R})$, $g \in C(\mathbb{R}^d, \mathbb{R})$ satisfy for all $t \in [0,T)$, $x \in \mathbb{R}^d$, $v, w \in \mathbb{R}^{d+1}$ that $|f(t,x,v) - f(t,x,w)| \leqslant c|\!|\!|v-w|\!|\!|$ and $\max\{|g(x)|, |f(t,x,0)|\} \leqslant c(\|x\|^c + 1)$. Then there exists a unique $v \in C([0,T)\times\mathbb{R}^d, \mathbb{R}^{d+1})$ such that $(v(t,x)\sqrt{T-t})_{t\in[0,T),x\in\mathbb{R}^d}$ grows at most polynomially and for all $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$v(t,x) = \mathbb{E}\left[g(X_{t,T}^x)Z_{t,T}^x + \int_t^T f(r, X_{t,r}^x, v(r, X_{t,r}^x))Z_{t,r}^x\,\mathrm{d}r\right]. \tag{4}$$

Theorem 1 is an immediate consequence of Corollary 1.2.6 in Section 1.2 below. Corollary 1.2.6, in turn, is a special case of Theorem 1.2.5. In this regard, Chapter 1 generalizes the findings in [13] from gradient-independent to gradient-dependent nonlinearities.

In Chapter 2 we further analyse the established connection between certain SFPEs and PDEs. In particular, we prove in Theorem 2.3.1 in Section 2.3 below that suitable solutions to certain SFPEs are also viscosity solutions of the corresponding semilinear PDEs. To achieve this, we establish a Bismut-Elworthy-Li type formula (cf. Theorem 2.2.3) and apply this to a certain class of semilinear PDEs with gradient-dependent nonlinearities.

This leads to a one-to-one correspondence between viscosity solutions of certain semilinear PDEs and solutions of the connected SFPEs. To illustrate the main findings of Chapter 2, we provide in the following theorem, Theorem 2, a special case of Theorem 2.3.1.

Theorem 2. Let $d \in \mathbb{N}$, $\alpha, c, L, T \in (0, \infty)$, let $\langle \cdot, \cdot \rangle \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the standard Euclidean scalar product on $\mathbb{R}^d$, let $\|\cdot\| \colon \mathbb{R}^d \to [0, \infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, let $|\!|\!|\cdot|\!|\!| \colon \mathbb{R}^{d+1} \to [0, \infty)$ be the standard Euclidean norm on $\mathbb{R}^{d+1}$, let $\|\cdot\|_F \colon \mathbb{R}^{d\times d} \to [0, \infty)$ be the Frobenius norm on $\mathbb{R}^{d\times d}$, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_s)_{s\in[0,T]})$ be a filtered probability space satisfying the usual conditions, let $W \colon [0, T] \times \Omega \to \mathbb{R}^d$ be a standard $(\mathbb{F}_s)_{s\in[0,T]}$-Brownian motion, let $\mu \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, $\sigma \in C^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$ satisfy for all $s \in [0, T]$, $x, y \in \mathbb{R}^d$, $v \in \mathbb{R}^d$ that

$$\max\left\{\langle x - y, \mu(x) - \mu(y)\rangle, \tfrac{1}{2}\|\sigma(x) - \sigma(y)\|_F^2\right\} \leqslant \tfrac{c}{2}\|x - y\|^2, \tag{5}$$

$\max\{\langle x, \mu(x)\rangle, \|\sigma(x)\|_F^2\} \leqslant c(1 + \|x\|^2)$, and $v^*\sigma(x)(\sigma(x))^*v \geqslant \alpha\|v\|^2$, assume for all $j \in \{1, 2, \ldots, d\}$ that $\frac{\partial\mu}{\partial x}$ and $\frac{\partial\sigma}{\partial x_j}$ are locally Lipschitz continuous, for every $t \in [0, T]$, $x \in \mathbb{R}^d$ let $X_t^x = (X_{t,s}^x)_{s\in[t,T]} \colon [t, T] \times \Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in [t, T]$ it holds a.s. that

$$X_{t,s}^x = x + \int_t^s \mu(X_{t,r}^x)\,\mathrm{d}r + \int_t^s \sigma(X_{t,r}^x)\,\mathrm{d}W_r, \tag{6}$$

assume for all $t \in [0, T]$, $\omega \in \Omega$ that $\big([t, T] \times \mathbb{R}^d \ni (s, x) \mapsto X_{t,s}^x(\omega) \in \mathbb{R}^d\big) \in C^{0,1}([t, T] \times\mathbb{R}^d, \mathbb{R}^d)$, for every $t \in [0, T]$, $x \in \mathbb{R}^d$ let $Z_t^x = (Z_{t,s}^x)_{s\in(t,T]} \colon (t, T] \times \Omega \to \mathbb{R}^{d+1}$ be an $(\mathbb{F}_s)_{s\in(t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in (t, T]$ it holds a.s. that

$$Z_{t,s}^x = \begin{pmatrix} 1 \\ \frac{1}{s-t}\int_t^s (\sigma(X_{t,r}^x))^{-1}\left(\frac{\partial}{\partial x}X_{t,r}^x\right)\mathrm{d}W_r \end{pmatrix}, \tag{7}$$

let $f \in C([0, T]\times\mathbb{R}^d\times\mathbb{R}^{d+1}, \mathbb{R})$, $g \in C(\mathbb{R}^d, \mathbb{R})$ be at most polynomially growing, and assume for all $t \in [0, T]$, $s \in [t, T]$ $x \in \mathbb{R}^d$, $w_1, w_2 \in \mathbb{R}^{d+1}$ that $\mathbb{E}[|g(X_{t,T}^x)|^2] + \mathbb{E}[|f(s, X_{t,s}^x, w_1)|^2] < \infty$ and $|f(t, x, w_1) - f(t, x, w_2)| \leqslant L|\!|\!|w_1 - w_2|\!|\!|$. Then

(i) there exists a unique $v \in C([0, T] \times \mathbb{R}^d, \mathbb{R}) \cap C^{0,1}([0, T) \times \mathbb{R}^d, \mathbb{R})$ which satisfies that $((v, \nabla_x v)(t, x)\sqrt{T - t})_{t\in[0,T),x\in\mathbb{R}^d}$ grows at most polynomially and for all $t \in [0, T)$, $x \in \mathbb{R}^d$ it holds that $\mathbb{E}[|g(X_{t,T}^x)||\!|\!|Z_{t,T}^x|\!|\!| + \int_t^T|f(r, X_{t,r}^x, v(r, X_{t,r}^x), (\nabla_x v)(r, X_{t,r}^x))| \cdot|\!|\!|Z_{t,r}^x|\!|\!|\,\mathrm{d}r] < \infty$ and

$$\begin{aligned}&(v, \nabla_x v)(t, x)\\ &= \mathbb{E}\left[g(X_{t,T}^x)Z_{t,T}^x + \int_t^T f(r, X_{t,r}^x, (v, \nabla_x v)(r, X_{t,r}^x))Z_{t,r}^x\,\mathrm{d}r\right],\end{aligned} \tag{8}$$

(ii) there exists a unique at most polynomially growing viscosity solution $u \in C([0, T] \times \mathbb{R}^d, \mathbb{R})$ of

$$\begin{aligned}&(\tfrac{\partial u}{\partial t})(t, x) + \langle\mu(x), (\nabla_x u)(t, x)\rangle\\ &\quad+ \tfrac{1}{2}\operatorname{Tr}(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t, x)) + f(t, x, (u, \nabla_x u)(t, x)) = 0\end{aligned} \tag{9}$$

with $u(T, x) = g(x)$ for $(t, x) \in (0, T) \times \mathbb{R}^d$, and

(iii) for all $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that $u(t, x) = v(t, x)$.

Theorem 2 is an immediate consequence of Corollary 2.3.3 in Section 2.3 below. Corollary 2.3.3 follows from Corollary 2.3.2 which is a special version of Theorem 2. The findings in Chapter 2 expand the results in [10] which covers the gradient-independent case.

In Chapter 3 we convert these theoretical results from Section 1 and 2 to a numerical approximation scheme by introducing a new class of Multilevel Picard (MLP) approximation schemes which are suitable for a certain class of semilinear PDES with Lipschitz coefficients and gradient-dependent nonlinearities. The findings of Chapters 1 and 2 imply that MLP approximation schemes are able to numerically compute viscosity solutions of certain semilinear PDEs with gradient-dependent nonlinearities. Combining this with the fact that MLP approximation schemes have been shown to overcome the curse of dimensionality justifies the construction of MLP approximation schemes as approach to determine solutions of semilinear PDEs with gradient-dependent nonlinearities. More precisely, Theorem 2.3.1 in Chapter 2 obtains a concrete representation of PDE solutions as solution of associated SFPEs. These solutions can then be approximated by Monte Carlo methods and a suitable discretization of the SDE leading to our approximation scheme. We are able to show in Corollary 3.2.8 in Section 3.2 below that - under the assumption of a smooth PDE solution - the overall complexity analysis for the proposed approximation scheme grows as $O(d\varepsilon^{-(6+\varepsilon)})$ for all $\delta \in (0,\infty)$ provided that d denotes the dimension of the problem and ε the prescribed accuracy. In this sense, the approximation scheme overcomes the curse of dimensionality and is therefore suitable for PDEs of high dimensions. To illustrate the main result of Chapter 3, we provide the following theorem, Theorem 3.

Theorem 3. Let $d, n, M, Q \in \mathbb{N}$, $\alpha, \delta, T \in (0,\infty)$, $L, K \in [1,\infty)$, let $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, let $|||\cdot|||\colon \mathbb{R}^{d+1} \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^{d+1}$, let $\mu = (\mu_1, \ldots, \mu_d) \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, $\sigma = (\sigma_{ij})_{i,j\in\{1,2,\ldots,d\}} \in C^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$ satisfy for all $x, y \in \mathbb{R}^d$ that

$$\max\{\textstyle\sum_{i,j=1}^d |\frac{\partial \mu_i}{\partial x_j}(x)|, \sum_{i,j=1}^d (\sum_{k=1}^d |\frac{\partial \sigma_{ij}}{\partial x_k}(x)|)^2\} \leqslant K \tag{10}$$

and $y^*\sigma(x)(\sigma(x))^*y \geqslant \alpha\|y\|^2$, let $f \in C(\mathbb{R}^{d+1}, \mathbb{R})$, $g \in C(\mathbb{R}^d, \mathbb{R})$ satisfy for all $x_1, x_2 \in \mathbb{R}^d$, $y_1, y_2 \in \mathbb{R}^{d+1}$, that $\max\{|f(y_1)-f(y_2)|, |g(x_1)-g(x_2)|\} \leqslant L|||y_1 - y_2|||$, let $u^\infty \in C^\infty([0,T]\times\mathbb{R}^d, \mathbb{R})$ satisfy for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that $u(T,x) = g(x)$ and

$$\begin{aligned}&(\tfrac{\partial u^\infty}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x u^\infty)(t,x)\rangle \\ &+ \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x u^\infty)(t,x)) + f(t,x,u^\infty(t,x),(\nabla_x u^\infty)(t,x)) = 0,\end{aligned} \tag{11}$$

for every $N \in \mathbb{N}$, $t \in [0,T]$ let $(\mathfrak{c}_i^N)_{i\in\{1,2,\ldots,N\}} \subseteq [-1,1]$ be the N distinct roots of the Legendre polynomial $[-1,1] \ni x \mapsto \frac{1}{2^N N!}\frac{\partial^N}{\partial x^N}[(x^2-1)^N] \in \mathbb{R}$ and $q^{N,[t,T]}\colon [t,T] \to \mathbb{R}$ be the function which satisfies for all $s \in [t,T]$ that

$$q^{N,[t,T]}(s) = \begin{cases} \int_t^T \left[\prod_{\substack{i\in\{1,2,\ldots,N\} \\ \mathfrak{c}_i^N \neq \frac{2s-(t+T)}{T-t}}} \frac{2x-(T-t)\mathfrak{c}_i^N-(t+T)}{2s-(T-t)\mathfrak{c}_i^N-(t+T)} \right] dx & : (t<T), (\frac{2s-(t+T)}{T-t} \in \{\mathfrak{c}_1^N, \ldots, \mathfrak{c}_N^N\}) \\ 0 & : \text{else}, \end{cases} \tag{12}$$

let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_s)_{s\in[0,T]})$ be a filtered probability space satisfying the usual conditions, let $\Theta = \bigcup_{n\in\mathbb{N}} \mathbb{Z}^n$, let $W^\theta\colon [0,T]\times\Omega \to \mathbb{R}^d$, $\theta \in \Theta$, be i.i.d. standard $(\mathbb{F}_s)_{s\in[0,T]}$-Brownian motions, let $\lfloor\cdot\rfloor_t\colon \mathbb{R} \to \mathbb{R}$, $t \in [0,T]$, satisfy for all $s \in [0,T]$ that $\lfloor s\rfloor_t = \sup([0,s] \cap (\frac{(T-t)}{n}\mathbb{N}))$, for every $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ let $Y_t^{\theta,x,n} = (Y_{t,s}^{\theta,x,n})_{s\in[t,T]}\colon [t,T]\times\Omega \to \mathbb{R}^d$

satisfy for all $s \in [t,T]$ that $Y_{t,t}^{\theta,x,n} = x$ and

$$\begin{aligned} &Y_{t,s}^{\theta,x,n} - Y_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n} \\ &= \mu(Y_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n})(s - \max\{t,\lfloor s \rfloor_t\}) + \sigma(Y_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n})(W_s^\theta - W_{\max\{t,\lfloor s \rfloor_t\}}^\theta), \end{aligned} \tag{13}$$

for every $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ let $D_t^{\theta,x,n} = ((D_{t,s}^{\theta,x,n})_{ij})_{s\in[t,T],i,j\in\{1,2,\ldots,d\}} \colon [t,T]\times\Omega \to \mathbb{R}^{d\times d}$ satisfy for all $i,j \in \{1,2,\ldots,d\}$, $s \in [t,T]$ that $(D_{t,t}^{\theta,x,n})_{ij} = \delta_{ij}$ and

$$\begin{aligned} &(D_{t,s}^{\theta,x,n})_{ij} - (D_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n})_{ij} \\ &= \sum_{k=1}^d \Bigg[\frac{\partial \mu_i}{\partial x_k}(Y_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n})(D_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n})_{kj}(s - \max\{t,\lfloor s \rfloor_t\}) \\ &\quad + \sum_{m=1}^d \Bigg[\frac{\partial \sigma_{im}}{\partial x_k}(Y_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n})(D_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n})_{kj}(W_s^{\theta,m} - W_{\max\{t,\lfloor s \rfloor_t\}}^{\theta,m}) \Bigg] \Bigg], \end{aligned} \tag{14}$$

let $\mathbf{e}_1 \in \mathbb{R}^{d+1}$ satisfy $\mathbf{e}_1 = (1,0,0,\ldots,0)$, for every $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ let $Z_t^{\theta,x,n} = (Z_{t,s}^{\theta,x,n})_{s\in(t,T]} \colon (t,T] \times \Omega \to \mathbb{R}^{d+1}$ satisfy for all $s \in (t,T]$ that $Z_{t,t}^{\theta,x,n} = \mathbf{e}_1$ and

$$Z_{t,s}^{\theta,x,n} - Z_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n} = \begin{pmatrix} 0 \\ \frac{1}{s-t}(\sigma(Y_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n}))^{-1} D_{t,\max\{t,s\}}^{\theta,x,n}(W_s^\theta - W_{\max\{t,\lfloor s \rfloor_t\}}^\theta) \end{pmatrix}, \tag{15}$$

let $V_{n,m,M,Q}^\theta \colon [0,T) \times \mathbb{R}^d \times \Omega \to \mathbb{R}^{d+1}$, $m \in \mathbb{Z}$, $\theta \in \Theta$, satisfy for all $\theta \in \Theta$, $m \in \mathbb{Z}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} V_{n,m,M,Q}^\theta(t,x) &= g(x)\mathbf{e}_1 + \frac{\mathbb{1}_{\mathbb{N}}(n)}{M^m} \sum_{i=1}^{M^m} \big(g(Y_{t,T}^{(\theta,0,-i),x,n}) - g(x)\big) Z_{t,T}^{(\theta,0,-i),x,n} \\ &+ \sum_{l=0}^{m-1} \sum_{s\in(t,T)} \frac{q^{Q,[t,T]}(s)}{M^{m-l}} \Bigg[\sum_{i=1}^{M^{m-l}} (f(V_{n,l,M,Q}^{(\theta,l,i)}(s, Y_{t,s}^{(\theta,l,i),x,n})) \\ &- \mathbb{1}_{\mathbb{N}}(l) f(V_{n,l-1,M,Q}^{(\theta,-l,i)}(s, Y_{t,s}^{(\theta,l,i),x,n}))) Z_{t,s}^{(\theta,l,i),x,n} \Bigg], \end{aligned} \tag{16}$$

let $C \in [0,\infty)$ satisfy that

$$\begin{aligned} C = \max\Bigg\{ &\sqrt{\pi T}, L, \big(\sqrt{T} + (1 + \tfrac{1}{\sqrt{\alpha}}\sqrt{3KT}\exp\big(\tfrac{5}{2}KT\big))\big), \\ &K\sqrt{\tfrac{1}{\alpha}\big(1 + 3KT\exp(5KT)\big)}, \Big(2K\big([\sup_{x\in\mathbb{R}^d}(8(\|\mu(x)\|^4 + 4\|\sigma(x)\|_F^4))] \\ &\cdot T\exp\big(6 + 32K^4T\big)\big)^{\frac{1}{4}} + [\sup_{x\in\mathbb{R}^d}(\|\mu(x)\| + \|\sigma(x)\|_F)]\Big) \\ &\cdot \Big[\sqrt{T} + \sqrt{\tfrac{6}{\alpha}}\big(1 + \sqrt{6K(1+T)}\exp(40K^2(1+T)^2)\big)\Big], \\ &\sup_{k\in\mathbb{N}_0}\sup_{s\in[0,T]}\sup_{y\in\mathbb{R}^d} \frac{\big\|(1,\nabla_x)((\frac{\partial u^\infty}{\partial t}) + \langle \mu, (\nabla_x u^\infty)\rangle + \frac{1}{2}\operatorname{Tr}(\sigma\sigma^*(\operatorname{Hess}_x u^\infty)))^k(s,y)\big\|}{(k!)^{\frac{3}{4}}}, \\ &\sup_{y\in\mathbb{R}^d}|g(y)|, \sup_{s\in[0,T)}\sup_{y\in\mathbb{R}^d}|f(0)| \Bigg\}, \end{aligned} \tag{17}$$

for every $n,m,M,Q \in \mathbb{N}$ let $RN_{n,m,M,Q} \subseteq \mathbb{N}_0$ be number of realizations of standard normal random variables required to compute on realization of the random variable $V_{n,m,M,Q}^0(0,0) \colon \Omega \to \mathbb{R}$ and and $FE_{n,m,M,Q} \subseteq \mathbb{N}_0$ be the number of function evaluations of f

and g required to compute one realization of of the random variable $V^0_{n,m,M,Q}(0,0)\colon \Omega \to \mathbb{R}$ (cf. Corollary 3.2.8 for the precise definitions). Then it holds for all $N \in \mathbb{N}$ that

$$\begin{aligned} &RN_{N^N,N,N,N} + FE_{N^N,N,N,N} \\ &\leqslant \left[\sup_{t\in[0,T)} \sup_{x\in\mathbb{R}^d} \| V^0_{N^N,N,N,N}(t,x) - v^\infty(t,x) \|_{L^2(\mathbb{P};\mathbb{R}^d)} \right]^{-(6+\delta)} \\ &\quad \cdot 16d \sum_{\mathfrak{n}\in\mathbb{N}} \sqrt{\mathfrak{n}^{-\delta\mathfrak{n}}} \left(56 C^{4\mathfrak{n}} 2^{5\mathfrak{n}} T^{3\mathfrak{n}} \right)^{(6+\delta)}. \end{aligned} \tag{18}$$

Theorem 3 is a direct consequence of Corollary 3.2.8 in Section 3.2 below. Corollary 3.2.8 follows from Corollary 3.2.7 which is a special version of Corollary 3.2.6. Corollary 3.2.6, in turn, is a consequence of Theorem 3.2.4. The findings in Chapter 3 generalize the results of [54] from the case of the heat equation with gradient-dependent nonlinearities to the setting of general semilinear PDEs with gradient-dependent nonlinearities.

As already mentioned above, a different approach to overcome the curse of dimensionality is to reformulate the PDEs as stochastic learning problems and then solve them with DNNs. In this thesis, we focus on stochastic gradient descent (SGD) optimization methods as they can be shown to perform very effectively in the training of DNNs. Although SGD optimization methods are already used in numerical simulations, the mathematical proofs for the convergence of these optimization methods are still an open issue. Over the last few years first achievements in providing mathematical proofs for the convergence analysis of SGD optimization methods have been accomplished. In particular, there are abstract convergence results for GD type optimization methods under convexity assumptions on the function which intends to minimize the optimization problem (see, e.g., [6, 80, 81, 82, 89, 66, 94] and the references mentioned therein), abstract convergence results for GD type optimization methods without this type of convexity assumption (see, e.g., [1, 2, 5, 18, 30, 41, 70, 71, 74, 76, 87, 99, 93, 99, 29, 71, 100, 74, 5]), convergence results for GD type optimization in the so-called overparametrized regime (see, e.g., [4, 31, 22, 101, 34, 62, 3, 32, 93, 102]), convergence results for GD type optimization methods in the training of shallow artificial neural networks under specific assumptions on the target function (see, e.g., [64, 38, 24, 63, 63]), results for lower bounds for approximation errors for GD type optimization methods (see [65]), and certain non-convergence results for GD type optimization methods in the training of artificial neural networks (see, e.g., [23, 77, 95]). In Chapter 4 we consider SGD type optimization methods in the training of fully-connected feed-forward deep artificial neural networks with Rectified Linear Unit activation function. We start by proving general regularity properties and representation results for the risk functions and their generalized gradient functions in the setting of a general measurable target function in Section 4.1. These results ensure in Section 4.2 that the risk of the gradient flow (GF) processes converge to zero with convergence rate 1 provided that the target function is constant. In Section 4.3 we use some of the key findings from Sections 4.1 and 4.2 to prove that under the assumption of a constant target function and sufficiently small but not L^1-summable step sizes the risk functions of gradient descent (GD) processes converge to zero. We expand these results in Section 4.4 to establish that the expected risk of the SGD process converges to zero as the number of SGD steps goes to infinity provided that the target function is constant and the step sizes are sufficiently small but not L^1-summable. To illustrate this result, we provide in the following theorem, Theorem 4, a special case of Corollary 4.4.12 in Section 4.4.7 below.

Theorem 4. Let $L, \mathfrak{d} \in \mathbb{N}$, $(\ell_k)_{k\in\mathbb{N}_0} \subseteq \mathbb{N}$, $a \in \mathbb{R}$, $b \in (a,\infty)$, $\xi \in \mathbb{R}^{\ell_L}$ satisfy $\mathfrak{d} = \sum_{k=1}^{L} \ell_k(\ell_{k-1}+1)$, for every $\theta = (\theta_1,\dots,\theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ let $\mathfrak{w}^{k,\theta} = (\mathfrak{w}^{k,\theta}_{i,j})_{(i,j)\in\{1,\dots,\ell_k\}\times\{1,\dots,\ell_{k-1}\}} \in$

$\mathbb{R}^{\ell_k \times \ell_{k-1}}$, $k \in \mathbb{N}$, and $\mathfrak{b}^{k,\theta} = (\mathfrak{b}^{k,\theta}_i)_{i \in \{1,\dots,\ell_k\}} \in \mathbb{R}^{\ell_k}$, $k \in \mathbb{N}$, satisfy for all $k \in \{1,\dots,L\}$, $i \in \{1,\dots,\ell_k\}$, $j \in \{1,\dots,\ell_{k-1}\}$ that

$$\mathfrak{w}^{k,\theta}_{i,j} = \theta_{(i-1)\ell_{k-1}+j+\sum_{h=1}^{k-1}\ell_h(\ell_{h-1}+1)} \qquad \text{and} \qquad \mathfrak{b}^{k,\theta}_i = \theta_{\ell_k\ell_{k-1}+i+\sum_{h=1}^{k-1}\ell_h(\ell_{h-1}+1)}, \tag{19}$$

for every $k \in \mathbb{N}$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ let $\mathcal{A}^\theta_k \colon \mathbb{R}^{\ell_{k-1}} \to \mathbb{R}^{\ell_k}$ satisfy for all $x \in \mathbb{R}^{\ell_{k-1}}$ that $\mathcal{A}^\theta_k(x) = \mathfrak{b}^{k,\theta} + \mathfrak{w}^{k,\theta}x$, let $\mathscr{R}_r \colon \mathbb{R} \to \mathbb{R}$, $r \in [1,\infty]$, satisfy for all $r \in [1,\infty)$, $x \in (-\infty, 2^{-1}r^{-1}]$, $y \in \mathbb{R}$, $z \in [r^{-1},\infty)$ that

$$\mathscr{R}_r \in C^1(\mathbb{R},\mathbb{R}), \quad \mathscr{R}_r(x) = 0, \quad 0 \leqslant \mathscr{R}_r(y) \leqslant \mathscr{R}_\infty(y) = \max\{y,0\}, \text{ and } \mathscr{R}_r(z) = z, \tag{20}$$

assume $\sup_{r\in[1,\infty)} \sup_{x\in\mathbb{R}} |(\mathscr{R}_r)'(x)| < \infty$, let $\|\cdot\| \colon (\cup_{n\in\mathbb{N}}\mathbb{R}^n) \to \mathbb{R}$ and $\mathfrak{M}_r \colon (\cup_{n\in\mathbb{R}}\mathbb{R}^n) \to (\cup_{n\in\mathbb{N}}\mathbb{R}^n)$, $r \in [1,\infty]$, satisfy for all $r \in [1,\infty]$, $n \in \mathbb{N}$, $x = (x_1,\dots,x_n) \in \mathbb{R}^n$ that $\|x\| = (\sum_{i=1}^n |x_i|^2)^{1/2}$ and $\mathfrak{M}_r(x) = (\mathscr{R}_r(x_1),\dots,\mathscr{R}_r(x_n))$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $X^{n,m} \colon \Omega \to [a,b]^{\ell_0}$, $n,m \in \mathbb{N}_0$, be i.i.d. random variables, let $(M_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{N}$, for every $r \in [1,\infty]$ let $\mathcal{N}^{k,\theta}_r \colon \mathbb{R}^{\ell_0} \to \mathbb{R}^{\ell_L}$, $\theta \in \mathbb{R}^{\mathfrak{d}}$, $k \in \mathbb{N}$, and $\mathfrak{L}^n_r \colon \mathbb{R}^{\mathfrak{d}} \times \Omega \to \mathbb{R}$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}_0$, $\theta \in \mathbb{R}^{\mathfrak{d}}$, $k \in \mathbb{N}$, $\omega \in \Omega$ that

$$\mathcal{N}^{1,\theta}_r(x) = \mathcal{A}^\theta_1(x), \qquad \mathcal{N}^{k+1,\theta}_r(x) = \mathcal{A}^\theta_{k+1}(\mathfrak{M}_{r^{1/k}}(\mathcal{N}^{k,\theta}_r(x))), \tag{21}$$

and $\mathfrak{L}^n_r(\theta,\omega) = \frac{1}{M_n}\sum_{m=1}^{M_n} \|\mathcal{N}^{L,\theta}_r(X^{n,m}(\omega)) - \xi\|^2$, let $\mathcal{L} \colon \mathbb{R}^{\mathfrak{d}} \to \mathbb{R}$ satisfy for all $\theta \in \mathbb{R}^{\mathfrak{d}}$ that $\mathcal{L}(\theta) = \mathbb{E}[\|\mathcal{N}^{L,\theta}_\infty(X^{0,0}) - \xi\|^2]$, for every $n \in \mathbb{N}_0$ let $\mathfrak{G}^n \colon \mathbb{R}^{\mathfrak{d}} \times \Omega \to \mathbb{R}^{\mathfrak{d}}$ satisfy for all $\theta \in \mathbb{R}^{\mathfrak{d}}$, $\omega \in \{w \in \Omega \colon ((\nabla_\theta \mathfrak{L}^n_r)(\theta,w))_{r\in[1,\infty)} \text{ is convergent}\}$ that

$$\mathfrak{G}^n(\theta,\omega) = \lim_{r\to\infty} (\nabla_\theta \mathfrak{L}^n_r)(\theta,\omega), \tag{22}$$

let $\Theta = (\Theta_n)_{n\in\mathbb{N}_0} \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^{\mathfrak{d}}$ be a stochastic process, let $(\gamma_n)_{n\in\mathbb{N}_0} \subseteq [0,\infty)$, assume that Θ_0 and $(X^{n,m})_{(n,m)\in(\mathbb{N}_0)^2}$ are independent, and assume for all $n \in \mathbb{N}_0$ that $\Theta_{n+1} = \Theta_n - \gamma_n \mathfrak{G}^n(\Theta_n)$, $(4L\mathfrak{d}\max\{|a|,|b|,\|\xi\|,1\})^{2L}\gamma_n \leqslant (\|\Theta_0\|+1)^{-2L}$, and $\sum_{k=0}^\infty \gamma_k = \infty$. Then

(i) there exists $\mathfrak{C} \in \mathbb{R}$ such that $\mathbb{P}(\sup_{n\in\mathbb{N}_0} \|\Theta_n\| \leqslant \mathfrak{C}) = 1$,

(ii) it holds that $\mathbb{P}(\limsup_{n\to\infty} \mathcal{L}(\Theta_n) = 0) = 1$, and

(iii) it holds that $\limsup_{n\to\infty} \mathbb{E}[\mathcal{L}(\Theta_n)] = 0$.

Theorem 4 is an immediate consequence of Corollary 4.4.12 in Section 4.4.7 below. Corollary 4.4.12, in turn, follows from Theorem 4.4.11. To further illustrate the main findings of Chapter 4, we provide in Section 4.5 two numerical applications of the results presented in Theorem 4. The results in Chapter 4 generalize the findings in [63] from the setting of shallow ANNs with just one hidden layer to deep ANNs with an arbitrary number of hidden layers. The research reported in Chapter 4 was a joint work with Martin Hutzenthaler, Arnulf Jentzen, Adrian Riekert, and Luca Scarpa (cf. [57]).

Notation

The following notation is used throughout this thesis. For all sets A, B, C, D with $C \subseteq D$ and every function $f\colon A \mapsto B$ we denote by $\mathbb{1}_D\colon C \to \{0,1\}$ the function satisfying for all $x \in C$ that

$$\mathbb{1}_D(x) = \begin{cases} 1, & x \in D, \\ 0, & x \neq D \end{cases} \tag{23}$$

and by $f|_C\colon C \cap A \to B$ the function satisfying for all $x \in C \cap A$ that $f|_C(x) = f(x)$. For every $i, j \in \mathbb{N}$ we denote by $\delta_{ij} \in \{0,1\}$ the real number satisfying that

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases} \tag{24}$$

For every $d \in \mathbb{N}$ we denote by $I_d = ((I_d)_{ij})_{i,j \in \{1,\ldots,d\}} \subseteq \mathbb{R}^{d\times d}$ the matrix satisfying for all $i, j \in \{1, \ldots, d\}$ that $(I_d)_{ij} = \delta_{ij}$, we denote for every $A \in \mathbb{R}^{d\times d}$ by $A^* \in \mathbb{R}^{d\times d}$ the transpose of A, and we denote by $\mathbb{S}_d \subseteq \mathbb{R}^{d\times d}$ the set satisfying $\mathbb{S}_d = \{A \in \mathbb{R}^{d\times d}\colon A^* = A\}$. For Hilbert spaces $\mathbb{H}_1 = (H_1, \langle\cdot,\cdot\rangle_{H_1}, \|\cdot\|_{H_1})$, $\mathbb{H}_2 = (H_2, \langle\cdot,\cdot\rangle_{H_2}, \|\cdot\|_{H_2})$ we denote for every $n \in \mathbb{N}_0 \cup \{\infty\}$ by $C^n(H_1, H_2)$ the set satisfying that

$$C^n(H_1, H_2) = \{f\colon H_1 \to H_2\colon f \text{ is } n \text{ times continuously differentiable}\}, \tag{25}$$

and we denote by $C_c^\infty(H_1, H_2)$ the set satisfying that

$$C_c^\infty(H_1, H_2) = \{f \in C^\infty(H_1, H_2)\colon \operatorname{supp}(f) \text{ is compact}\}. \tag{26}$$

For every $d \in \mathbb{N}$, $O \subseteq \mathbb{R}^d$, $\mathbb{R}$-vector space $\mathcal{V}$ we denote by $C^{0,1}([0,T] \times O, \mathcal{V})$ the set of continuous functions $f\colon [0,T] \times O \to \mathcal{V}$ whose first order spatial partial derivatives exist and are jointly continuous in time and space and by $C^{1,2}([0,T] \times O, \mathcal{V})$ the set of functions $f\colon [0,T] \times O \to \mathcal{V}$ whose first temporal derivative and second order spatial partial derivatives exist and are jointly continuous in time and space. For every $d \in \mathbb{N}$ we call the map $\langle\cdot,\cdot\rangle\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ that satisfies for all $x = (x_1, x_2, \ldots, x_d), y = (y_1, y_2, \ldots, y_d) \in \mathbb{R}^d$ that $\langle x, y\rangle = \sum_{i=1}^d x_i y_i$ the standard Euclidean scalar product on $\mathbb{R}^d$ and we call the map $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ which satisfies for all $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $\|x\| = (\sum_{i=1}^d |x_i|^2)^{\frac{1}{2}}$ the standard Euclidean norm on $\mathbb{R}^d$. For every $d \in \mathbb{N}$ and $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ the standard Euclidean norm on $\mathbb{R}^d$ we denote by $L(\mathbb{R}^d)$ the vector space $L(\mathbb{R}^d) = \{A\colon \mathbb{R}^d \to \mathbb{R}^d\colon A \text{ is linear}\}$ and by $\|\cdot\|_{L(\mathbb{R}^d)}\colon L(\mathbb{R}^d) \to [0,\infty)$ the norm satisfying for all $A \in L(\mathbb{R}^d)$ that

$$\|A\|_{L(\mathbb{R}^d)} = \sup_{x \in \mathbb{R}^d \setminus \{0\}} \frac{\|Ax\|}{\|x\|}. \tag{27}$$

For every topological space $(E, \mathcal{E})$ we denote by $\mathcal{B}(E)$ the Borel-sigma-algebra on $(E, \mathcal{E})$. For all measure spaces $(A, \mathcal{A})$, $(B, \mathcal{B})$ we denote by $\mathcal{M}(\mathcal{A}, \mathcal{B})$ the set given by $\mathcal{M}(\mathcal{A}, \mathcal{B}) = \{f\colon A \to B\colon f \text{ is } \mathcal{A}/\mathcal{B}\text{-measurable}\}$. For a measure space $(\Omega, \mathcal{A}, \mu)$, a normed vector space $(V, \|\cdot\|)$, and a real number $q \in (0,\infty)$ we denote by $L^0(\mu; \mathbb{R}^d)$ the set satisfying that $L^0(\mu; \mathbb{R}^d) = \mathcal{M}(\mathcal{A}, \mathcal{B}(\mathbb{R}^d))$, we denote by $\|\cdot\|_{L^q(\mu;\mathbb{R}^d)}\colon L^0(\mu; \mathbb{R}^d) \to [0,\infty]$ the mapping which satisfies for all $f \in L^0(\mu; \mathbb{R}^d)$ that

$$\|f\|_{L^q(\mu;\mathbb{R}^d)} = \left(\int_\Omega \|f(\omega)\|^q \mu(d\omega) \right)^{\frac{1}{q}} \tag{28}$$

and we denote by $L^q(\mu;\mathbb{R}^d)$ the set given by

$$L^q(\mu;\mathbb{R}^d) = \{f \in L^0(\mu;\mathbb{R}^d) \colon \|f\|_{L^q(\mu;\mathbb{R}^d)} < \infty\}. \tag{29}$$

We say a filtered measure space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_s)_{s\in[0,T]})$ satisfies the usual conditions if and only if for all $t \in [0,T)$ it holds that $\{A \subseteq \mathcal{F} \colon \mathbb{P}(A) = 0\} \subseteq \mathbb{F}_0$ and $\mathbb{F}_t = \cap_{s\in(t,T]}\mathbb{F}_s$. We denote by $\frac{0}{0}$, $0 \cdot \infty$, 0^0, and $\sqrt{\infty}$ the extended real numbers given by $\frac{0}{0} = 0$, $0 \cdot \infty = 0$, $0^0 = 1$, and $\sqrt{\infty} = \infty$. For every $a \in (0,\infty)$, $b \in \mathbb{R}$ we denote by $\frac{a}{0}$, $\frac{-a}{0}$, 0^a, 0^{-a}, $\frac{1}{0^a}$, and $\frac{b}{\infty}$ the extended real numbers given by $\frac{a}{0} = \infty$, $\frac{-a}{0} = -\infty$, $0^a = 0$, $0^{-a} = \infty$, $\frac{1}{0^a} = \infty$, and $\frac{b}{\infty} = 0$. For every $A \subseteq \mathbb{Z}$, $f \colon A \to \mathbb{R}$, $k \in \mathbb{Z}$ we denote by $\prod_{i=k}^{k-1} f(i)$ and $\sum_{i=k}^{k-1} f(i)$ the real numbers given by $\prod_{i=k}^{k-1} f(i) = 1$ and $\sum_{i=k}^{k-1} f(i) = 0$.

Chapter 1

Existence and Uniqueness of Stochastic fixed point equations (SFPEs)

In this chapter we analyse a certain class of SFPEs (see, e.g., (1.86) in Theorem 1.1.11 below) with the goal to prove existence and uniqueness of solutions of these SFPEs. The approach to construct these solutions arises from the Feynman-Kac formula which forges a strong connection between classical solutions of semilinear Kolmogorov PDEs and SDEs. To transfer this idea to the setting of semilinear PDEs with gradient-dependent nonlinearities we apply a combination of Itô's formula and a Bismut-Elworthy-Li formula. In particular, if $X^x_t\colon [t,T]\times\mathbb{R}^d\to\mathbb{R}^d$, $t\in[0,T]$, $x\in\mathbb{R}^d$, solve the SDE

$$dX^x_{t,s}=\mu(s,X^x_{t,s})\,\mathrm{d}s+\sigma(s,X^x_{t,s})\,\mathrm{d}W_s,\qquad s\in[t,T],\ X^x_{t,t}=x, \tag{1.1}$$

and $u\in C^{1,2}([0,T]\times\mathbb{R}^d,\mathbb{R})$ satisfies for all $s\in[0,T]$, $y\in\mathbb{R}^d$ that $u(T,y)=g(y)$ and

$$\begin{aligned}&(\tfrac{\partial u}{\partial t})(t,x)+\langle\mu(t,x),(\nabla_x u)(t,x)\rangle\\&\quad+\tfrac12\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x u)(t,x))+f(t,x,u(t,x),(\nabla_x u)(t,x))=0\end{aligned} \tag{1.2}$$

then the Itô formula demonstrates for all $t\in[0,T]$, $s\in[t,T]$, $x\in\mathbb{R}^d$ that it holds a.s. that

$$\begin{aligned}u(s,X^x_{t,s})=g(X^x_{t,T})&+\int_s^T f(r,X^x_{t,r},u(r,X^x_{t,r}),(\nabla_x u)(r,X^x_{t,r}))\,\mathrm{d}r\\&-\int_s^T((\nabla_x u)\sigma)(r,X^x_{t,r})\,\mathrm{d}W_r.\end{aligned} \tag{1.3}$$

Applying expectations leads to the following connected stochastic fixed point equation

$$\begin{aligned}&u(t,x)=\mathbb{E}\Big[g(X^x_{t,T})+\int_t^T f(r,X^x_{t,r},u(r,X^x_r),(\nabla_x u)(r,X^x_{t,r}))\,\mathrm{d}r\Big],\\&t\in[0,T],\ X^x_{t,t}=x.\end{aligned} \tag{1.4}$$

Since PDEs of the form (1.2) often do not admit a classical solution (cf. [45]), we concentrate on finding solutions of the related SFPE (1.4) instead. If the nonlinearity f of the SFPE in (1.4) is gradient-independent, then the SFPE in (1.4) is closed and can be solved with fixed point methods. In this case and under the additional assumption of Lipschitz continuity of the nonlinearity, the results in [13] show the existence of a unique solution of the SFPE in (1.4).

In this chapter, we extend the results in [13] to gradient-dependent, Lipschitz continuous nonlinearities. In this case, the SFPE in (1.4) is not closed. Our attempt to still solve the SFPE in (1.4) follows the idea that - under suitable assumptions - applying the Bismut-Elworthy-Li formula to (1.4) shows for all $t \in [0, T)$ that

$$\begin{aligned}(\nabla_x u)(t,x) &= \mathbb{E}\Big[g(X^x_{t,T})\tfrac{1}{T-t}\int_t^T (\sigma(r,X^x_{t,r}))^{-1}\big(\tfrac{\partial}{\partial x}X^x_{t,r}\big)\,\mathrm{d}W_r\Big] \\ &\quad + \mathbb{E}\Big[\int_t^T f(r,X^x_{t,r},u(r,X^x_{t,r}),(\nabla_x u)(r,X^x_{t,r}))\tfrac{1}{s-t}\int_t^s (\sigma(r,X^x_{t,r}))^{-1}\big(\tfrac{\partial}{\partial x}X^x_{t,r}\big)\,\mathrm{d}W_r\,\mathrm{d}s\Big]\end{aligned} \tag{1.5}$$

(cf., e.g., [40, Theorem 2.1]). The goal of this chapter is to study the closed system of SFPEs (1.4) and (1.5) and to prove the existence of a unique solution for Lipschitz continuous, gradient-independent nonlinearities. The main challenge of this approach arises from the fact that a non-differentiable terminal condition and the Bismut-Elworthy-Li formula lead to a singularity of (1.5) in the last time point. To still follow the idea of [13] to prove the existence and uniqueness of the solution of the SFPE by Banach's fixed point theorem, it is necessary to define a new, suitable norm of the Banach space on which Banach's fixed point theorem is applied (cf. Theorem 1.1.11). Another challenge comes from allowing non-Lipschitz continuous drift coefficients which only satisfy a one-sided Lipschitz condition. This generalized assumption allows our results to solve an even bigger class of SFPEs and their associated PDEs.

This chapter is structured in the following way. In Section 1.1 we study SFPEs in an abstract setting for general stochastic processes X and Z in (1.175) below. The main result of Section 1.1 is Theorem 1.1.11 in Subsection 1.1.5 below which proves the existence and uniqueness of solutions to SFPEs in an abstract setting by using Banach's fixed point theorem. In Subsections 1.1.1 - 1.1.4 we ensure that the assumption for Banach's fixed point theorem are satisfied. In Section 1.2 we investigate SFPEs of the form (1.175) where X is an SDE solution and Z is a specific stochastic process arising from the Bismut-Elworthy-Li formula. In Theorem 1.2.5, the main result of this chapter, we apply the abstract result Theorem 1.1.11 to these stochastic processes. In Subsections 1.2.1 and 1.2.2 we establish several auxiliary results to ensure the assumptions of Theorem 1.1.11.

1.1 Abstract stochastic fixed point equations (SFPEs)

In this section we study existence, uniqueness, and further general properties of abstract SFPEs. The main result of this section is Theorem 1.1.11 in Subsection 1.1.5 below whose proof is an application of Banach's fixed point theorem to the function in (1.86). Lemma 1.1.8 in Subsection 1.1.3 below ensures well-definedness of this function, Lemma 1.1.9 in Subsection 1.1.4 below establishes the contractivity property, and Lemma 1.1.10 proves that the space to which we apply Banach's fixed point theorem is indeed a Banach space. In this approach Lemma 1.1.2 is a helpful tool to overcome the challenges of the singularity of the stochastic process arising from the Bismut-Elworthy-Li formula, especially in in Lemma 1.1.7, Lemma 1.1.8, and Lemma 1.1.9. Throughout this section we frequently use the following setting.

Setting 1.1.1. *Let $d, m \in \mathbb{N}$, $c, T \in (0,\infty)$, let $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$, $|||\cdot|||\colon \mathbb{R}^m \to [0,\infty)$ be norms, let $O \subseteq \mathbb{R}^d$ be a non-empty open set, for every $r \in (0,\infty)$ let $K_r \subseteq [0,T)$, $O_r \subseteq O$ satisfy $K_r = [0, \max\{T - \frac{1}{r}, 0\}]$ and $O_r = \{x \in O\colon \|x\| \leqslant r$ and $\{y \in \mathbb{R}^d\colon \|y - x\| < \frac{1}{r}\} \subseteq O\}$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, for every $t \in [0,T]$, $x \in O$ let*

$X_t^x = (X_{t,s}^x)_{s\in[t,T]}\colon [t,T]\times\Omega\to O$ and $Z_t^x = (Z_{t,s}^x)_{s\in(t,T]}\colon (t,T]\times\Omega\to\mathbb{R}^m$ be measurable, and let $V\in C([0,T]\times O,(0,\infty))$ satisfy for all $t\in[0,T)$, $s\in(t,T]$, $x\in O$ that $\mathbb{E}[V(s,X_{t,s}^x)\|Z_{t,s}^x\|]\leqslant\frac{c}{\sqrt{s-t}}V(t,x)$.

1.1.1 Integrability of certain stochastic processes

In the following lemma we calculate two specific integrals which will be used to prove upper bounds and ensure the well-definedness of the SFPE (1.86) in Theorem 1.1.11 below.

Lemma 1.1.2. *Let $a\in\mathbb{R}$, $b\in(a,\infty)$. Then*

(i) for all $\lambda\in[0,\infty)$ it holds that

$$\int_a^b \frac{1}{\sqrt{(b-x)(x-a)}}\, e^{-\lambda x}\,\mathrm{d}x = e^{-\frac{\lambda(b+a)}{2}}\int_0^\pi e^{\frac{\lambda(b-a)}{2}\cos(\theta)}\,\mathrm{d}\theta \tag{1.6}$$

and

(ii) for all $\lambda\in(0,\infty)$ it holds that

$$\int_a^b \frac{1}{\sqrt{(b-x)(x-a)}}\, e^{-\lambda x}\,\mathrm{d}x \leqslant \sqrt{\frac{\pi^3}{4\lambda(b-a)}}\, e^{-\lambda a}. \tag{1.7}$$

Proof of Lemma 1.1.2. First note that the substitution $(0,\pi)\ni\theta\mapsto x(\theta)=\frac{c}{2}(1-\cos(\theta))\in(0,c)$ and the fact that for all $x\in\mathbb{R}$ it holds that $(\sin(x))^2+(\cos(x))^2=1$ ensure that for all $\lambda\in[0,\infty)$, $c\in(0,\infty)$ it holds that

$$\begin{aligned}
&e^{\frac{\lambda c}{2}}\int_0^c \frac{1}{\sqrt{x(c-x)}}\, e^{-\lambda x}\,\mathrm{d}x = \int_0^c \frac{1}{\sqrt{x(c-x)}}\, e^{\lambda(\frac{c}{2}-x)}\,\mathrm{d}x\\
&= \int_0^\pi \frac{1}{\sqrt{\frac{c}{2}(1-\cos(\theta))(c-\frac{c}{2}(1-\cos(\theta)))}}\, e^{\frac{\lambda c}{2}\cos(\theta)}\,\frac{c}{2}\sin(\theta)\,\mathrm{d}\theta\\
&= \int_0^\pi \frac{1}{\sqrt{(1-\cos(\theta))(1+\cos(\theta))}}\, e^{\frac{\lambda c}{2}\cos(\theta)}\sin(\theta)\,\mathrm{d}\theta\\
&= \int_0^\pi \frac{1}{\sqrt{1-(\cos(\theta))^2}}\, e^{\frac{\lambda c}{2}\cos(\theta)}\sin(\theta)\,\mathrm{d}\theta = \int_0^\pi e^{\frac{\lambda c}{2}\cos(\theta)}\,\mathrm{d}\theta.
\end{aligned}\tag{1.8}$$

Hence, we obtain for all $\lambda\in[0,\infty)$ that

$$\begin{aligned}
&\int_a^b \frac{1}{\sqrt{(b-x)(x-a)}}\, e^{-\lambda x}\,\mathrm{d}x = e^{-\lambda a}\int_0^{b-a}\frac{1}{\sqrt{(b-a-y)y}}\, e^{-\lambda y}\,\mathrm{d}y\\
&= e^{-\frac{\lambda(b-a)}{2}-\lambda a}\int_0^\pi e^{\frac{\lambda(b-a)}{2}\cos(\theta)}\,\mathrm{d}\theta = e^{-\frac{\lambda(b+a)}{2}}\int_0^\pi e^{\frac{\lambda(b-a)}{2}\cos(\theta)}\,\mathrm{d}\theta.
\end{aligned}\tag{1.9}$$

This establishes item (i). Next, observe that the fact that $\int_0^\infty e^{-z^2}\mathrm{d}z=\frac{\sqrt{\pi}}{2}$ and the fact that for all $y\in[0,\frac{\pi}{2}]$ it holds that $\sin(y)\geqslant\frac{2y}{\pi}$ yield that for all $c\in(0,\infty)$ it holds that

$$\begin{aligned}
&\int_0^\pi e^{c\cos(\theta)}\,\mathrm{d}\theta = 2\int_0^{\frac{\pi}{2}} e^{c\cos(2y)}\,\mathrm{d}y = 2\int_0^{\frac{\pi}{2}} e^{c(1-2(\sin(y))^2)}\,\mathrm{d}y\\
&= 2e^c\int_0^{\frac{\pi}{2}} e^{-2c(\sin(y))^2}\,\mathrm{d}y \leqslant 2e^c\int_0^{\frac{\pi}{2}} e^{-\frac{8cy^2}{\pi^2}}\,\mathrm{d}y \leqslant \frac{2\pi}{\sqrt{8c}}e^c\int_0^\infty e^{-z^2}\,\mathrm{d}z = \sqrt{\frac{\pi^3}{8c}}\,e^c.
\end{aligned}\tag{1.10}$$

Item (i) therefore shows that for all $\lambda \in (0, \infty)$ it holds that

$$\begin{aligned} \int_a^b \frac{1}{\sqrt{(b-x)(x-a)}} e^{-\lambda x} \, dx &= e^{-\frac{\lambda(b+a)}{2}} \int_0^\pi e^{\frac{\lambda(b-a)}{2} \cos(\theta)} \, d\theta \\ &\leqslant e^{-\frac{\lambda(b+a)}{2}} \sqrt{\frac{2\pi^3}{8\lambda(b-a)}} e^{\frac{\lambda(b-a)}{2}} = \sqrt{\frac{\pi^3}{4\lambda(b-a)}} e^{-\lambda a}. \end{aligned} \tag{1.11}$$

The proof of Lemma 1.1.2 is thus complete. □

The next lemma uses Lemma 1.1.2 to establish integrability properties for a certain class of stochastic processes. Lemma 1.1.3 is a generalization of the results in [13, Lemma 2.1].

Lemma 1.1.3. *Let $d, m \in \mathbb{N}$, $c, T \in (0, \infty)$, let $|||\cdot|||\colon \mathbb{R}^m \to [0, \infty)$ be a norm, let $O \subseteq \mathbb{R}^d$ be a non-empty open set, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, for every $t \in [0, T]$, $x \in O$ let $X_t^x = (X_{t,s}^x)_{s \in [t,T]}\colon [t, T] \times \Omega \to O$ and $Z_t^x = (Z_{t,s}^x)_{s \in (t,T]}\colon (t, T] \times \Omega \to \mathbb{R}^m$ be measurable, let $g\colon O \to \mathbb{R}$, $h\colon [0, T] \times O \to \mathbb{R}$, and $V\colon [0, T] \times O \to (0, \infty)$ be measurable, and assume for all $t \in [0, T)$, $s \in (t, T]$, $x \in O$ that $\mathbb{E}[V(s, X_{t,s}^x)|||Z_{t,s}^x|||] \leqslant \frac{c}{\sqrt{s-t}} V(t, x)$ and $\sup_{t \in [0,T)} \sup_{x \in O} \left[\frac{|g(x)|}{V(T,x)} + \frac{|h(t,x)|}{V(t,x)} \sqrt{T-t} \right] < \infty$. Then it holds for all $t \in [0, T)$, $x \in O$ that*

$$\mathbb{E}\left[|g(X_{t,T}^x)| |||Z_{t,T}^x||| + \int_t^T |h(r, X_{t,r}^x)| |||Z_{t,r}^x||| \, dr \right] < \infty. \tag{1.12}$$

Proof of Lemma 1.1.3. Throughout this proof let $\alpha \in [0, \infty)$ satisfy for all $t \in [0, T)$, $x \in O$ that

$$|g(x)| \leqslant \alpha V(T, x) \qquad \text{and} \qquad |h(t, x)| \sqrt{T-t} \leqslant \alpha V(t, x). \tag{1.13}$$

Observe that the assumption that $g\colon O \to \mathbb{R}$ and $h\colon [0, T] \times O \to \mathbb{R}$ are measurable and the fact that for all $t \in [0, T]$, $x \in O$ it holds that $X_t^x\colon [t, T] \times \Omega \to O$ and $Z_t^x\colon (t, T] \times \Omega \to \mathbb{R}^m$ are measurable imply that for all $t \in [0, T)$, $x \in O$ it holds that $\Omega \ni \omega \mapsto g(X_{t,T}^x(\omega)) Z_{t,T}^x(\omega) \in \mathbb{R}^m$ and $(t, T] \times \Omega \ni (s, \omega) \mapsto h(s, X_{t,s}^x(\omega)) Z_{t,s}^x(\omega) \in \mathbb{R}^m$ are measurable. Furthermore, note that Fubini's theorem, (1.13), and the hypothesis that for all $t \in [0, T)$, $s \in (t, T]$, $x \in O$ it holds that $\mathbb{E}[V(s, X_{t,s}^x)|||Z_{t,s}^x|||] \leqslant \frac{c}{\sqrt{s-t}} V(t, x)$ demonstrate that for all $t \in [0, T)$, $x \in O$ it holds that

$$\begin{aligned} &\mathbb{E}\left[|g(X_{t,T}^x)| |||Z_{t,T}^x||| + \int_t^T |h(r, X_{t,r}^x)| |||Z_{t,r}^x||| \, dr \right] \\ &= \mathbb{E}\left[|g(X_{t,T}^x)| |||Z_{t,T}^x||| \right] + \int_t^T \mathbb{E}\left[|h(r, X_{t,r}^x)| |||Z_{t,r}^x||| \right] dr \\ &\leqslant \mathbb{E}\left[\alpha V(T, X_{t,T}^x) |||Z_{t,T}^x||| \right] + \int_t^T \frac{1}{\sqrt{T-r}} \mathbb{E}\left[\alpha V(r, X_{t,r}^x) |||Z_{t,r}^x||| \right] dr \\ &\leqslant \frac{\alpha c}{\sqrt{T-t}} V(t, x) + \int_t^T \frac{\alpha c}{\sqrt{(T-r)(r-t)}} V(t, x) \, dr. \end{aligned} \tag{1.14}$$

Item (i) of Lemma 1.1.2 (applied for every $t \in [0, T)$ with $a \curvearrowleft t$, $b \curvearrowleft T$, $\lambda \curvearrowleft 0$ in the notation of Lemma 1.1.2) therefore shows that for all $t \in [0, T)$, $x \in O$ it holds that

$$\begin{aligned} &\mathbb{E}\left[|g(X_{t,T}^x)| |||Z_{t,T}^x||| + \int_t^T |h(r, X_{t,r}^x)| |||Z_{t,r}^x||| \, dr \right] \\ &\leqslant \alpha c \left(\frac{1}{\sqrt{T-t}} + \pi \right) V(t, x) < \infty. \end{aligned} \tag{1.15}$$

The proof of Lemma 1.1.3 is thus complete. □

1.1.2 Continuity of SFPEs with respect to coefficient functions

In this subsection we establish some general approximation results with the goal to approximate certain SFPEs in Lemma 1.1.8. We achieve this by approximating the terminal conditions and nonlinearities of the SFPEs properly. For this, the following lemma demonstrates several properties of approximating functions. Lemma 1.1.4 extends [13, Lemma 2.2] by considering the $\sqrt{T-t}$ term and replacing $[0,T]$ by $[0,T]\backslash K_r$.

Lemma 1.1.4. *Assume Setting 1.1.1, let $g_n \in C(O,\mathbb{R})$, $n \in \mathbb{N}_0$, and $h_n \in C([0,T)\times O,\mathbb{R})$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}$ that $\inf_{r\in(0,\infty)}[\sup_{t\in[0,T)\backslash K_r} \sup_{x\in O\backslash O_r} (\frac{|g_n(x)|}{V(T,x)} + \frac{|h_n(t,x)|}{V(t,x)}\sqrt{T-t})] = 0$, and assume that*

$$\limsup_{n\to\infty}\left[\sup_{t\in[0,T)}\sup_{x\in O}\left(\frac{|g_n(x)-g_0(x)|}{V(T,x)} + \frac{|h_n(t,x)-h_0(t,x)|}{V(t,x)}\sqrt{T-t}\right)\right] = 0. \tag{1.16}$$

Then

(i) it holds for every $n \in \mathbb{N}_0$ that

$$\sup_{t\in[0,T)}\sup_{x\in O}\left[\frac{|g_n(x)|}{V(T,x)} + \frac{|h_n(t,x)|}{V(t,x)}\sqrt{T-t}\right] < \infty, \tag{1.17}$$

(ii) for every $n \in \mathbb{N}_0$ there exists a unique $v_n\colon [0,T)\times O \to \mathbb{R}^m$ which satisfies for all $t \in [0,T)$, $x \in O$ that

$$v_n(t,x) = \mathbb{E}\left[g_n(X^x_{t,T})Z^x_{t,T} + \int_t^T h_n(r,X^x_{t,r})Z^x_{t,r}\,\mathrm{d}r\right], \tag{1.18}$$

(iii) it holds that

$$\limsup_{n\to\infty}\left[\sup_{t\in[0,T)}\sup_{x\in O}\left(\frac{|||v_n(t,x)-v_0(t,x)|||}{V(t,x)}\sqrt{T-t}\right)\right] = 0, \tag{1.19}$$

and

(iv) it holds for every compact set $\mathcal{K} \subseteq [0,T)\times O$ that

$$\limsup_{n\to\infty}\left[\sup_{(t,x)\in\mathcal{K}}\left(|||v_n(t,x)-v_0(t,x)|||\sqrt{T-t}\right)\right] = 0. \tag{1.20}$$

Proof of Lemma 1.1.4. First note that for all $r \in (0,\infty)$ it holds that K_r and O_r are compact sets. Combining this with the fact that for all $n \in \mathbb{N}_0$ it holds that $O \ni x \mapsto \frac{g_n(x)}{V(T,x)} \in \mathbb{R}$ and $[0,T)\times O \ni (t,x) \mapsto \frac{h_n(t,x)}{V(t,x)}\sqrt{T-t} \in \mathbb{R}$ are continuous ensures that for all $n \in \mathbb{N}_0$, $r \in (0,\infty)$ it holds that

$$\sup_{t\in K_r}\sup_{x\in O_r}\left[\frac{|g_n(x)|}{V(T,x)} + \frac{|h_n(t,x)|}{V(t,x)}\sqrt{T-t}\right] < \infty. \tag{1.21}$$

The assumption that for all $n \in \mathbb{N}$ it holds that $\inf_{r\in(0,\infty)}[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}(\frac{|g_n(x)|}{V(T,x)} + \frac{|h_n(t,x)|}{V(t,x)}\sqrt{T-t})] = 0$ hence implies that for all $n \in \mathbb{N}$ it holds that

$$\sup_{t\in[0,T)}\sup_{x\in O}\left[\frac{|g_n(x)|}{V(T,x)} + \frac{|h_n(t,x)|}{V(t,x)}\sqrt{T-t}\right] < \infty. \tag{1.22}$$

This and (1.16) establish item (i). Moreover, note that combining item (i) with Lemma 1.1.3 proves item (ii). Next observe that the hypothesis that for all $t \in [0,T)$, $s \in (t,T]$, $x \in O$ it holds that $\mathbb{E}[V(s, X^x_{t,s})|||Z^x_{t,s}|||] \leqslant \frac{c}{\sqrt{s-t}} V(t,x)$ implies that for all $n \in \mathbb{N}$, $t \in [0,T)$, $x \in O$ it holds that

$$\begin{aligned}
&\frac{\mathbb{E}\left[|||g_n(X^x_{t,T})Z^x_{t,T} - g_0(X^x_{t,T})Z^x_{t,T}|||\right]}{V(t,x)} \\
&= \mathbb{E}\left[\frac{|g_n(X^x_{t,T}) - g_0(X^x_{t,T})|}{V(T, X^x_{t,T})} \cdot \frac{V(T, X^x_{t,T})|||Z^x_{t,T}|||}{V(t,x)}\right] \\
&\leqslant \left[\sup_{y \in O}\left(\frac{|g_n(y) - g_0(y)|}{V(T,y)}\right)\right] \frac{\mathbb{E}\left[V(T, X^x_{t,T})|||Z^x_{t,T}|||\right]}{V(t,x)} \\
&\leqslant \sup_{y \in O}\left(\frac{|g_n(y) - g_0(y)|}{V(T,y)}\right) \frac{c}{\sqrt{T-t}}.
\end{aligned} \tag{1.23}$$

This and (1.16) show that

$$\begin{aligned}
&\limsup_{n\to\infty}\left[\sup_{t\in[0,T)}\sup_{x\in O}\left(\frac{|||\mathbb{E}\left[g_n(X^x_{t,T})Z^x_{t,T}\right] - \mathbb{E}\left[g_0(X^x_{t,T})Z^x_{t,T}\right]|||}{V(t,x)}\sqrt{T-t}\right)\right] \\
&= 0.
\end{aligned} \tag{1.24}$$

In addition, observe that Fubini's theorem and the assumption that for all $t \in [0,T)$, $s \in (t,T]$, $x \in O$ it holds that $\mathbb{E}[V(s, X^x_{t,s})|||Z^x_{t,s}|||] \leqslant \frac{c}{\sqrt{s-t}} V(t,x)$ ensure that for all $n \in \mathbb{N}$, $t \in [0,T)$, $x \in O$ it holds that

$$\begin{aligned}
&\frac{\mathbb{E}\left[\int_t^T |||h_n(r, X^x_{t,r})Z^x_{t,r} - h_0(r, X^x_{t,r})Z^x_{t,r}|||\,\mathrm{d}r\right]}{V(t,x)} \\
&= \int_t^T \mathbb{E}\left[\frac{|h_n(r, X^x_{r,t}) - h_0(r, X^x_{t,r})|}{V(r, X^x_{t,r})} \cdot \frac{\sqrt{T-r}}{\sqrt{T-r}} \cdot \frac{V(r, X^x_{t,r})\,|||Z^x_{t,r}|||}{V(t,x)}\right] \mathrm{d}r \\
&\leqslant \int_t^T \left[\sup_{s\in[0,T)}\sup_{y\in O}\left(\frac{|h_n(s,y) - h_0(s,y)|}{V(s,y)}\sqrt{T-s}\right)\right] \frac{\mathbb{E}\left[V(r, X^x_{t,r})\,|||Z^x_{t,r}|||\right]}{\sqrt{T-r}\,V(t,x)}\,\mathrm{d}r \\
&\leqslant \left[\sup_{s\in[0,T)}\sup_{y\in O}\left(\frac{|h_n(s,y) - h_0(s,y)|}{V(s,y)}\sqrt{T-s}\right)\right] \int_t^T \frac{c}{\sqrt{(T-r)(r-t)}}\,\mathrm{d}r.
\end{aligned} \tag{1.25}$$

Item (i) of Lemma 1.1.2 (applied for every $t \in [0,T)$ with $a \curvearrowleft t$, $b \curvearrowleft T$, $\lambda \curvearrowleft 0$ in the notation of Lemma 1.1.2) hence demonstrates that for all $n \in \mathbb{N}$, $t \in [0,T)$, $x \in O$ it holds that

$$\begin{aligned}
&\frac{\mathbb{E}\left[\int_t^T |||h_n(r, X^x_{t,r})Z^x_{t,r} - h_0(r, X^x_{t,r})Z^x_{t,r}|||\,\mathrm{d}r\right]}{V(t,x)} \\
&\leqslant c\pi\left[\sup_{s\in[0,T)}\sup_{y\in O}\left(\frac{|h_n(s,y) - h_0(s,y)|}{V(s,y)}\sqrt{T-s}\right)\right].
\end{aligned} \tag{1.26}$$

Combining this with (1.16) shows that

$$\limsup_{n\to\infty}\left[\sup_{t\in[0,T)}\sup_{x\in O}\left(\frac{|||\mathbb{E}[\int_t^T h_n(r,X^x_{t,r})Z^x_{t,s}\,\mathrm{d}r] - \mathbb{E}[\int_t^T h_0(r,X^x_{t,r})Z^x_{t,r}\,\mathrm{d}r]|||}{V(t,x)}\sqrt{T-t}\right)\right] = 0. \tag{1.27}$$

The triangle inequality, item (ii), and (1.24) hence ensure that

$$\limsup_{n\to\infty}\left[\sup_{t\in[0,T)}\sup_{x\in O}\left(\frac{|||v_n(t,x)-v_0(t,x)|||}{V(t,x)}\sqrt{T-t}\right)\right]=0. \tag{1.28}$$

This establishes item (iii). Next observe that item (iii) and the hypothesis that $V\colon[0,T]\times O\to(0,\infty)$ is continuous demonstrate that for all $\mathcal{K}\subseteq[0,T)\times O$ compact it holds that

$$\begin{aligned}&\limsup_{n\to\infty}\left[\sup_{(t,x)\in\mathcal{K}}\left(|||v_n(t,x)-v_0(t,x)|||\sqrt{T-t}\right)\right]\\&\leqslant\limsup_{n\to\infty}\left[\sup_{(t,x)\in\mathcal{K}}\left(\frac{|||v_n(t,x)-v_0(t,x)|||}{V(t,x)}\sqrt{T-t}\right)\right]\left[\sup_{(t,x)\in\mathcal{K}}V(t,x)\right]=0.\end{aligned} \tag{1.29}$$

This establishes item (iv). The proof of Lemma 1.1.4 is thus complete. □

The next lemma shows the construction of an approximating series of compactly supported, continuous functions for a certain class of continuous functions. Lemma 1.1.5 is a generalization of the results in [13, Lemma 2.3] to our setting on $[0,T)$ instead of $[0,T]$ and with the $\sqrt{T-t}$ term in (1.30).

Lemma 1.1.5. *Let $d\in\mathbb{N}$, $T\in(0,\infty)$, let $\|\cdot\|\colon\mathbb{R}^d\to[0,\infty)$ be a norm, let $O\subseteq\mathbb{R}^d$ be a non-empty open set, for every $r\in(0,\infty)$ let $K_r\subseteq[0,T)$, $O_r\subseteq O$ satisfy $K_r=[0,\max\{T-\frac{1}{r},0\}]$ and $O_r=\{x\in O\colon\|x\|\leqslant r$ and $\{y\in\mathbb{R}^d\colon\|y-x\|<\frac{1}{r}\}\subseteq O\}$, and let $h\in C([0,T)\times O,\mathbb{R})$ satisfy $\inf_{r\in(0,\infty)}[\sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}(|h(t,x)|\sqrt{T-t})]=0$. Then there exist compactly supported $h_n\in C([0,T)\times O,\mathbb{R})$, $n\in\mathbb{N}$, which satisfy that*

$$\limsup_{n\to\infty}\left[\sup_{t\in[0,T)}\sup_{x\in O}\left(|h_n(t,x)-h(t,x)|\sqrt{T-t}\right)\right]=0. \tag{1.30}$$

Proof of Lemma 1.1.5. Throughout this proof let $U_n\subseteq[0,T)\times O$, $n\in\mathbb{N}$, satisfy for all $n\in\mathbb{N}$ that $U_n=\{(t,x)\in[0,T)\times O\colon(\exists(s,y)\in K_n\times O_n\colon\max\{|s-t|,\|y-x\|\}<\frac{1}{2n})\}$. Observe that for every $n\in\mathbb{N}$ it holds that $K_n\times O_n\subseteq[0,T)\times O$ is a compact set, $U_n\subseteq[0,T)\times O$ is an open set, and $K_n\times O_n\subseteq U_n$. Urysohn's lemma (cf., e.g., [92, Lemma 2.12]) therefore demonstrates that for all $n\in\mathbb{N}$ there exists $\varphi_n\in C([0,T)\times O,\mathbb{R})$ which satisfies for all $t\in[0,T)$, $x\in O$ that $\mathbb{1}_{K_n\times O_n}(t,x)\leqslant\varphi_n(t,x)\leqslant\mathbb{1}_{U_n}(t,x)$. Let $h_n\colon[0,T)\times O\to\mathbb{R}$, $n\in\mathbb{N}$, satisfy for all $n\in\mathbb{N}$, $t\in[0,T)$, $x\in O$ that $h_n(t,x)=\varphi(t,x)h(t,x)$. This and the fact that for all $n\in\mathbb{N}$ it holds that φ_n has compact support implies that for all $n\in\mathbb{N}$ it holds that $h_n\colon[0,T)\times O\to\mathbb{R}$ is a continuous function with compact support. In addition, observe that

$$\begin{aligned}&\limsup_{n\to\infty}\left[\sup_{t\in[0,T)}\sup_{x\in O}\left(|h_n(t,x)-h(t,x)|\sqrt{T-t}\right)\right]\\&=\limsup_{n\to\infty}\left[\sup_{t\in[0,T)}\sup_{x\in O}\left([\varphi_n(t,x)-1]|h(t,x)|\sqrt{T-t}\right)\right]\\&\leqslant\limsup_{n\to\infty}\left[\sup_{t\in[0,T)\setminus K_n}\sup_{x\in O\setminus O_n}\left(|h(t,x)|\sqrt{T-t}\right)\right]=0.\end{aligned} \tag{1.31}$$

The proof of Lemma 1.1.5 is thus complete. □

The following corollary is a consequence of Lemma 1.1.5 and proves for a certain class of continuous functions the existence of a series of compactly supported, continuous functions which satisfies a specific approximation property. Corollary 1.1.6 is a generalization of the results in [13, Corollary 2.4].

Corollary 1.1.6. *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, let $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ be a norm, let $O \subseteq \mathbb{R}^d$ be a non-empty open set, for every $r \in (0,\infty)$ let $K_r \subseteq [0,T)$, $O_r \subseteq O$ satisfy $K_r = [0, \max\{T - \frac{1}{r}, 0\}]$ and $O_r = \{x \in O \colon \|x\| \leqslant r$ and $\{y \in \mathbb{R}^d \colon \|y - x\| < \frac{1}{r}\} \subseteq O\}$, let $h \in C([0,T) \times O, \mathbb{R})$, $V \in C([0,T] \times O, (0,\infty))$, and assume that $\inf_{r\in(0,\infty)}[\sup_{t\in[0,T)\backslash K_r} \sup_{x\in O\backslash O_r}(\frac{|h(t,x)|}{V(t,x)}\sqrt{T-t})] = 0$. Then there exist compactly supported $h_n \in C([0,T) \times O, \mathbb{R})$, $n \in \mathbb{N}$, which satisfy that*

$$\limsup_{n\to\infty}\left[\sup_{t\in[0,T)}\sup_{x\in O}\left(\frac{|h_n(t,x) - h(t,x)|}{V(t,x)}\sqrt{T-t}\right)\right] = 0. \tag{1.32}$$

Proof of Corollary 1.1.6. Throughout this proof let $g\colon [0,T) \times O \to \mathbb{R}$ satisfy for all $t \in [0,T)$, $x \in O$ that $g(t,x) = \frac{h(t,x)}{V(t,x)}$. Note that the fact that $h \in C([0,T) \times O, \mathbb{R})$ and $V \in C([0,T] \times O, (0,\infty))$ and the assumption that $\inf_{r\in(0,\infty)}[\sup_{t\in[0,T)\backslash K_r} \sup_{x\in O\backslash O_r}(\frac{|h(t,x)|}{V(t,x)}\sqrt{T-t})] = 0$ ensure that $g \in C([0,T) \times O, \mathbb{R})$ and

$$\inf\nolimits_{r\in(0,\infty)}\left[\sup\nolimits_{t\in[0,T)\backslash K_r}\sup\nolimits_{x\in O\backslash O_r}\left(|g(t,x)|\sqrt{T-t}\right)\right] = 0. \tag{1.33}$$

Lemma 1.1.5 (applied with $h \curvearrowleft g$ in the notation of Lemma 1.1.5) hence demonstrates that there exist compactly supported $g_n \in C([0,T) \times O, \mathbb{R})$, $n \in \mathbb{N}$, which satisfy that

$$\limsup\nolimits_{n\to\infty}\left[\sup\nolimits_{t\in[0,T)}\sup\nolimits_{x\in O}\left(|g_n(t,x) - g(t,x)|\sqrt{T-t}\right)\right] = 0. \tag{1.34}$$

In the next step let $h_n\colon [0,T) \times O \to \mathbb{R}$, $n \in \mathbb{N}$, satisfy for all $n \in \mathbb{N}$, $t \in [0,T)$, $x \in O$ that $h_n(t,x) = g_n(t,x)V(t,x)$. Observe that (1.34) demonstrates that for all $n \in \mathbb{N}$ it holds that $h_n \in C([0,T) \times O, \mathbb{R})$ and

$$\begin{aligned}
&\limsup_{n\to\infty}\left[\sup_{t\in[0,T)}\sup_{x\in O}\left(\frac{|h_n(t,x) - h(t,x)|}{V(t,x)}\sqrt{T-t}\right)\right] \\
&= \limsup_{n\to\infty}\left[\sup_{t\in[0,T)}\sup_{x\in O}\left(|g_n(t,x) - g(t,x)|\sqrt{T-t}\right)\right] = 0.
\end{aligned} \tag{1.35}$$

The proof of Corollary 1.1.6 is thus complete. □

1.1.3 Continuity of solutions of SFPEs

The following lemma shows well-definedness and continuity of certain SFPE solutions under the assumption that the terminal condition and the nonlinearity are bounded. Lemma 1.1.7 is a generalization of the results in [13, Lemma 2.5].

Lemma 1.1.7. *Assume Setting 1.1.1, let $g \in C(O, \mathbb{R})$, $h \in C([0,T) \times O, \mathbb{R})$ be bounded, assume $\inf_{t\in[0,T]}\inf_{x\in O} V(t,x) > 0$, and assume for all $\varepsilon \in (0,\infty)$, $t \in [0,T)$, $s \in (t,T]$, $x \in O$ that*

$$\limsup\nolimits_{[0,s)\times O \ni (u,y)\to(t,x)}\left[\mathbb{P}(\|X^y_{u,s} - X^x_{t,s}\| > \varepsilon) + \mathbb{E}[\|Z^y_{u,s} - Z^x_{t,s}\|]\right] = 0. \tag{1.36}$$

Then

$$[0,T) \times O \ni (t,x) \mapsto \mathbb{E}\left[g(X^x_{t,T})Z^x_{t,T} + \int_t^T h(r, X^x_{t,r})Z^x_{t,r}\,dr\right] \in \mathbb{R}^m \tag{1.37}$$

is well-defined and continuous.

Proof of Lemma 1.1.7. Throughout this proof let $M \in [0, \infty)$ satisfy that

$$M = \sup_{t\in[0,T)} \sup_{x\in O} [|g(x)| + |h(t,x)|]. \tag{1.38}$$

Note that this and the assumption that $\inf_{t\in[0,T]} \inf_{x\in O} V(t,x) > 0$ ensure that

$$\sup_{t\in[0,T)} \sup_{x\in O} \left[\frac{|g(x)|}{V(T,x)} + \frac{|h(t,x)|}{V(t,x)} \sqrt{T-t} \right] \leqslant \frac{M(1+\sqrt{T})}{\inf_{t\in[0,T]} \inf_{x\in O} V(t,x)} < \infty. \tag{1.39}$$

Lemma 1.1.3 and the fact that for all $t \in [0,T)$, $s \in (t,T]$, $x \in O$ it holds that $\mathbb{E}[V(s, X^x_{t,s}) \|Z^x_{t,s}\|] \leqslant \frac{c}{\sqrt{s-t}} V(t,x)$ hence demonstrate that for all $t \in [0,T)$, $x \in O$ it holds that

$$\mathbb{E}\left[|g(X^x_{t,T})| \|Z^x_{t,T}\| + \int_t^T |h(r, X^x_{t,r})| \|Z^x_{t,r}\| \, dr \right] < \infty. \tag{1.40}$$

This proves that (1.37) is well-defined. In the next step for every $t \in [0,T]$, $x \in O$ let $Z^x_{t,t} = 0$, $h(T,x) = g(x)$, and let $(t_n, x_n) \in [0,T) \times O$, $n \in \mathbb{N}_0$, satisfy $\limsup_{n\to\infty} [|t_n - t_0| + \|x_n - x_0\|] = 0$. Note that the fact that $\limsup_{n\to\infty}[|t_n - t_0| + \|x_n - x_0\|] = 0$ ensures that there exists $q \in (0,\infty)$ which satisfies for all $n \in \mathbb{N}_0$ that $(t_n, x_n) \in K_q \times O_q$. Furthermore, observe that the assumption that $\sup_{t\in[0,T)} \sup_{x\in O} |h(t,x)| \leqslant M$ proves that for all $n \in \mathbb{N}$, $s \in [0,T)$ it holds that

$$\begin{aligned}
&\mathbb{E}\left[\|h(s, X^{x_n}_{t_n,\max\{s,t_n\}}) Z^{x_n}_{t_n,\max\{s,t_n\}} - h(s, X^{x_0}_{t_0,\max\{s,t_0\}}) Z^{x_0}_{t_0,\max\{s,t_0\}}\| \right] \\
&\leqslant \mathbb{E}\left[\left\| h(s, X^{x_n}_{t_n,\max\{s,t_n\}}) Z^{x_n}_{t_n,\max\{s,t_n\}} - h(s, X^{x_n}_{t_n,\max\{s,t_n\}}) Z^{x_0}_{t_0,\max\{s,t_0\}} \right\| \right] \\
&\quad + \mathbb{E}\left[\left\| h(s, X^{x_n}_{t_n,\max\{s,t_n\}}) Z^{x_0}_{t_0,\max\{s,t_0\}} - h(s, X^{x_0}_{t_0,\max\{s,t_0\}}) Z^{x_0}_{t_0,\max\{s,t_0\}} \right\| \right] \\
&\leqslant \mathbb{E}\left[|h(s, X^{x_n}_{t_n,\max\{s,t_n\}})| \left\| Z^{x_n}_{t_n,\max\{s,t_n\}} - Z^{x_0}_{t_0,\max\{s,t_0\}} \right\| \right] \\
&\quad + \mathbb{E}\left[\left\| h(s, X^{x_n}_{t_n,\max\{s,t_n\}}) Z^{x_0}_{t_0,\max\{s,t_0\}} - h(s, X^{x_0}_{t_0,\max\{s,t_0\}}) Z^{x_0}_{t_0,\max\{s,t_0\}} \right\| \right] \\
&\leqslant M \mathbb{E}\left[\left\| Z^{x_n}_{t_n,\max\{s,t_n\}} - Z^{x_0}_{t_0,\max\{s,t_0\}} \right\| \right] \\
&\quad + \mathbb{E}\left[\left\| h(s, X^{x_n}_{t_n,\max\{s,t_n\}}) Z^{x_0}_{t_0,\max\{s,t_0\}} - h(s, X^{x_0}_{t_0,\max\{s,t_0\}}) Z^{x_0}_{t_0,\max\{s,t_0\}} \right\| \right].
\end{aligned} \tag{1.41}$$

Next note that the fact that $\sup_{t\in[0,T)} \sup_{x\in O} |h(t,x)| \leqslant M$ ensures that for all $n \in \mathbb{N}$, $s \in [0,T)$ it holds a.s. that

$$\begin{aligned}
&\left\| h(s, X^{x_n}_{t_n,\max\{s,t_n\}}) Z^{x_0}_{t_0,\max\{s,t_0\}} - h(s, X^{x_0}_{t_0,\max\{s,t_0\}}) Z^{x_0}_{t_0,\max\{s,t_0\}} \right\| \\
&\leqslant 2M \left\| Z^{x_0}_{t_0,\max\{s,t_0\}} \right\|.
\end{aligned} \tag{1.42}$$

Moreover, observe that the hypothesis that for all $t \in [0,T)$, $s \in (t,T]$, $x \in O$ it holds that $\mathbb{E}[V(s, X^x_{t,s}) \|Z^x_{t,s}\|] \leqslant \frac{c}{\sqrt{s-t}} V(t,x)$ and the assumption that $\inf_{t\in[0,T]} \inf_{x\in O} V(t,x) > 0$ show that for all $n \in \mathbb{N}$, $s \in (t_0, T]$ it holds that

$$\begin{aligned}
\mathbb{E}\left[\|Z^{x_0}_{t_0,s}\| \right] &= \mathbb{E}\left[\frac{V(s, X^{x_0}_{t_0,s})}{V(s, X^{x_0}_{t_0,s})} \|Z^{x_0}_{t_0,s}\| \right] \\
&\leqslant \left(\sup_{r\in[0,T]} \sup_{y\in O} \frac{1}{V(r,y)} \right) \mathbb{E}\left[V(s, X^{x_0}_{t_0,s}) \|Z^{x_0}_{t_0,s}\| \right] \\
&\leqslant \frac{1}{[\inf_{r\in[0,T]} \inf_{y\in O} V(r,y)]} \frac{c}{\sqrt{s-t_0}} V(t_0, x_0) < \infty.
\end{aligned} \tag{1.43}$$

Lebesgue's dominated convergence theorem, (1.36), (1.41), (1.42), and the fact that $h \in C([0,T) \times O, \mathbb{R})$ hence demonstrate that for all $s \in [0,T)$ it holds that

$$\limsup_{n\to\infty} \mathbb{E}\left[\left\|h(s, X^{x_n}_{t_n,\max\{s,t_n\}}) Z^{x_n}_{t_n,\max\{s,t_n\}} - h(s, X^{x_0}_{t_0,\max\{s,t_0\}}) Z^{x_0}_{t_0,\max\{s,t_0\}}\right\|\right] = 0. \tag{1.44}$$

Moreover, observe that (1.41) and (1.42) imply that for all $n \in \mathbb{N}$, $s \in [0,T)$ it holds that

$$\begin{aligned}&\mathbb{E}\left[\left\|h(s, X^{x_n}_{t_n,\max\{s,t_n\}}) Z^{x_n}_{t_n,\max\{s,t_n\}} - h(s, X^{x_0}_{t_0,\max\{s,t_0\}}) Z^{x_0}_{t_0,\max\{s,t_0\}}\right\|\right] \\ &\leqslant M\mathbb{E}\left[\left\|Z^{x_n}_{t_n,\max\{s,t_n\}} - Z^{x_0}_{t_0,\max\{s,t_0\}}\right\|\right] + 2M\mathbb{E}\left[\left\|Z^{x_0}_{t_0,\max\{s,t_0\}}\right\|\right].\end{aligned} \tag{1.45}$$

The dominated convergence theorem, (1.36), (1.43), and (1.44) therefore show that

$$\begin{aligned}\limsup_{n\to\infty} \bigg(\int_{t_0}^{T} \mathbb{E}\Big[\Big\| & h(r, X^{x_n}_{t_n,\max\{r,t_n\}}) Z^{x_n}_{t_n,\max\{r,t_n\}} \\ & - h(r, X^{x_0}_{t_0,\max\{r,t_0\}}) Z^{x_0}_{t_0,\max\{r,t_0\}}\Big\|\Big] \mathrm{d}r \bigg) = 0.\end{aligned} \tag{1.46}$$

In the next step note that for all $k, n \in \mathbb{N}_0$ with $t_n \leqslant t_k$ it holds that

$$\begin{aligned}&\left\|\mathbb{E}\left[\int_{t_n}^{T} h(r, X^{x_n}_{t_n,r}) Z^{x_n}_{t_n,r}\,\mathrm{d}r\right] - \mathbb{E}\left[\int_{t_k}^{T} h(r, X^{x_k}_{t_k,r}) Z^{x_k}_{t_k,r}\,\mathrm{d}r\right]\right\| \\ &= \bigg\|\mathbb{E}\left[\int_{t_n}^{t_k} h(r, X^{x_n}_{t_n,r}) Z^{x_n}_{t_n,r}\,\mathrm{d}r\right] \\ &\quad + \mathbb{E}\left[\int_{t_k}^{T} \left(h(r, X^{x_n}_{t_n,r}) Z^{x_n}_{t_n,r} - h(r, X^{x_k}_{t_k,r}) Z^{x_k}_{t_k,r}\right) \mathrm{d}r\right]\bigg\| \\ &\leqslant \mathbb{E}\left[\int_{t_n}^{t_k} |h(r, X^{x_n}_{t_n,r})|\, \|Z^{x_n}_{t_n,r}\|\,\mathrm{d}r\right] \\ &\quad + \mathbb{E}\left[\int_{t_k}^{T} \left\|h(r, X^{x_n}_{t_n,r}) Z^{x_n}_{t_n,r} - h(r, X^{x_k}_{t_k,r}) Z^{x_k}_{t_k,r}\right\| \mathrm{d}r\right].\end{aligned} \tag{1.47}$$

Furthermore, observe that Fubini's theorem, the fact that $\sup_{t\in[0,T)}\sup_{x\in O}|h(t,x)| \leqslant M$, and the assumption that for all $t \in [0,T)$, $s \in (t,T]$, $x \in O$ it holds that $\mathbb{E}[V(s, X^x_{t,s})\|Z^x_{t,s}\|] \leqslant \frac{c}{\sqrt{s-t}} V(t,x)$ ensure that for all $k, n \in \mathbb{N}_0$ with $t_n \leqslant t_k$ it holds that

$$\begin{aligned}&\mathbb{E}\left[\int_{t_n}^{t_k} |h(r, X^{x_n}_{t_n,r})| \|Z^{x_n}_{t_n,r}\|\,\mathrm{d}r\right] \leqslant M \int_{t_n}^{t_k} \mathbb{E}\left[\frac{V(r, X^{x_n}_{t_n,r})}{V(r, X^{x_n}_{t_n,r})} \|Z^{x_n}_{t_n,r}\|\right] \mathrm{d}r \\ &\leqslant \frac{M}{\inf_{t\in[0,T]}\inf_{x\in O} V(t,x)} \int_{t_n}^{t_k} \mathbb{E}\left[V(r, X^{x_n}_{t_n,r})\|Z^{x_n}_{t_n,r}\|\right] \mathrm{d}r \\ &\leqslant \frac{M}{\inf_{t\in[0,T]}\inf_{x\in O} V(t,x)} \int_{t_n}^{t_k} V(t_n, x_n) \frac{c}{\sqrt{r-t_n}}\,\mathrm{d}r \\ &\leqslant \frac{cM}{\inf_{t\in[0,T]}\inf_{x\in O} V(t,x)} \left[\sup_{u\in K_q}\sup_{y\in O_q} V(u,y)\right] \int_0^{t_k - t_n} \frac{1}{\sqrt{z}}\,\mathrm{d}z.\end{aligned} \tag{1.48}$$

Combining this with (1.46), (1.47), and the fact that $\lim_{\varepsilon\to 0}\int_0^{|\varepsilon|} \frac{1}{\sqrt{z}}\,\mathrm{d}z = 0$ proves that

$$\begin{aligned}&\limsup_{n\to\infty} \left\|\mathbb{E}\left[\int_{t_n}^{T} h(r, X^{x_n}_{t_n,r}) Z^{x_n}_{t_n,r}\,\mathrm{d}r\right] - \mathbb{E}\left[\int_{t_0}^{T} h(r, X^{x_0}_{t_0,r}) Z^{x_0}_{t_0,r}\,\mathrm{d}r\right]\right\| \\ &\leqslant \limsup_{n\to\infty} \bigg(\frac{cM}{\inf_{t\in[0,T]}\inf_{x\in O} V(t,x)} \left[\sup_{u\in K_q}\sup_{y\in O_q} V(u,y)\right] \int_0^{|t_0-t_n|} \frac{1}{\sqrt{z}}\,\mathrm{d}z \\ &\quad + \mathbb{E}\left[\int_{\max\{t_0,t_n\}}^{T} \left\|h(r, X^{x_n}_{t_n,r}) Z^{x_n}_{t_n,r} - h(r, X^{x_0}_{t_0,r}) Z^{x_0}_{t_0,r}\right\| \mathrm{d}r\right]\bigg) = 0.\end{aligned} \tag{1.49}$$

This and (1.44) (applied with $h \curvearrowleft ([0,T) \times O \ni (t,x) \mapsto g(x) \in \mathbb{R})$) show that

$$\limsup_{n\to\infty} \bigg(\mathbb{E}\Big[|||g(X^{x_n}_{t_n,T}) Z^{x_n}_{t_n,T} - g(X^{x_0}_{t_0,T}) Z^{x_0}_{t_0,T}||| \Big] + \bigg|\bigg|\bigg| \mathbb{E}\bigg[\int_{t_n}^T h(r, X^{x_n}_{t_n,r}) Z^{x_n}_{t_n,r}\,\mathrm{d}r \bigg] - \mathbb{E}\bigg[\int_{t_0}^T h(r, X^{x_0}_{t_0,r}) Z^{x_0}_{t_0,r}\,\mathrm{d}r \bigg] \bigg|\bigg|\bigg| \bigg) = 0. \tag{1.50}$$

This proves that (1.37) is continuous. The proof of Lemma 1.1.7 is thus complete. □

The next lemma extends the results of Lemma 1.1.7 by demonstrating well-definedness and continuity of certain SFPE solutions for continuous terminal conditions and nonlinearities. For this, we use the approximation results proven in Corollary 1.1.6. Lemma 1.1.8 is a generalization of the results in [13, Lemma 2.6].

Lemma 1.1.8. *Assume Setting 1.1.1, assume* $\inf_{t\in[0,T]} \inf_{x\in O} V(t,x) > 0$*, assume for all* $\varepsilon \in (0,\infty)$, $t \in [0,T)$, $s \in (t,T]$, $x \in O$ *that*

$$\limsup\nolimits_{[0,s)\times O \ni (u,y)\to(t,x)} \Big[\mathbb{P}\big(\|X^y_{u,s} - X^x_{t,s}\| > \varepsilon \big) + \mathbb{E}\big[|||Z^y_{u,s} - Z^x_{t,s}||| \big] \Big] = 0, \tag{1.51}$$

let $g \in C(O,\mathbb{R})$, $h \in C([0,T)\times O,\mathbb{R})$, *and* $v\colon [0,T)\times O \to \mathbb{R}^m$ *satisfy* $\inf_{r\in(0,\infty)} [\sup_{t\in[0,T)\setminus K_r} \sup_{x\in O\setminus O_r} (\frac{|g(x)|}{V(T,x)} + \frac{|h(t,x)|}{V(t,x)}\sqrt{T-t})] = 0$ *and for all* $t \in [0,T)$, $x \in O$ *that*

$$v(t,x) = \mathbb{E}\bigg[g(X^x_{t,T}) Z^x_{t,T} + \int_t^T h(r, X^x_{t,r}) Z^x_{t,r}\,\mathrm{d}r \bigg] \tag{1.52}$$

(cf. Lemma 1.1.3). Then

(i) it holds that $v \in C([0,T)\times O, \mathbb{R}^m)$ *and*

(ii) if - in addition to the above assumptions - it holds for all $u \in (0,\infty)$ *that*

$$\limsup_{r\to\infty} \left[\sup_{t\in[0,T)\setminus K_r} \sup_{x\in O\setminus O_r} \left(\frac{\mathbb{E}\Big[\int_t^T \mathbb{1}_{O_u}(X^x_{t,s})\, |||Z^x_{t,s}|||\,(\mathrm{d}s + \delta_T(\mathrm{d}s)) \Big] \sqrt{T-t}}{V(t,x)} \right) \right] = 0, \tag{1.53}$$

then it holds that

$$\lim_{r\to\infty} \sup_{t\in[0,T)\setminus K_r} \sup_{x\in O\setminus O_r} \left(\frac{|||v(t,x)|||}{V(t,x)} \sqrt{T-t} \right) = 0. \tag{1.54}$$

Proof of Lemma 1.1.8. First observe that Corollary 1.1.6 and the assumption that $\inf_{r\in(0,\infty)} [\sup_{t\in[0,T)\setminus K_r} \sup_{x\in O\setminus O_r} (\frac{|g(x)|}{V(T,x)} + \frac{|h(t,x)|}{V(t,x)}\sqrt{T-t})] = 0$ demonstrate that there exists compactly supported $g_n \in C(O,\mathbb{R})$, $n \in \mathbb{N}$, and $h_n \in C([0,T)\times O,\mathbb{R})$, $n \in \mathbb{N}$, which satisfy

$$\limsup_{n\to\infty} \left[\sup_{t\in[0,T)} \sup_{x\in O} \left(\frac{|g_n(x) - g(x)|}{V(T,x)} + \frac{|h_n(t,x) - h(t,x)|}{V(t,x)} \sqrt{T-t} \right) \right] = 0. \tag{1.55}$$

For every $n \in \mathbb{N}$ let $v_n\colon [0,T)\times O \to \mathbb{R}^m$ satisfy for all $t \in [0,T)$, $x \in O$ that

$$v_n(t,x) = \mathbb{E}\bigg[g_n(X^x_{t,T}) Z^x_{t,T} + \int_t^T h_n(s, X^x_{t,s}) Z^x_{t,s}\,\mathrm{d}s \bigg] \tag{1.56}$$

(cf. Lemma 1.1.3). Observe that Lemma 1.1.7, (1.51), and the fact that for all $n \in \mathbb{N}$ it holds that g_n and h_n are compactly supported and continuous show that for all $n \in \mathbb{N}$

it holds that $v_n\colon [0,T)\times O \to \mathbb{R}^m$ is continuous. Furthermore, note that the fact that $g_n\colon O \to \mathbb{R}$, $n \in \mathbb{N}$, and $h_n\colon [0,T)\times O \to \mathbb{R}$, $n \in \mathbb{N}$, are compactly supported ensures that for every $n \in \mathbb{N}$ there exists $r_n \in (0,\infty)$ which satisfies that for all $r \in [r_n,\infty)$, $t \in [0,T)\backslash K_r$, $x \in O\backslash O_r$ it holds that $g_n(x) = 0 = h_n(t,x)$. This demonstrates that for all $n \in \mathbb{N}$ it holds that

$$\inf_{r\in(0,\infty)}\left[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}\left(\frac{|g_n(x)|}{V(T,x)}+\frac{|h_n(t,x)|}{V(t,x)}\sqrt{T-t}\right)\right]=0. \tag{1.57}$$

Item (iv) of Lemma 1.1.4, (1.55), and the fact that for all $n \in \mathbb{N}$ it holds that $v_n\colon [0,T)\times O \to \mathbb{R}^m$ is continuous therefore imply that $v\colon [0,T)\times O \to \mathbb{R}^m$ is continuous. This establishes item (i). To prove item (ii) assume for all $u \in (0,\infty)$ that

$$\limsup_{r\to\infty}\left[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}\left(\frac{\mathbb{E}[\int_t^T \mathbb{1}_{O_u}(X^x_{t,s})\,\|Z^x_{t,s}\|(\mathrm{d}s+\delta_T(\mathrm{d}s))]\sqrt{T-t}}{V(t,x)}\right)\right]=0. \tag{1.58}$$

This and the fact that for all $n \in \mathbb{N}$, $x \in O\backslash O_{r_n}$ it holds that $g_n(x) = 0$ ensure that for all $n \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\limsup_{r\to\infty}\left[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}\left(\frac{\mathbb{E}[|g_n(X^x_{t,T})|\,\|Z^x_{t,T}\|]}{V(t,x)}\sqrt{T-t}\right)\right]\\
&\leq \limsup_{r\to\infty}\left[\left[\sup_{y\in O_{r_n}}|g_n(y)|\right]\right.\\
&\quad\left.\cdot\left[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}\left(\frac{\mathbb{E}\left[\mathbb{1}_{O_{r_n}}(X^x_{t,T})\|Z^x_{t,T}\|\right]\sqrt{T-t}}{V(t,x)}\right)\right]\right]=0.
\end{aligned} \tag{1.59}$$

Moreover, note that (1.58) and the fact that for all $n \in \mathbb{N}$, $t \in [0,T)\backslash K_{r_n}$, $x \in O\backslash O_{r_n}$ it holds that $h_n(t,x) = 0$ demonstrate that for all $n \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\limsup_{r\to\infty}\left[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}\left(\frac{\mathbb{E}\Big[\int_t^T|h_n(s,X^x_{t,s})|\,\|Z^x_{t,s}\|\,\mathrm{d}s\Big]\sqrt{T-t}}{V(t,x)}\right)\right]\\
&\leq \limsup_{r\to\infty}\left[\left[\sup_{u\in K_{r_n}}\sup_{y\in O_{r_n}}|h_n(u,y)|\right]\right.\\
&\quad\left.\cdot\left[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}\left(\frac{\mathbb{E}\Big[\int_t^T\mathbb{1}_{O_{r_n}}(X^x_{t,s})\,\|Z^x_{t,s}\|\,\mathrm{d}s\Big]\sqrt{T-t}}{V(t,x)}\right)\right]\right]=0.
\end{aligned} \tag{1.60}$$

Combining this, (1.56), and (1.59) shows that for all $n \in \mathbb{N}$ it holds that

$$\limsup_{r\to\infty}\left[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}\left(\frac{\|v_n(t,x)\|}{V(t,x)}\sqrt{T-t}\right)\right]=0. \tag{1.61}$$

Item (iii) of Lemma 1.1.4 and the triangle inequality therefore prove that

$$\begin{aligned}
&\limsup_{r\to\infty}\left[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}\left(\tfrac{\|v(t,x)\|}{V(t,x)}\sqrt{T-t}\right)\right]\\
&\leqslant \inf_{n\in\mathbb{N}}\left(\limsup_{r\to\infty}\left[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}\left(\tfrac{\|v(t,x)-v_n(t,x)\|+\|v_n(t,x)\|}{V(t,x)}\sqrt{T-t}\right)\right]\right)\\
&= \inf_{n\in\mathbb{N}}\left(\limsup_{r\to\infty}\left[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}\left(\tfrac{\|v(t,x)-v_n(t,x)\|}{V(t,x)}\sqrt{T-t}\right)\right]\right)\\
&\leqslant \inf_{n\in\mathbb{N}}\left(\sup_{t\in[0,T)}\sup_{x\in O}\left(\tfrac{\|v(t,x)-v_n(t,x)\|}{V(t,x)}\sqrt{T-t}\right)\right)\\
&\leqslant \limsup_{n\to\infty}\left(\sup_{t\in[0,T)}\sup_{x\in O}\left(\tfrac{\|v(t,x)-v_n(t,x)\|}{V(t,x)}\sqrt{T-t}\right)\right)=0.
\end{aligned}\tag{1.62}$$

This establishes item (ii). The proof of Lemma 1.1.8 is thus complete. □

1.1.4 Contractivity of SFPEs

The following lemma derives a contractivity property for SFPEs under a Lipschitz assumption on f. This contractivity property will be used in Theorem 1.1.11 to apply Banach's fixed point theorem. Lemma 1.1.9 is a generalization of the results in [13, Lemma 2.8].

Lemma 1.1.9. *Assume Setting 1.1.1, let $L\in(0,\infty)$, let $f\colon[0,T)\times O\times\mathbb{R}^m\to\mathbb{R}$ be measurable, assume for all $t\in[0,T)$, $x\in O$, $y,z\in\mathbb{R}^m$ that $|f(t,x,y)-f(t,x,z)|\leqslant L|\!|\!|y-z|\!|\!|$, let $v,w\colon[0,T)\times O\to\mathbb{R}^m$ be measurable, and assume that*

$$\sup_{t\in[0,T)}\sup_{x\in O}\left[\frac{|\!|\!|v(t,x)|\!|\!|+|\!|\!|w(t,x)|\!|\!|}{V(t,x)}\sqrt{T-t}\right]<\infty. \tag{1.63}$$

Then it holds for all $\lambda\in(0,\infty)$, $t\in[0,T)$, $x\in O$ that

$$\begin{aligned}
&\mathbb{E}\left[\int_t^T|f(r,X^x_{t,r},v(r,X^x_{t,r}))-f(r,X^x_{t,r},w(r,X^x_{t,r}))||\!|\!|Z^x_{t,r}|\!|\!|\,\mathrm{d}r\right]\\
&\leqslant cL\sqrt{\tfrac{\pi^3}{4\lambda(T-t)}}V(t,x)e^{-\lambda t}\left[\sup_{s\in[0,T)}\sup_{y\in O}\left(\frac{e^{\lambda s}|\!|\!|v(s,y)-w(s,y)|\!|\!|}{V(s,y)}\sqrt{T-s}\right)\right].
\end{aligned}\tag{1.64}$$

Proof of Lemma 1.1.9. First note that the assumption that f, v, and w are measurable and the fact that for all $t\in[0,T]$, $x\in O$ it holds that X^x_t and Z^x_t are measurable ensure that for all $t\in[0,T)$, $x\in O$ it holds that $(t,T]\times\Omega\ni(s,\omega)\mapsto|f(s,X^x_{t,s}(\omega),v(s,X^x_{t,s}(\omega)))-f(s,X^x_{t,s}(\omega),w(s,X^x_{t,s}(\omega)))||\!|\!|Z^x_{t,s}|\!|\!|\in\mathbb{R}$ is measurable. Fubini's theorem and the assumption that for all $t\in[0,T)$, $x\in O$, $y,z\in\mathbb{R}^m$ it holds that $|f(t,x,y)-f(t,x,z)|\leqslant L|\!|\!|y-z|\!|\!|$ therefore imply that for all $\lambda\in(0,\infty)$, $t\in[0,T)$, $x\in O$ it holds that

$$\begin{aligned}
&\mathbb{E}\left[\int_t^T|f(r,X^x_{t,r},v(r,X^x_{t,r}))-f(r,X^x_{t,r},w(r,X^x_{t,r}))||\!|\!|Z^x_{t,r}|\!|\!|\,\mathrm{d}r\right]\\
&\leqslant \mathbb{E}\left[\int_t^T L|\!|\!|v(r,X^x_{t,r})-w(r,X^x_{t,r})|\!|\!|\,|\!|\!|Z^x_{t,r}|\!|\!|\,\mathrm{d}r\right]\\
&= L\int_t^T\mathbb{E}\left[\tfrac{e^{\lambda r}|\!|\!|v(r,X^x_{t,r})-w(r,X^x_{t,r})|\!|\!|}{V(r,X^x_{t,r})}\sqrt{T-r}\,\tfrac{V(r,X^x_{t,r})}{\sqrt{T-r}}|\!|\!|Z^x_{t,r}|\!|\!|\right]e^{-\lambda r}\,\mathrm{d}r\\
&\leqslant L\left[\sup_{s\in[0,T)}\sup_{y\in O}\left(\tfrac{e^{\lambda s}|\!|\!|v(s,y)-w(s,y)|\!|\!|}{V(s,y)}\sqrt{T-s}\right)\right]\int_t^T\tfrac{\mathbb{E}\left[V(r,X^x_{t,r})|\!|\!|Z^x_{t,r}|\!|\!|\right]}{\sqrt{T-r}}e^{-\lambda r}\,\mathrm{d}r.
\end{aligned}\tag{1.65}$$

This, item (ii) of Lemma 1.1.2 (applied for every $t \in [0,T)$ with $a \curvearrowleft t$, $b \curvearrowleft T$ in the notation of Lemma 1.1.2), and the fact that for all $t \in [0,T)$, $r \in (t,T]$, $x \in O$ it holds that $\mathbb{E}[V(r, X^x_{t,r}) |||Z^x_{t,r}|||] \leqslant \frac{c}{\sqrt{r-t}} V(t,x)$ demonstrate that for all $\lambda \in (0,\infty)$, $t \in [0,T)$, $x \in O$ it holds that

$$
\begin{aligned}
&\mathbb{E}\left[\int_t^T |f(r, X^x_{t,r}, v(r, X^x_{t,r})) - f(r, X^x_{t,r}, w(r, X^x_{t,r}))| |||Z^x_{t,r}||| \, \mathrm{d}r\right] \\
&\leqslant L \left[\sup_{s\in[0,T)} \sup_{y\in O} \left(\tfrac{e^{\lambda s}|||v(s,y)-w(s,y)|||}{V(s,y)} \sqrt{T-s}\right)\right] \int_t^T \tfrac{cV(t,x)}{\sqrt{(T-r)(r-t)}} e^{-\lambda r} \, \mathrm{d}r \\
&\leqslant cL\sqrt{\tfrac{\pi^3}{4\lambda(T-t)}} V(t,x) e^{-\lambda t} \left[\sup_{s\in[0,T)} \sup_{y\in O} \left(\tfrac{e^{\lambda s}|||v(s,y)-w(s,y)|||}{V(s,y)} \sqrt{T-s}\right)\right].
\end{aligned} \tag{1.66}
$$

The proof of Lemma 1.1.9 is thus complete. □

1.1.5 Existence and uniqueness of solutions of SFPEs

In this section we use the results of Sections 1.1.2-1.1.4 to obtain our main result of this section, Theorem 1.1.11. To make use of Banach's fixed point theorem in the proof of Theorem 1.1.11 we construct in the following lemma a vector space of SFPE solutions and show that it is a Banach space. In this sense, Lemma 1.1.10 generalizes a result in the proof of [13, Theorem 2.9] to the case where $[0,T)\backslash K_r$ is considered in (1.67) instead of $[0,T]$ and where (1.67) includes the $\sqrt{T-t}$ term.

Lemma 1.1.10. *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, let $\|\cdot\| \colon \mathbb{R}^d \to [0,\infty)$ be a norm, let $O \subseteq \mathbb{R}^d$ be a non-empty open set, for every $r \in (0,\infty)$ let $K_r \subseteq [0,T)$, $O_r \subseteq O$ satisfy $K_r = [0, \max\{T - \frac{1}{r}, 0\}]$, $O_r = \{x \in O \colon \|x\| \leqslant r$ and $\{y \in \mathbb{R}^d \colon \|y - x\| < \frac{1}{r}\} \subseteq O\}$, let $(B, |||\cdot|||)$ be an $\mathbb{R}$-Banach space, let $V \in C([0,T] \times O, (0,\infty))$ satisfy $\sup_{r\in(0,\infty)}[\inf_{t\in[0,T)\backslash K_r} \inf_{x\in O\backslash O_r} V(t,x)] = \infty$, let $\mathcal{V}$ satisfy*

$$
\begin{aligned}
\mathcal{V} = \Bigg\{ & v \in C([0,T) \times O, B) \colon \\
& \limsup_{r\to\infty} \left[\sup_{t\in[0,T)\backslash K_r} \sup_{x\in O\backslash O_r} \left(\frac{|||v(t,x)|||}{V(t,x)} \sqrt{T-t}\right)\right] = 0 \Bigg\},
\end{aligned} \tag{1.67}
$$

and let $\|\cdot\|_\lambda \colon \mathcal{V} \to [0,\infty)$, $\lambda \in \mathbb{R}$, satisfy for all $\lambda \in \mathbb{R}$, $w \in \mathcal{V}$ that

$$
\|w\|_\lambda = \sup_{t\in[0,T)} \sup_{x\in O} \left(\frac{e^{\lambda t}|||w(t,x)|||}{V(t,x)} \sqrt{T-t}\right). \tag{1.68}
$$

Then for all $\lambda \in \mathbb{R}$ it holds that $(\mathcal{V}, \|\cdot\|_\lambda)$ is an $\mathbb{R}$-Banach space.

Proof of Lemma 1.1.10. Throughout this proof let $\mathcal{W}_1, \mathcal{W}_2 \subseteq C([0,T) \times O, B)$ satisfy that

$$
\mathcal{W}_1 = \left\{v \in C([0,T) \times O, B) \colon \sup_{t\in[0,T)} \sup_{x\in O} \left(|||v(t,x)||| \sqrt{T-t}\right) < \infty\right\} \tag{1.69}
$$

and

$$
\begin{aligned}
\mathcal{W}_2 = \Big\{ & v \in C([0,T) \times O, B) \colon \\
& \limsup_{r\to\infty} \left[\sup_{t\in[0,T)\backslash K_r} \sup_{x\in O\backslash O_r} \left(|||v(t,x)||| \sqrt{T-t}\right)\right] = 0 \Big\},
\end{aligned} \tag{1.70}
$$

and let $\|\cdot\|_{\mathcal{W}_i}\colon \mathcal{W}_i \to [0,\infty)$, $i \in \{1,2\}$, satisfy for all $i \in \{1,2\}$, $w \in \mathcal{W}_i$ that

$$\|w\|_{\mathcal{W}_i} = \sup_{t\in[0,T)} \sup_{x\in O} \left(|||w(t,x)||| \sqrt{T-t}\right). \tag{1.71}$$

First we claim that $\|\cdot\|_{\mathcal{W}_2}\colon \mathcal{W}_2 \to [0,\infty)$ is well-defined. For this, let $w \in \mathcal{W}_2$. Observe that the assumption that $w \in C([0,T)\times O, B)$ and the fact that for all $r \in (0,\infty)$ it holds that K_r and O_r are compact ensures that for all $r \in (0,\infty)$ it holds that

$$\sup_{t\in K_r} \sup_{x\in O_r} \left(|||w(t,x)|||\sqrt{T-t}\right) < \infty. \tag{1.72}$$

Combining this with the fact that $\limsup_{r\to\infty}[\sup_{t\in[0,T)\setminus K_r} \sup_{x\in O\setminus O_r}(|||w(t,x)||| \sqrt{T-t})] = 0$ shows that

$$\sup_{t\in[0,T)} \sup_{x\in O} \left(|||w(t,x)|||\sqrt{T-t}\right) < \infty. \tag{1.73}$$

This demonstrates that $\|\cdot\|_{\mathcal{W}_2}$ is well-defined. In particular, this demonstrates that $\mathcal{W}_2 \subseteq \mathcal{W}_1$. Next note that $(\mathcal{W}_1, \|\cdot\|_{\mathcal{W}_1})$ is a normed $\mathbb{R}$-vector space. Let $(w_n)_{n\in\mathbb{N}} \subseteq \mathcal{W}_1$ be a Cauchy sequence. Observe that for all $\varepsilon \in (0,\infty)$ there exists $N_\varepsilon \in \mathbb{N}$ which satisfies for all $m,n \in \mathbb{N}\cap[N_\varepsilon,\infty)$ that $\|w_n - w_m\|_{\mathcal{W}_1} < \varepsilon$. This implies that for all $\varepsilon \in (0,\infty)$, $k,n \in \mathbb{N}\cap[N_\varepsilon,\infty)$, $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$|||w_n(t,x) - w_k(t,x)||| \sqrt{T-t} \leqslant \|w_n - w_k\|_{\mathcal{W}_1} < \varepsilon. \tag{1.74}$$

This ensures for all $t \in [0,T)$, $x \in O$ that $(w_n(t,x))_{n\in\mathbb{N}}$ is a Cauchy sequence in B. The fact that B is complete hence demonstrates that for all $t \in [0,T)$, $x \in O$ there exists $w(t,x) \in B$ which satisfies $w(t,x) = \lim_{n\to\infty} w_n(t,x)$. Moreover, note that (1.74) demonstrates that for all $\varepsilon \in (0,\infty)$, $n \in \mathbb{N}\cap[N_\varepsilon,\infty)$ it holds that

$$\begin{aligned}
&\sup_{t\in[0,T)} \sup_{x\in O} \left(|||w(t,x) - w_n(t,x)||| \sqrt{T-t}\right)\\
&= \sup_{t\in[0,T)} \sup_{x\in O} \left(\left|\left|\left| \lim_{m\to\infty} w_m(t,x) - w_n(t,x)\right|\right|\right| \sqrt{T-t}\right)\\
&\leqslant \limsup_{m\to\infty} \sup_{t\in[0,T)} \sup_{x\in O} \left(|||w_m(t,x) - w_n(t,x)||| \sqrt{T-t}\right) < \varepsilon.
\end{aligned} \tag{1.75}$$

This and the fact that $w_{N_1} \in \mathcal{W}_1$ demonstrate that

$$\begin{aligned}
&\sup_{t\in[0,T)} \sup_{x\in O} \left(|||w(t,x)|||\sqrt{T-t}\right)\\
&\leqslant \sup_{t\in[0,T)} \sup_{x\in O} \left(|||w(t,x) - w_{N_1}(t,x)|||\sqrt{T-t}\right)\\
&\quad + \sup_{t\in[0,T)} \sup_{x\in O} \left(|||w_{N_1}(t,x)|||\sqrt{T-t}\right)\\
&\leqslant 1 + \|w_{N_1}\|_{\mathcal{W}_1} < \infty.
\end{aligned} \tag{1.76}$$

Combining this with the fact that the uniform limit of continuous functions is continuous shows that $w \in \mathcal{W}_1$. This proves that $\mathcal{W}_1$ is a Banach space. Next we prove that $\mathcal{W}_2$ is a closed subset of $(\mathcal{W}_1, \|\cdot\|_{\mathcal{W}_1})$. For this, let $(w_n)_{n\in\mathbb{N}} \subseteq \mathcal{W}_2$ and $w \in \mathcal{W}_1$ satisfy that

$$\lim_{n\to\infty} \|w_n - w\|_{\mathcal{W}_1} = 0. \tag{1.77}$$

For every $\varepsilon \in (0,\infty)$ let $n_\varepsilon \in \mathbb{N}$ and $r_\varepsilon \in (0,\infty)$ satisfy $\|w - w_{n_\varepsilon}\|_{\mathcal{W}_1} < \frac{\varepsilon}{2}$ and $\sup_{t\in[0,T)\setminus K_{r_\varepsilon}} \sup_{x\in O\setminus O_{r_\varepsilon}} (|||w_{n_\varepsilon}(t,x)|||\sqrt{T-t}) < \frac{\varepsilon}{2}$. The triangle inequality hence ensures that

$$\begin{aligned}
&\limsup_{r\to\infty} \left[\sup_{t\in[0,T)\setminus K_r} \sup_{x\in O\setminus O_r} \left(|||w(t,x)|||\sqrt{T-t}\right)\right]\\
&\leqslant \limsup_{\varepsilon\to 0} \left[\sup_{t\in[0,T)\setminus K_{r_\varepsilon}} \sup_{x\in O\setminus O_{r_\varepsilon}} \left(|||w(t,x)|||\sqrt{T-t}\right)\right] \leqslant \limsup_{\varepsilon\to 0} \varepsilon = 0.
\end{aligned} \tag{1.78}$$

This implies that $w \in \mathcal{W}_2$. Combining this with the fact that $\mathcal{W}_2 \subseteq \mathcal{W}_1$ demonstrates that $\mathcal{W}_2$ is a complete subset of $(\mathcal{W}_1, \|\cdot\|_{\mathcal{W}_1})$. Next note that for all $\lambda \in \mathbb{R}$ it holds that $(\mathcal{V}, \|\cdot\|_\lambda)$ is an $\mathbb{R}$-vector space. We claim that $(\mathcal{V}, \|\cdot\|_0)$ is an $\mathbb{R}$-Banach space. For this, let $v_n \in \mathcal{V}$, $n \in \mathbb{N}$, satisfy

$$\limsup_{n\to\infty} \left[\sup_{k,l\in\mathbb{N}\cap[n,\infty)} \|v_k - v_l\|_0\right] = 0. \tag{1.79}$$

This implies that $\frac{v_n}{V} \colon [0,T) \times O \to \mathbb{R}$, $n \in \mathbb{N}$, is a Cauchy sequence in $(\mathcal{W}_2, \|\cdot\|_{\mathcal{W}_2})$. The fact that $(\mathcal{W}_2, \|\cdot\|_{\mathcal{W}_2})$ is a Banach space therefore implies that there exists a unique $\phi \in \mathcal{W}_2$ which satisfies

$$\limsup_{n\to\infty} \left\|\frac{v_n}{V} - \phi\right\|_{\mathcal{W}_2} = 0. \tag{1.80}$$

Observe that $\phi V = ([0,T) \times O \ni (t,x) \mapsto \phi(t,x)V(t,x) \in \mathbb{R}^m) \in \mathcal{V}$ and

$$\limsup_{n\to\infty} \|v_n - \phi V\|_0 = \limsup_{n\to\infty} \left\|\frac{v_n}{V} - \phi\right\|_{\mathcal{W}_2} = 0. \tag{1.81}$$

This proves that $(\mathcal{V}, \|\cdot\|_0)$ is an $\mathbb{R}$-Banach space. The fact that for all $\Lambda \in \mathbb{R}$, $\lambda \in [\Lambda, \infty)$, $v \in \mathcal{V}$ it holds that

$$\|v\|_\Lambda \leqslant \|v\|_\lambda \leqslant e^{(\lambda-\Lambda)T} \|v\|_\Lambda \tag{1.82}$$

hence shows that for all $\lambda \in \mathbb{R}$ it holds that $(\mathcal{V}, \|\cdot\|_\lambda)$ is an $\mathbb{R}$-Banach space. The proof of Lemma 1.1.10 is thus complete. □

Next we can combine the above findings and Banach's fixed point theorem to obtain this section's main result which establishes existence and uniqueness of certain SFPE solution under the assumption of a Lipschitz continuous nonlinearity. Theorem 1.1.11 is a generalization of [13, Theorem 2.9] to the case of gradient-dependent nonlinearities.

Theorem 1.1.11. *Assume Setting 1.1.1, let $L \in (0,\infty)$, assume for all $\varepsilon \in (0,\infty)$, $t \in [0,T)$, $s \in (t,T]$, $x \in O$ that*

$$\limsup_{[0,s)\times O\ni(u,y)\to(t,x)} \left[\mathbb{P}\big(\|X^y_{u,s} - X^x_{t,s}\| > \varepsilon\big) + \mathbb{E}\left[\|\!|Z^y_{u,s} - Z^x_{t,s}|\!\|\right]\right] = 0, \tag{1.83}$$

let $f \in C([0,T) \times O \times \mathbb{R}^m, \mathbb{R})$, $g \in C(O, \mathbb{R})$ satisfy for all $t \in [0,T)$, $x \in O$, $v, w \in \mathbb{R}^m$ that $|f(t,x,v) - f(t,x,w)| \leqslant L\|\!|v - w|\!\|$, and assume that $\inf_{r\in(0,\infty)}[\sup_{t\in[0,T)\setminus K_r} \sup_{x\in O\setminus O_r}(\frac{|g(x)|}{V(T,x)} + \frac{|f(t,x,0)|}{V(t,x)}\sqrt{T-t})] = 0$, $\inf_{t\in[0,T]} \inf_{x\in O} V(t,x) > 0$, $\sup_{r\in(0,\infty)}[\inf_{t\in[0,T)\setminus K_r} \inf_{x\in O\setminus O_r} V(t,x)] = \infty$, and for all $u \in (0,\infty)$ that

$$\limsup_{r\to\infty} \left[\sup_{t\in[0,T)\setminus K_r} \sup_{x\in O\setminus O_r} \left(\frac{\mathbb{E}[\int_t^T \mathbb{1}_{O_u}(X^x_{t,s})\,\|\!|Z^x_{t,s}|\!\|(\mathrm{d}s + \delta_T(\mathrm{d}s))]\sqrt{T-t}}{V(t,x)}\right)\right] = 0. \tag{1.84}$$

Then there exists a unique $v \in C([0,T) \times O, \mathbb{R}^m)$ such that

(i) it holds that

$$\limsup_{r\to\infty} \left[\sup_{t\in[0,T)\setminus K_r} \sup_{x\in O\setminus O_r} \left(\frac{\|\!|v(t,x)|\!\|}{V(t,x)} \sqrt{T-t}\right)\right] = 0, \tag{1.85}$$

(ii) it holds that $[0,T) \times O \ni (t,x) \mapsto \mathbb{E}\left[g(X^x_{t,T})Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, v(r, X^x_{t,r}))Z^x_{t,r}\,\mathrm{d}r\right] \in \mathbb{R}^m$ is well-defined and continuous, and

(iii) it holds for all $t \in [0,T)$, $x \in O$ that

$$v(t,x) = \mathbb{E}\left[g(X^x_{t,T})Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, v(r, X^x_{t,r}))Z^x_{t,r}\,\mathrm{d}r\right]. \tag{1.86}$$

Proof of Theorem 1.1.11. Let $\mathcal{V}$ satisfy

$$\mathcal{V} = \Big\{ v \in C([0,T) \times O, \mathbb{R}^m) \colon \limsup_{r\to\infty} \Big[\sup_{t\in[0,T)\backslash K_r} \sup_{x\in O\backslash O_r} \Big(\frac{|||v(t,x)|||}{V(t,x)} \sqrt{T-t} \Big) \Big] = 0 \Big\}, \tag{1.87}$$

and let $\|\cdot\|_\lambda \colon \mathcal{V} \to [0,\infty)$, $\lambda \in \mathbb{R}$, satisfy for all $\lambda \in \mathbb{R}$, $w \in \mathcal{V}$ that

$$\|w\|_\lambda = \sup_{t\in[0,T)} \sup_{x\in O} \left(\frac{e^{\lambda t} |||w(t,x)|||}{V(t,x)} \sqrt{T-t} \right). \tag{1.88}$$

Observe that Lemma 1.1.10 proves that for all $\lambda \in \mathbb{R}$ it holds that $(\mathcal{V}, \|\cdot\|_\lambda)$ is an $\mathbb{R}$-Banach space. Note that the triangle inequality, the assumption that $f \in C([0,T) \times O \times \mathbb{R}^m, \mathbb{R})$, and the assumption that for all $t \in [0,T)$, $x \in O$, $v, w \in \mathbb{R}^m$ it holds that $|f(t,x,v) - f(t,x,w)| \leqslant L|||v-w|||$ ensure that for all $w \in \mathcal{V}$ it holds that $[0,T) \times O \ni (t,x) \mapsto f(t,x,w(t,x)) \in \mathbb{R}$ is a continuous function which satisfies that for all $t \in [0,T)$, $x \in O$ it holds that

$$\begin{aligned} |f(t,x,w(t,x))| &\leqslant |f(t,x,0)| + |f(t,x,w(t,x)) - f(t,x,0)| \\ &\leqslant |f(t,x,0)| + L|||w(t,x)|||. \end{aligned} \tag{1.89}$$

The assumption that $\inf_{r\in(0,\infty)}[\sup_{t\in[0,T)\backslash K_r} \sup_{x\in O\backslash O_r} (\frac{|f(t,x,0)|}{V(t,x)} \sqrt{T-t})] = 0$ and (1.87) hence show that for all $w \in \mathcal{V}$ it holds that

$$\begin{aligned} &\inf_{r\in(0,\infty)} \left[\sup_{t\in[0,T)\backslash K_r} \sup_{x\in O\backslash O_r} \left(\frac{|f(t,x,w(t,x))|}{V(t,x)} \sqrt{T-t} \right) \right] \\ &\leqslant \inf_{r\in(0,\infty)} \left[\sup_{t\in[0,T)\backslash K_r} \sup_{x\in O\backslash O_r} \left(\frac{|f(t,x,0)|}{V(t,x)} \sqrt{T-t} + L \frac{|||w(t,x)|||}{V(t,x)} \sqrt{T-t} \right) \right] = 0. \end{aligned} \tag{1.90}$$

Combining this with Lemma 1.1.3 and item (i) of Lemma 1.1.4 demonstrates that for all $w \in \mathcal{V}$ it holds that

$$[0,T) \times O \ni (t,x) \mapsto \mathbb{E}\left[g(X^x_{t,T}) Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, w(r, X^x_{t,r})) Z^x_{t,r} \, dr \right] \in \mathbb{R}^m \tag{1.91}$$

is well-defined. Furthermore, note that items (i) and (ii) of Lemma 1.1.8 (applied for every $w \in \mathcal{V}$ with $h \curvearrowleft ([0,T) \times O \ni (t,x) \mapsto f(t,x,w(t,x)) \in \mathbb{R})$ in the notation of Lemma 1.1.8) and (1.90) prove that for every $w \in \mathcal{V}$ it holds that the function in (1.91) is in $\mathcal{V}$. This shows that there exists $\Phi \colon \mathcal{V} \to \mathcal{V}$ which satisfies for all $t \in [0,T)$, $x \in O$, $w \in \mathcal{V}$ that

$$(\Phi(w))(t,x) = \mathbb{E}\left[g(X^x_{t,T}) Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, w(r, X^x_{t,r})) Z^x_{t,r} \, dr \right]. \tag{1.92}$$

Furthermore, note that Lemma 1.1.9 demonstrates that for all $\lambda \in (0,\infty)$, $w, \tilde{w} \in \mathcal{V}$ it holds that

$$\begin{aligned} &\|\Phi(w) - \Phi(\tilde{w})\|_\lambda \\ &\leqslant \sup_{t\in[0,T)} \sup_{x\in O} \left(\frac{e^{\lambda t} \mathbb{E}\left[\int_t^T |f(r, X^x_{t,r}, v(r, X^x_{t,r})) - f(r, X^x_{t,r}, w(r, X^x_{t,r}))| \, |||Z^x_{t,r}||| \, dr \right]}{V(t,x)} \sqrt{T-t} \right) \\ &\leqslant cL \sqrt{\frac{\pi^3}{4\lambda}} \|w - \tilde{w}\|_\lambda. \end{aligned} \tag{1.93}$$

This implies that for all $\lambda \in [c^2L^2\pi^3, \infty)$, $w, \tilde{w} \in \mathcal{V}$ it holds that

$$\|\Phi(w) - \Phi(\tilde{w})\|_\lambda \leqslant \frac{1}{2}\|w - \tilde{w}\|_\lambda. \tag{1.94}$$

Banach's fixed point theorem therefore shows that there exists a unique $v \in \mathcal{V}$ which satisfies $\Phi(v) = v$. The proof of Theorem 1.1.11 is thus complete. □

1.2 SFPEs associated with stochastic differential equations (SDEs)

The aim of this section is to apply the abstract existence and uniqueness result in Theorem 1.1.11 to the case where X is an SDE solution and Z is the stochastic process in (1.99) below arising from the Bismut-Elworthy-Li formula. Subsections 1.2.1-1.2.2 provide several boundedness and convergence properties of SDE solutions which will be needed to prove our main result, Theorem 1.2.5. Throughout this section we frequently use the following setting.

Setting 1.2.1. *Let $d \in \mathbb{N}$, $\alpha, c, T \in (0, \infty)$, let $\langle\cdot,\cdot\rangle\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the standard Euclidean scalar product on $\mathbb{R}^d$, let $\|\cdot\|\colon \mathbb{R}^d \to [0, \infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, let $\|\cdot\|_F\colon \mathbb{R}^{d\times d} \to [0, \infty)$ be the Frobenius norm on $\mathbb{R}^{d\times d}$, let $O \subseteq \mathbb{R}^d$ be an open set, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_s)_{s\in[0,T]})$ be a filtered probability space satisfying the usual conditions, let $W\colon [0,T]\times\Omega \to \mathbb{R}^d$ be a standard $(\mathbb{F}_s)_{s\in[0,T]}$-Brownian motion, let $\mu \in C^{0,1}([0,T]\times O, \mathbb{R}^d)$, $\sigma \in C^{0,1}([0,T] \times O, \mathbb{R}^{d\times d})$ satisfy for all $s \in [0,T]$, $x, y \in O$, $v \in \mathbb{R}^d$ that*

$$\max\left\{\langle x - y, \mu(s,x) - \mu(s,y)\rangle, \tfrac{1}{2}\|\sigma(s,x) - \sigma(s,y)\|_F^2\right\} \leqslant \tfrac{c}{2}\|x - y\|^2 \tag{1.95}$$

and $v^*\sigma(s,x)(\sigma(s,x))^*v \geqslant \alpha\|v\|^2$, for every $t \in [0,T]$, $x \in O$ let $X_t^x = ((X_{t,s}^{x,1}, \ldots, X_{t,s}^{x,d}))_{s\in[t,T]}\colon [t,T] \times \Omega \to O$ be an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in [t,T]$ it holds a.s. that

$$X_{t,s}^x = x + \int_t^s \mu(r, X_{t,r}^x)\,\mathrm{d}r + \int_t^s \sigma(r, X_{t,r}^x)\,\mathrm{d}W_r, \tag{1.96}$$

and assume for all $t \in [0,T]$, $\omega \in \Omega$ that $\big([t,T] \times O \ni (s,x) \mapsto X_{t,s}^x(\omega) \in O\big) \in C^{0,1}([t,T]\times O, O)$.

1.2.1 Moment estimates

The next lemma establishes several moment estimates for SDE solutions and the stochastic process in (1.99) below arising from the Bismut-Elworthy-Li formula.

Lemma 1.2.2. *Let $d \in \mathbb{N}$, $\alpha, c, T \in (0, \infty)$, let $t \in [0,T]$, let $\langle\cdot,\cdot\rangle\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the standard Euclidean scalar product on $\mathbb{R}^d$, let $\|\cdot\|\colon \mathbb{R}^d \to [0, \infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, let $\|\cdot\|_F\colon \mathbb{R}^{d\times d} \to [0, \infty)$ be the Frobenius norm on $\mathbb{R}^{d\times d}$, let $O \subseteq \mathbb{R}^d$ be an open set, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_s)_{s\in[0,T]})$ be a filtered probability space satisfying the usual conditions, let $W\colon [0,T] \times \Omega \to \mathbb{R}^d$ be a standard $(\mathbb{F}_s)_{s\in[0,T]}$-Brownian motion, let $\mu \in C^{0,1}([0,T] \times O, \mathbb{R}^d)$, $\sigma \in C^{0,1}([0,T] \times O, \mathbb{R}^{d\times d})$ satisfy for all $s \in [t,T]$, $x, y \in O$, $v \in \mathbb{R}^d$ that*

$$\max\left\{\langle x - y, \mu(s,x) - \mu(s,y)\rangle, \tfrac{1}{2}\|\sigma(s,x) - \sigma(s,y)\|_F^2\right\} \leqslant \tfrac{c}{2}\|x - y\|^2 \tag{1.97}$$

and $v^*\sigma(s,x)(\sigma(s,x))^*v \geqslant \alpha\|v\|^2$, *let* $\mathfrak{m} \in [0,\infty)$ *satisfy* $\mathfrak{m} = \max_{s\in[0,T]}[\frac{1}{2}\|\mu(s,0)\|^2 + \|\sigma(s,0)\|_F^2]$, *for every* $x \in O$ *let* $X^x = ((X^x_{s,1},\dots,X^x_{s,d}))_{s\in[t,T]} \colon [t,T]\times\Omega \to O$ *be an* $(\mathbb{F}_s)_{s\in[t,T]}$*-adapted stochastic process with continuous sample paths satisfying that for all* $s \in [t,T]$ *it holds a.s. that*

$$X^x_s = x + \int_t^s \mu(r, X^x_r)\,\mathrm{d}r + \int_t^s \sigma(r, X^x_r)\,\mathrm{d}W_r, \tag{1.98}$$

assume for all $\omega \in \Omega$ *that* $([t,T]\times O \ni (s,x) \mapsto X^x_s(\omega) \in O) \in C^{0,1}([t,T]\times O, O)$, *and for every* $x \in O$ *let* $Z^x = (Z^x_s)_{s\in(t,T]} \colon (t,T]\times\Omega \to \mathbb{R}^d$ *be an* $(\mathbb{F}_s)_{s\in(t,T]}$*-adapted stochastic process with continuous sample paths satisfying that for all* $s \in (t,T]$ *it holds a.s. that*

$$Z^x_s = \frac{1}{s-t}\int_t^s (\sigma(r, X^x_r))^{-1}\left(\frac{\partial}{\partial x}X^x_r\right)\mathrm{d}W_r. \tag{1.99}$$

Then

(i) *for all* $x \in O$ *and all stopping times* $\tau \colon \Omega \to [t,T]$ *it holds that*

$$\Big(\mathbb{E}\big[\|X^x_\tau\|^2\big]\Big)^{\frac{1}{2}} \leqslant \exp((2c+1)T)\Big(\|x\|^2 + \frac{\mathfrak{m}}{2c+1}\Big)^{\frac{1}{2}}, \tag{1.100}$$

(ii) *for all* $j \in \{1,\dots,d\}$, $x \in O$, *and all stopping times* $\tau \colon \Omega \to [t,T]$ *it holds that*

$$\left(\mathbb{E}\left[\Big\|\frac{\partial}{\partial x_j}X^x_\tau\Big\|^2\right]\right)^{\frac{1}{2}} \leqslant \mathbb{E}\big[\exp\big(c(\tau-t)\big)\big], \tag{1.101}$$

(iii) *for all* $x \in O$ *and all stopping times* $\tau \colon \Omega \to [t,T]$ *it holds that*

$$\mathbb{E}\left[\Big\|\frac{\partial}{\partial x}X^x_\tau\Big\|_F^2\right] \leqslant d\exp\big(2c(T-t)\big), \tag{1.102}$$

(iv) *for all* $s \in [t,T]$, $x \in O$ *it holds that*

$$\|(\sigma(s, X^x_s))^{-1}\|^2_{L(\mathbb{R}^d)} \leqslant \frac{1}{\alpha}, \tag{1.103}$$

(v) *for all* $x \in O$ *and all stopping times* $\tau \colon \Omega \to [t,T]$ *it holds that*

$$\mathbb{E}\left[\Big\|\int_t^\tau (\sigma(r, X^x_r))^{-1}\left(\frac{\partial}{\partial x}X^x_r\right)\mathrm{d}W_r\Big\|^2\right] \leqslant \frac{dT}{\alpha}\exp(2cT), \tag{1.104}$$

and

(vi) *for all* $s \in (t,T]$, $x \in O$ *it holds that*

$$\mathbb{E}\Big[\|Z^x_s\|^2\Big] \leqslant \frac{d}{\alpha(s-t)^2}\int_t^s \exp(2(r-t)c)\,\mathrm{d}r. \tag{1.105}$$

Proof of Lemma 1.2.2. Throughout this proof let $\mathbf{e}_1,\mathbf{e}_2,\dots,\mathbf{e}_d \in \mathbb{R}^d$ satisfy that $\mathbf{e}_1 = (1,0,\dots,0)$, $\mathbf{e}_2 = (0,1,0,\dots,0)$, …, $\mathbf{e}_d = (0,\dots,0,1)$. First observe that (1.97), the

Cauchy-Schwarz inequality, and the fact that for all $a, b \in [0, \infty)$ it holds that $2ab \leqslant a^2 + b^2$ imply that for all $x \in O$, $s \in [t, T]$ it holds that

$$\begin{aligned}
&\langle X_s^x, \mu(s, X_s^x)\rangle + \frac{1}{2}\|\sigma(s, X_s^x)\|_F^2 \\
&\leqslant \langle X_s^x, \mu(s, X_s^x) - \mu(s,0)\rangle + \langle X_s^x, \mu(s,0)\rangle \\
&\qquad + \|\sigma(s, X_s^x) - \sigma(s,0)\|_F^2 + \|\sigma(s,0)\|_F^2 \\
&\leqslant \frac{c}{2}\|X_s^x\|^2 + \|X_s^x\|\|\mu(s,0)\| + c\|X_s^x\|^2 + \|\sigma(s,0)\|_F^2 \\
&\leqslant \frac{c}{2}\|X_s^x\|^2 + \frac{1}{2}\|X_s^x\|^2 + \frac{1}{2}\|\mu(s,0)\|^2 + c\|X_s^x\|^2 + \|\sigma(s,0)\|_F^2 \\
&\leqslant (2c+1)\|X_s^x\|^2 + \mathfrak{m}.
\end{aligned} \tag{1.106}$$

Combining this with [50, Corollary 2.5 (i)] (applied for every $x \in \mathbb{R}^d$, $j \in \{1, 2, \ldots, d\}$ with $H \curvearrowleft \mathbb{R}^d$, $U \curvearrowleft \mathbb{R}^d$, $T \curvearrowleft T - t$, $a \curvearrowleft ([0, T-t] \times \Omega \ni (s, \omega) \mapsto \mu(t+s, x) \in \mathbb{R}^d)$, $b \curvearrowleft ([0, T-t] \times \Omega \ni (s, \omega) \mapsto \sigma(t+s, x) \in \mathbb{R}^{d \times d})$, $X \curvearrowleft ([0, T-t] \times \Omega \ni (s, \omega) \mapsto X_{t+s}^x(\omega) \in O)$, $p \curvearrowleft 2$, $\alpha \curvearrowleft ([0, T-t] \times \Omega \ni (s, \omega) \mapsto (2c+1) \in \mathbb{R})$, $\beta \curvearrowleft ([0, T-t] \times \Omega \ni (s, \omega) \mapsto \sqrt{2\mathfrak{m}} \in \mathbb{R})$, $q_1 \curvearrowleft 2$, $q_2 \curvearrowleft \infty$ in the notation of [50, Corollary 2.5 (i)]) shows that for all $x \in O$ and all stopping times $\tau \colon \Omega \to [t, T]$ it holds that

$$\begin{aligned}
\Big(\mathbb{E}\big[\|X_\tau^x\|^2\big]\Big)^{\frac{1}{2}} &\leqslant \exp((2c+1)T)\Big(\|x\|^2 + 2\mathfrak{m}\int_0^T \exp(-2(2c+1)s)\, ds\Big)^{\frac{1}{2}} \\
&\leqslant \exp((2c+1)T)\Big(\|x\|^2 + \frac{2\mathfrak{m}}{4c+2}(1 - \exp(-(4c+2)T))\Big)^{\frac{1}{2}} \\
&\leqslant \exp((2c+1)T)\Big(\|x\|^2 + \frac{\mathfrak{m}}{2c+1}\Big)^{\frac{1}{2}}.
\end{aligned} \tag{1.107}$$

This establishes item (i). Next note that (1.98), the Leibniz integral rule, the chain rule, and the fact that $\mu \in C^{0,1}([0,T] \times O, \mathbb{R}^d)$ and $\sigma \in C^{0,1}([0,T] \times O, \mathbb{R}^{d \times d})$ show that for all $s \in [t, T]$, $x \in O$, $j \in \{1, 2, \ldots, d\}$ it holds a.s. that

$$\begin{aligned}
\frac{\partial}{\partial x_j} X_s^x = \mathbf{e}_j + \sum_{i=1}^d \bigg[&\int_t^s \Big(\frac{\partial \mu}{\partial x_i}\Big)(r, X_r^x)\Big(\frac{\partial}{\partial x_j} X_{r,i}^x\Big)\, dr \\
&+ \int_t^s \Big(\frac{\partial \sigma}{\partial x_i}\Big)(r, X_r^x)\Big(\frac{\partial}{\partial x_j} X_{r,i}^x\Big)\, dW_r\bigg].
\end{aligned} \tag{1.108}$$

Moreover, observe that the chain rule, (1.97), and the assumption that for all $\omega \in \Omega$ it holds that $([t,T] \times O \ni (s,x) \mapsto X_s^x(\omega) \in O) \in C^{0,1}([t,T] \times O, O)$ ensure that for all $r \in [t, T]$, $x \in O$, $j \in \{1, 2, \ldots, d\}$ it holds a.s. that

$$\begin{aligned}
&\Big\langle \Big(\frac{\partial}{\partial x_j} X_r^x\Big), \sum_{i=1}^d \Big(\frac{\partial \mu}{\partial x_i}\Big)(r, X_r^x)\frac{\partial}{\partial x_j} X_{r,i}^x \Big\rangle + \frac{1}{2}\Big\| \sum_{k=1}^d \Big(\frac{\partial \sigma}{\partial x_k}\Big)(r, X_r^x)\frac{\partial}{\partial x_j} X_{r,k}^x \Big\|_F^2 \\
&= \lim_{(0,\infty) \ni h \to 0} \frac{1}{h^2}\bigg[\Big\langle X_r^{x+h\mathbf{e}_j} - X_r^x, \mu(r, X_r^{x+h\mathbf{e}_j}) - \mu(r, X_r^x)\Big\rangle \\
&\qquad + \frac{1}{2}\Big\|\sigma(r, X_r^{x+h\mathbf{e}_j}) - \sigma(r, X_r^x)\Big\|_F^2\bigg] \\
&\leqslant \Big(\frac{c}{2} + \frac{c}{2}\Big) \lim_{(0,\infty) \ni h \to 0} \frac{1}{h^2}\Big[\big\|X_r^{x+h\mathbf{e}_j} - X_r^x\big\|^2\Big] = c\Big\|\frac{\partial}{\partial x_j} X_r^x\Big\|^2.
\end{aligned} \tag{1.109}$$

This, (1.108), and e.g., [50, Corollary 2.5(i)] (applied for every $x \in \mathbb{R}^d$, $j \in \{1, 2, \ldots, d\}$ with $H \curvearrowleft \mathbb{R}^d$, $U \curvearrowleft \mathbb{R}^d$, $T \curvearrowleft T - t$, $a \curvearrowleft ([0, T-t] \times \Omega \ni (s, \omega) \mapsto \sum_{i=1}^d \frac{\partial \mu}{\partial x_i}(t +$

$s, x)\frac{\partial}{\partial x_j} X^x_{t+s,i} \in \mathbb{R}^d)$, $b \hookleftarrow ([0, T-t] \times \Omega \ni (s,\omega) \mapsto \sum_{i=1}^d \frac{\partial \sigma}{\partial x_i}(t+s,x)\frac{\partial}{\partial x_j} X^x_{t+s,i} \in \mathbb{R}^{d\times d})$, $X \hookleftarrow (\frac{\partial}{\partial x_j} X^x_{t+r})_{r\in[0,T-t]}$, $p \hookleftarrow 2$, $q_1 \hookleftarrow 2$, $q_2 \hookleftarrow \infty$, $\alpha \hookleftarrow ([0,T-t]\times\Omega \ni (s,\omega) \mapsto c \in [0,\infty])$, and $\beta \hookleftarrow 0$ in the notation of [50, Corollary 2.5(i)]) demonstrate that for all $j \in \{1, 2, \ldots, d\}$, $x \in O$, and every stoppping time $\tau\colon \Omega \to [t, T]$ it holds that

$$\Big(\mathbb{E}\Big[\big\|\frac{\partial}{\partial x_j} X^x_\tau\big\|^2\Big]\Big)^{\frac{1}{2}} \leqslant \exp\big(c(T-t)\big). \tag{1.110}$$

This establishes item (ii). In the next step note that item (ii) implies that for all $x \in O$ and all stopping times $\tau\colon \Omega \to [t,T]$ it holds that

$$\mathbb{E}\Big[\big\|\frac{\partial}{\partial x} X^x_\tau\big\|_F^2\Big] = \sum_{j=1}^d \mathbb{E}\Big[\big\|\frac{\partial}{\partial x_j} X^x_\tau\big\|^2\Big] \leqslant d\exp\big(2c(T-t)\big). \tag{1.111}$$

This establishes item (iii). Next observe that the assumption that for all $s \in [t,T]$, $x \in O$, $v \in \mathbb{R}^d$ it holds that $v^*\sigma(s,x)(\sigma(s,x))^*v \geqslant \alpha\|v\|^2$ ensures that for all $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that $\sigma(s,x)(\sigma(s,x))^*$ is invertible. Combining this with the fact that for all $s \in [t,T]$, $x \in O$ it holds that $\operatorname{rank}(\sigma(s,x)(\sigma(s,x))^*) = \operatorname{rank}(\sigma(s,x))$ shows that for all $s \in [t,T]$, $x \in O$ it holds that $\sigma(s,x)$ is invertible. The assumption that for all $s \in [t,T]$, $x \in O$, $y \in \mathbb{R}^d$ it holds that $y^*\sigma(s,x)(\sigma(s,x))^*y \geqslant \alpha\|y\|^2$ therefore demonstrates that for all $s \in [t,T]$, $x \in O$, $y \in \mathbb{R}^d$ it holds that

$$\begin{aligned}\|y\|^2 &= y^*\left(\sigma(s,X^x_s)\right)^{-1}\sigma(s,X^x_s)(\sigma(s,X^x_s))^*\left((\sigma(s,X^x_s))^{-1}\right)^* y\\ &\geqslant \alpha\big\|\left((\sigma(s,X^x_s))^{-1}\right)^* y\big\|^2.\end{aligned} \tag{1.112}$$

This implies that for all $s \in [t,T]$, $x \in O$ it holds that

$$\big\|(\sigma(s,X^x_s))^{-1}\big\|^2_{L(\mathbb{R}^d)} = \sup_{y\in\mathbb{R}^d\setminus\{0\}} \frac{\|((\sigma(s,X^x_s))^{-1})^* y\|^2}{\|y\|^2} \leqslant \frac{1}{\alpha}. \tag{1.113}$$

This establishes item (iv). In the next step note that the Burkholder-Davis-Gundy inequality, Fubini's theorem, and items (iii) and (iv) prove that for all $x \in O$ and all stopping times $\tau\colon \Omega \to [t,T]$ it holds that

$$\begin{aligned}&\mathbb{E}\left[\Big\|\int_t^\tau (\sigma(r,X^x_r))^{-1}\Big(\frac{\partial}{\partial x}X^x_r\Big)\,\mathrm{d}W_r\Big\|^2\right]\\ &\leqslant \int_t^T \mathbb{E}\left[\Big\|(\sigma(r,X^x_r))^{-1}\Big(\frac{\partial}{\partial x}X^x_r\Big)\Big\|_F^2\right]\mathrm{d}r\\ &\leqslant \int_t^T \mathbb{E}\left[\big\|(\sigma(r,X^x_r))^{-1}\big\|^2_{L(\mathbb{R}^d)}\Big\|\frac{\partial}{\partial x}X^x_r\Big\|_F^2\right]\mathrm{d}r \leqslant \frac{1}{\alpha}\int_t^T \mathbb{E}\left[\Big\|\frac{\partial}{\partial x}X^x_r\Big\|_F^2\right]\mathrm{d}r\\ &\leqslant \frac{d}{\alpha}\int_t^T \exp(2cT)\,\mathrm{d}r \leqslant \frac{dT}{\alpha}\exp(2cT).\end{aligned} \tag{1.114}$$

This establishes item (v). Next note that (1.108), (1.109), and e.g., [50, Corollary 2.5(i)] (applied for every $x \in \mathbb{R}^d$, $j \in \{1,2,\ldots,d\}$ with $H \hookleftarrow \mathbb{R}^d$, $U \hookleftarrow \mathbb{R}^d$, $T \hookleftarrow T-t$, $a \hookleftarrow ([0,T-t]\times\Omega \ni (s,\omega) \mapsto \sum_{i=1}^d \frac{\partial\mu}{\partial x_i}(t+s,x)\frac{\partial}{\partial x_j}X^x_{t+s,i} \in \mathbb{R}^d)$, $b \hookleftarrow ([0,T-t]\times\Omega \ni (s,\omega) \mapsto \sum_{i=1}^d \frac{\partial\sigma}{\partial x_i}(t+s,x)\frac{\partial}{\partial x_j}X^x_{t+s,i} \in \mathbb{R}^{d\times d})$, $X \hookleftarrow (\frac{\partial}{\partial x_j}X^x_{t+r})_{r\in[0,T-t]}$, $p \hookleftarrow 2$, $q_1 \hookleftarrow 2$, $q_2 \hookleftarrow \infty$, $\alpha \hookleftarrow ([0,T-t]\times\Omega \ni (s,\omega) \mapsto c \in [0,\infty])$, and $\beta \hookleftarrow 0$ in the notation of [50, Corollary 2.5(i)]) imply that for all $j \in \{1,2,\ldots,d\}$, $x \in O$, $r \in [t,T]$ it holds that

$$\Big(\mathbb{E}\Big[\big\|\frac{\partial}{\partial x_j} X^x_r\big\|^2\Big]\Big)^{\frac{1}{2}} \leqslant \exp\big(c(r-t)\big). \tag{1.115}$$

Combining this with the Burkholder-Davis-Gundy inequality, Fubini's theorem, and item (iv) demonstrates that for all $s \in (t, T]$, $x \in O$ it holds that

$$
\begin{aligned}
(s-t)^2 \mathbb{E}\left[\|Z_s^x\|^2\right] &= \mathbb{E}\left[\left\| \int_t^s (\sigma(r, X_r^x))^{-1} \Big(\tfrac{\partial}{\partial x} X_r^x\Big)\, \mathrm{d}W_r \right\|^2\right] \\
&\leqslant \int_t^s \mathbb{E}\left[\left\|(\sigma(r, X_r^x))^{-1} \Big(\tfrac{\partial}{\partial x} X_r^x\Big)\right\|_F^2\right] \mathrm{d}r \\
&\leqslant \int_t^s \mathbb{E}\left[\|(\sigma(r, X_r^x))^{-1}\|_{L(\mathbb{R}^d)}^2 \left\|\tfrac{\partial}{\partial x} X_r^x\right\|_F^2\right] \mathrm{d}r \\
&\leqslant \frac{1}{\alpha} \int_t^s \mathbb{E}\left[\left\|\tfrac{\partial}{\partial x} X_r^x\right\|_F^2\right] \mathrm{d}r \leqslant \frac{d}{\alpha} \int_t^s \exp(2(r-t)c)\, \mathrm{d}r.
\end{aligned}
\tag{1.116}
$$

This establishes item (vi). The proof of Lemma 1.2.2 is thus complete. □

1.2.2 Continuity in the starting point and starting time

The following lemma demonstrates convergence in probability of SDE solutions in the starting point. Lemma 1.2.3 is a generalization of [13, Lemma 3.7].

Lemma 1.2.3. *Let $d, m \in \mathbb{N}$, $T \in (0, \infty)$, let $\|\cdot\|\colon \mathbb{R}^d \to [0, \infty)$ and $|||\cdot|||\colon \mathbb{R}^{d\times m} \to [0, \infty)$ be norms, let $O \subseteq \mathbb{R}^d$ be a non-empty open set, for every $r \in (0, \infty)$ let $O_r \subseteq O$ satisfy $O_r = \{x \in O\colon \|x\| \leqslant r \text{ and } \{y \in \mathbb{R}^d\colon \|x-y\| < \frac{1}{r}\} \subseteq O\}$, let $\mu \in C([0, T] \times O, \mathbb{R}^d)$, $\sigma \in C([0, T] \times O, \mathbb{R}^{d\times m})$ satisfy for all $r \in (0, \infty)$ that*

$$
\sup\left(\left\{\tfrac{\|\mu(t,x)-\mu(t,y)\| + |||\sigma(t,x)-\sigma(t,y)|||}{\|x-y\|} : t \in [0, T], x, y \in O_r, x \neq y\right\} \cup \{0\}\right) < \infty, \tag{1.117}
$$

let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_s)_{s\in[0,T]})$ be a filtered probability space satisfying the usual conditions, let $W\colon [0, T] \times \Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_s)_{s\in[0,T]}$-Brownian motion, for every $t \in [0, T]$, $x \in O$ let $X_t^x = (X_{t,s}^x)_{s\in[t,T]}\colon [t, T] \times \Omega \to O$ be an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in [t, T]$ it holds a.s. that

$$
X_{t,s}^x = x + \int_t^s \mu(r, X_{t,r}^x)\, \mathrm{d}r + \int_t^s \sigma(r, X_{t,r}^x)\, \mathrm{d}W_r, \tag{1.118}
$$

and assume for all non-empty, compact $\mathcal{K} \subseteq [0, T] \times O$ that

$$
\inf_{k\in\mathbb{N}}\left[\sup_{(t,x)\in\mathcal{K}}\left(\sup_{\tau\colon \Omega\to[t,T] \text{ stopping time}} \mathbb{P}\big(\|X_{t,\tau}^x\| \geqslant k\big)\right)\right] = 0. \tag{1.119}
$$

Then it holds for all $\varepsilon \in (0, \infty)$, $s \in [0, T]$, and all $(t_n, x_n) \in [0, T] \times O$, $n \in \mathbb{N}_0$, with $\limsup_{n\to\infty}[|t_n - t_0| + \|x_n - x_0\|] = 0$ that

$$
\limsup_{n\to\infty}\left[\mathbb{P}\left(\left\|X_{t_n,\max\{s,t_n\}}^{x_n} - X_{t_0,\max\{s,t_0\}}^{x_0}\right\| \geqslant \varepsilon\right)\right] = 0. \tag{1.120}
$$

Proof of Lemma 1.2.3. Throughout this proof let $(t_n, x_n) \in [0, T] \times O$, $n \in \mathbb{N}_0$, satisfy $\limsup_{n\to\infty}[|t_n - t_0| + \|x_n - x_0\|] = 0$ and let $U_n \subseteq O$, $n \in \mathbb{N}$, satisfy for all $n \in \mathbb{N}$ that $U_n = \{x \in O\colon (\exists y \in O_n\colon \|y - x\| < \frac{1}{2n})\}$. Note that for every $n \in \mathbb{N}$ it holds that $O_n \subseteq O$ is a compact set, $U_n \subseteq O$ is an open set, and $O_n \subseteq U_n$. Combining this with [73, Theorem II.3.7] (applied for every $n \in \mathbb{N}$ with $E \curvearrowleft [0, T] \times O$, $A_1 \curvearrowleft [0, T] \times O \backslash U_n$, $A_2 \curvearrowleft [0, T] \times O_n$ in the notation of [73, Theorem II.3.7]) demonstrates that for all $n \in \mathbb{N}$ there exists $\varphi_n \in C_c^\infty([0, T] \times O, \mathbb{R})$ which satisfies for all $t \in [0, T]$, $x \in O$ that

$$
\mathbb{1}_{[0,T]\times O_n}(t, x) \leqslant \varphi_n(t, x) \leqslant \mathbb{1}_{[0,T]\times U_n}(t, x). \tag{1.121}
$$

Let $m_n\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $n \in \mathbb{N}$, and $s_n\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$, $n \in \mathbb{N}$, satisfy for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$m_n(t,x) = \begin{cases} \varphi_n(t,x)\mu(t,x) & : x \in O \\ 0 & : x \in \mathbb{R}^d \backslash O \end{cases} \tag{1.122}$$

and

$$s_n(t,x) = \begin{cases} \varphi_n(t,x)\sigma(t,x) & : x \in O \\ 0 & : x \in \mathbb{R}^d \backslash O. \end{cases} \tag{1.123}$$

Combining this with (1.117) and (1.121) shows that $m_n\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $n \in \mathbb{N}$, and $s_n\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$, $n \in \mathbb{N}$, are compactly supported, continuous functions which satisfy that

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

$$\sup_{t\in[0,T]} \sup_{\substack{x,y\in\mathbb{R}^d \\ x\neq y}} \left[\frac{\|m_n(t,x) - m_n(t,y)\| + \|\!|s_n(t,x) - s_n(t,y)\|\!|}{\|x-y\|} \right] < \infty, \tag{1.124}$$

(II) for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O_n$ it holds that

$$\left[\|m_n(t,x) - \mu(t,x)\| + \|\!|s_n(t,x) - \sigma(t,x)\|\!|\right] = 0, \tag{1.125}$$

and

(III) for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O \backslash U_n$ it holds that

$$\left[\|m_n(t,x)\| + \|\!|s_n(t,x)\|\!|\right] = 0. \tag{1.126}$$

Observe that [69, Theorem 5.2.9] (applied for all $n \in \mathbb{N}$ with $b \curvearrowleft m_n$, $\sigma \curvearrowleft s_n$ in the notation of [69, Theorem 5.2.9]) and item (I) demonstrate that for every $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O$ there exists an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process $X^{x,n}_t = (X^{x,n}_{t,s})_{s\in[t,T]}\colon [t,T] \times \Omega \to \mathbb{R}^d$ with continuous sample paths which satisfies that for all $s \in [t,T]$ it holds a.s. that

$$X^{x,n}_{t,s} = x + \int_t^s m_n(r, X^{x,n}_{t,r})\, \mathrm{d}r + \int_t^s s_n(r, X^{x,n}_{t,r})\, \mathrm{d}W_r. \tag{1.127}$$

Moreover, note that item (III) ensures that for all $n \in \mathbb{N}$ it holds that $\operatorname{supp}(m_n) \cup \operatorname{supp}(s_n) \subseteq [0,T] \times U_n$. This and [13, Lemma 3.4] (applied for every $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O$ with $T \curvearrowleft T-t$, $\mathcal{O} \curvearrowleft U_n$, $\mu \curvearrowleft ([0,T-t] \times O \ni (s,y) \mapsto m_n(t+s,y) \in \mathbb{R}^d)$, $\sigma \curvearrowleft ([0,T-t] \times O \ni (s,y) \mapsto s_n(t+s,y) \in \mathbb{R}^{d\times m})$, $\mathbb{F} \curvearrowleft (\mathbb{F}_{t+s})_{s\in[0,T-t]}$, $W \curvearrowleft ([0,T-t] \times \Omega \ni (s,\omega) \mapsto W_{t+s}(\omega) - W_t(\omega) \in \mathbb{R}^m)$, $X \curvearrowleft ([0,T-t] \times \Omega \ni (s,\omega) \mapsto X^{x,n}_{t,t+s} \in O)$ in the notation of [13, Lemma 3.4]) show that for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in U_n$ it holds that $\mathbb{P}(\forall s \in [t,T]\colon X^{x,n}_{t,s} \in \overline{U_n}) = 1$ and for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O \backslash U_n$ it holds that $\mathbb{P}(\forall s \in [t,T]\colon X^{x,n}_{t,s} = x) = 1$. This ensures that for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O$ there exists an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process $\mathcal{X}^{x,n}_t = (\mathcal{X}^{x,n}_{t,s})_{s\in[t,T]}\colon [t,T] \times \Omega \to O$ with continuous sample paths which satisfies that for all $s \in [t,T]$ it holds a.s. that

$$\mathcal{X}^{x,n}_{t,s} = x + \int_t^s m_n(r, \mathcal{X}^{x,n}_{t,r})\, \mathrm{d}r + \int_t^s s_n(r, \mathcal{X}^{x,n}_{t,r})\, \mathrm{d}W_r. \tag{1.128}$$

In the next step let $\tau^{n,t,x}\colon \Omega \to [t,T]$, $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O$, satisfy for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O$, $\omega \in \Omega$ that $\tau^{n,t,x}(\omega) = \inf(\{s \in [t,T]\colon \max\{\|\mathcal{X}^{x,n}_{t,s}\|, \|X^x_{t,s}\|\} > n\} \cup \{T\})$.

Note that for every $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O$ it holds that $\tau^{n,t,x} \colon \Omega \to [t,T]$ is an $(\mathbb{F}_s)_{s\in[t,T]}$-stopping time. Next observe that [13, Lemma 3.5] (applied for every $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O$ with $T \curvearrowleft T-t$, $\mathcal{C} \curvearrowleft [0,T-t]\times O_n$, $\mu_1 \curvearrowleft ([0,T-t]\times O \ni (s,y) \mapsto \mu(t+s,y) \in \mathbb{R}^d)$, $\mu_2 \curvearrowleft ([0,T-t]\times O \ni (s,y) \mapsto m_n(t+s,y) \in \mathbb{R}^d)$, $\sigma_1 \curvearrowleft ([0,T-t]\times O \ni (s,y) \mapsto \sigma(t+s,y) \in \mathbb{R}^{d\times m})$, $\sigma_2 \curvearrowleft ([0,T-t]\times O \ni (s,y) \mapsto s_n(t+s,y) \in \mathbb{R}^{d\times m})$, $\mathbb{F} \curvearrowleft (\mathbb{F}_{t+s})_{s\in[0,T-t]}$, $W \curvearrowleft ([0,T-t]\times\Omega \ni (s,\omega) \mapsto W_{t-s}(\omega) - W_t(\omega) \in \mathbb{R}^m)$, $X^{(1)} \curvearrowleft ([0,T-t]\times\Omega \ni (s,\omega) \mapsto X^x_{t,t+s} \in O)$, $X^{(2)} \curvearrowleft ([0,T-t]\times\Omega \ni (s,\omega) \mapsto \mathcal{X}^{x,n}_{t,t+s} \in \mathbb{R}^d)$, $\tau \curvearrowleft \tau^{n,t,x} - t$ in the notation of [13, Lemma 3.5]), item (II), (1.117), and the fact that O_n is a compact set demonstrate that for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O$ it holds that

$$\mathbb{P}(\forall s \in [t,T] \colon \mathbb{1}_{\{s\leqslant\tau^{n,t,x}\}}\|\mathcal{X}^{x,n}_{t,s} - X^x_{t,s}\| = 0) = 1. \tag{1.129}$$

This implies that for all $\varepsilon \in (0,\infty)$, $n \in \mathbb{N}$, $t \in [0,T]$, $s \in [t,T]$, $x \in O$ it holds that

$$\begin{aligned} &\mathbb{P}(\|\mathcal{X}^{x,n}_{t,s} - X^x_{t,s}\| \geqslant \varepsilon) \leqslant \mathbb{P}(\tau^{n,t,x} < s) \\ &\leqslant \mathbb{P}(\|X^x_{t,\tau^{n,t,x}}\| \geqslant n) \leqslant \sup_{\substack{\tau\colon\Omega\to[t,T]\\ \text{stopping time}}} \mathbb{P}(\|X^x_{t,\tau}\| \geqslant n). \end{aligned} \tag{1.130}$$

Moreover, observe that [13, Lemma 3.6] and items (I) and (III) prove that there exist c_k, $k \in \mathbb{N}$, which satisfy for all $k,n \in \mathbb{N}$, $s \in [t_0,T]$ that

$$\mathbb{E}\big[\|\mathcal{X}^{x_n,k}_{t_n,\max\{s,t_n\}} - \mathcal{X}^{x_0,k}_{t_0,\max\{s,t_0\}}\|^2\big] \leqslant c_k\big[|t_n - t_0| + \|x_n - x_0\|^2\big]. \tag{1.131}$$

Furthermore, note that (1.127) shows that for all $k,n \in \mathbb{N}$, $s \in [t_0,T]$ it holds a.s. that

$$\begin{aligned} &\mathcal{X}^{x_0,k}_{t_0,\max\{s,t_n\}} - \mathcal{X}^{x_0,k}_{t_0,\max\{s,t_0\}} \\ &= \int_s^{\max\{s,t_n\}} m_k(r,\mathcal{X}^{x_0,k}_{t_0,r})\,\mathrm{d}r + \int_s^{\max\{s,t_n\}} s_k(r,\mathcal{X}^{x_0,k}_{t_0,r})\,\mathrm{d}W_r. \end{aligned} \tag{1.132}$$

Combining this with Minkowski's inequality, the Burkholder-Davis-Gundy inequality, and the fact that m_n, $n \in \mathbb{N}$, and s_n, $n \in \mathbb{N}$, are continuous functions with compact support implies that for all $k,n \in \mathbb{N}$, $s \in [t_0,T]$ it holds that

$$\begin{aligned} &\Big(\mathbb{E}\Big[\big\|\mathcal{X}^{x_0,k}_{t_0,\max\{s,t_n\}} - \mathcal{X}^{x_0,k}_{t_0,\max\{s,t_0\}}\big\|^2\Big]\Big)^{\frac{1}{2}} \\ &\leqslant \int_s^{\max\{s,t_n\}} \Big(\mathbb{E}\Big[\big\|m_k(r,\mathcal{X}^{x_0,k}_{t_0,r})\big\|^2\Big]\Big)^{\frac{1}{2}}\,\mathrm{d}r \\ &\quad + \bigg(\int_s^{\max\{s,t_n\}} \mathbb{E}\Big[\big\|\big|s_k(r,\mathcal{X}^{x_0,k}_{t_0,r})\big\|\big|^2\Big]\,\mathrm{d}r\bigg)^{\frac{1}{2}} \\ &\leqslant |\max\{t_n - s, 0\}|^{\frac{1}{2}}\bigg[\sqrt{T}\Big(\sup_{t\in[0,T]}\sup_{x\in O}\|m_k(t,x)\|\Big) + \sup_{t\in[0,T]}\sup_{x\in O}|||s_k(t,x)|||\bigg] < \infty. \end{aligned} \tag{1.133}$$

The fact that for all $a,b \in \mathbb{R}$ it holds that $(a+b)^2 \leqslant 2(a^2+b^2)$ and (1.131) therefore show that there exist $\tilde{c}_k \in [0,\infty)$, $k \in \mathbb{N}$, which satisfy for all $k,n \in \mathbb{N}$, $s \in [t_0,T]$ that

$$\begin{aligned} &\mathbb{E}\Big[\big\|\mathcal{X}^{x_n,k}_{t_n,\max\{s,t_n\}} - \mathcal{X}^{x_0,k}_{t_0,\max\{s,t_0\}}\big\|^2\Big] \\ &\leqslant 2\mathbb{E}\Big[\big\|\mathcal{X}^{x_n,k}_{t_n,\max\{s,t_n\}} - \mathcal{X}^{x_0,k}_{t_0,\max\{s,t_n\}}\big\|^2\Big] + 2\mathbb{E}\Big[\big\|\mathcal{X}^{x_0,k}_{t_0,\max\{s,t_n\}} - \mathcal{X}^{x_0,k}_{t_0,\max\{s,t_0\}}\big\|^2\Big] \\ &\leqslant \tilde{c}_k\big[|t_n - t_0| + \|x_n - x_0\|^2 + \max\{t_n - s, 0\}\big]. \end{aligned} \tag{1.134}$$

In addition, observe that (1.127) shows that for all $k, n \in \mathbb{N}$, $s \in [0, t_0]$ it holds a.s. that

$$\begin{aligned}&\mathcal{X}^{x_n,k}_{t_n,\max\{s,t_n\}} - \mathcal{X}^{x_0,k}_{t_0,\max\{s,t_0\}} = \mathcal{X}^{x_n,k}_{t_n,\max\{s,t_n\}} - x_0\\&= x_n - x_0 + \int_{t_n}^{\max\{s,t_n\}} m_k(r, \mathcal{X}^{x_n,k}_{t_n,r})\,\mathrm{d}r + \int_{t_n}^{\max\{s,t_n\}} s_k(r, \mathcal{X}^{x_n,k}_{t_n,r})\,\mathrm{d}W_r.\end{aligned} \tag{1.135}$$

Minkowski's inequality, Itô's isometry and the fact that m_n, $n \in \mathbb{N}$, and s_n, $n \in \mathbb{N}$, are continuous functions with compact support ensure that for all $k, n \in \mathbb{N}$, $s \in [0, t_0]$ it holds that

$$\begin{aligned}&\left(\mathbb{E}\Big[\big\|\mathcal{X}^{x_n,k}_{t_n,\max\{s,t_n\}} - \mathcal{X}^{x_0,k}_{t_0,\max\{s,t_0\}}\big\|^2\Big]\right)^{\frac{1}{2}} - \|x_n - x_0\|\\&\leqslant \int_{t_n}^{\max\{s,t_n\}} \left(\mathbb{E}\Big[\big\|m_k(r, \mathcal{X}^{x_n,k}_{t_n,r})\big\|^2\Big]\right)^{\frac{1}{2}}\mathrm{d}r\\&\quad + \left(\int_{t_n}^{\max\{s,t_n\}} \mathbb{E}\Big[\big|\!\big|\!\big|s_k(r, \mathcal{X}^{x_n,k}_{t_n,r})\big|\!\big|\!\big|^2\Big]\,\mathrm{d}r\right)^{\frac{1}{2}}\\&\leqslant |\max\{t_n - s, 0\}|^{\frac{1}{2}}\left[\sqrt{T}\Big(\sup_{t\in[0,T]}\sup_{x\in O}\|m_k(t,x)\|\Big) + \sup_{t\in[0,T]}\sup_{x\in O}|\!|\!|s_k(t,x)|\!|\!|\right]\\&< \infty.\end{aligned} \tag{1.136}$$

This and (1.134) imply that there exist $\gamma_k \in [0, \infty)$, $k \in \mathbb{N}$, which satisfy for all $k, n \in \mathbb{N}$, $s \in [0, T]$ that

$$\begin{aligned}&\mathbb{E}\Big[\big\|\mathcal{X}^{x_n,k}_{t_n,\max\{s,t_n\}} - \mathcal{X}^{x_0,k}_{t_0,\max\{s,t_0\}}\big\|^2\Big]\\&\leqslant \gamma_k\Big[|t_n - t_0| + \|x_n - x_0\|^2 + \mathbb{1}_{[0,t_0]}(s)\max\{s - t_n, 0\}\\&\qquad + \mathbb{1}_{[t_0,T]}(s)\max\{t_n - s, 0\}\Big].\end{aligned} \tag{1.137}$$

Next note that the fact that $\limsup_{n\to\infty}[|t_n - t_0| + \|x_n - x_0\|] = 0$ ensures that there exists a compact set $\tilde{\mathcal{K}} \subseteq [0, T] \times O$ which satisfies for all $n \in \mathbb{N}_0$ that $(t_n, x_n) \in \tilde{\mathcal{K}}$. Markov's inequality, (1.119), (1.130), and (1.137) hence show that for all $\varepsilon \in (0, \infty)$, $s \in [0, T]$ it holds that

$$
\begin{aligned}
&\limsup_{n\to\infty}\Big[\mathbb{P}\Big(\|X^{x_n}_{t_n,\max\{s,t_n\}}-X^{x_0}_{t_0,\max\{s,t_0\}}\|\geqslant\varepsilon\Big)\Big]\\
&\leqslant\inf_{k\in N}\Bigg(\limsup_{n\to\infty}\Bigg[\mathbb{P}\Big(\|X^{x_n}_{t_n,\max\{s,t_n\}}-\mathcal{X}^{x_n,k}_{t_n,\max\{s,t_n\}}\|\geqslant\frac{\varepsilon}{3}\Big)\\
&\quad+\mathbb{P}\Big(\|\mathcal{X}^{x_n,k}_{t_n,\max\{s,t_n\}}-\mathcal{X}^{x_0,k}_{t_0,\max\{s,t_0\}}\|\geqslant\frac{\varepsilon}{3}\Big)\\
&\quad+\mathbb{P}\Big(\|\mathcal{X}^{x_0,k}_{t_0,\max\{s,t_0\}}-X^{x_0}_{t_0,\max\{s,t_0\}}\|\geqslant\frac{\varepsilon}{3}\Big)\Bigg]\Bigg)\\
&\leqslant\inf_{k\in N}\Bigg(\limsup_{n\to\infty}\Bigg[\sup_{\substack{\tau\colon\Omega\to[t_n,T]\\\text{stopping time}}}\mathbb{P}(\|X^{x_n}_{t_n,\tau}\|\geqslant k)\\
&\quad+\frac{9}{\varepsilon^2}\,\mathbb{E}\Big[\|\mathcal{X}^{x_n,k}_{t_n,\max\{s,t_n\}}-\mathcal{X}^{x_0,k}_{t_0,\max\{s,t_0\}}\|^2\Big]+\sup_{\substack{\tau\colon\Omega\to[t_0,T]\\\text{stopping time}}}\mathbb{P}(\|X^{x_0}_{t_0,\tau}\|\geqslant k)\Bigg]\Bigg)\\
&\leqslant\inf_{k\in N}\Bigg(\limsup_{n\to\infty}\Bigg[\sup_{\substack{\tau\colon\Omega\to[t_n,T]\\\text{stopping time}}}\mathbb{P}(\|X^{x_n}_{t_n,\tau}\|\geqslant k)+\sup_{\substack{\tau\colon\Omega\to[t_0,T]\\\text{stopping time}}}\mathbb{P}(\|X^{x_0}_{t_0,\tau}\|\geqslant k)\\
&\quad+\frac{9\gamma_k}{\varepsilon^2}(|t_n-t_0|+\|x_n-x_0\|^2+\mathbb{1}_{[0,t_0]}(s)\max\{s-t_n,0\}\\
&\quad+\mathbb{1}_{[t_0,T]}(s)\max\{t_n-s,0\})\Bigg]\Bigg)\\
&=\inf_{k\in\mathbb{N}}\Bigg(\limsup_{n\to\infty}\Bigg[\sup_{\substack{\tau\colon\Omega\to[t_n,T]\\\text{stopping time}}}\mathbb{P}(\|X^{x_n}_{t_n,\tau}\|\geqslant k)+\sup_{\substack{\tau\colon\Omega\to[t_0,T]\\\text{stopping time}}}\mathbb{P}(\|X^{x_0}_{t_0,\tau}\|\geqslant k)\Bigg]\Bigg)\\
&\leqslant\inf_{k\in\mathbb{N}}\Bigg(\sup_{(u,y)\in\tilde{\mathcal{K}}}\Bigg[2\sup_{\substack{\tau\colon\Omega\to[u,T]\\\text{stopping time}}}\mathbb{P}(\|X^{y}_{u,\tau}\|\geqslant k)\Bigg]\Bigg)=0.
\end{aligned}
\tag{1.138}
$$

This establishes (1.120). The proof of Lemma 1.2.3 is thus complete. □

The next lemma uses the results in Lemma 1.2.3 to prove convergence in mean for the stochastic process in (1.140) below which is the stochastic process arising from the Bismut-Elworthy-Li formula.

Lemma 1.2.4. *Assume Setting 1.2.1, for every $r \in (0,\infty)$ let $O_r \subseteq O$ satisfy $O_r = \{x \in O\colon \|x\| \leqslant r$ and $\{y \in \mathbb{R}^d\colon \|y-x\| < \frac{1}{r}\} \subseteq O\}$, assume for all $r \in (0,\infty)$, $j \in \{1,2,\ldots,d\}$ that*

$$
\sup\Bigg(\Bigg\{\frac{\|\frac{\partial\mu}{\partial x}(t,x)-\frac{\partial\mu}{\partial x}(t,y)\|_F+\|\frac{\partial\sigma}{\partial x_j}(t,x)-\frac{\partial\sigma}{\partial x_j}(t,y)\|_F}{\|x-y\|}\colon\\ t\in[0,T],x,y\in O_r,x\neq y\Bigg\}\cup\{0\}\Bigg)<\infty,
\tag{1.139}
$$

and for every $t \in [0,T]$, $x \in O$ let $Z^x_t = (Z^x_{t,s})_{s\in(t,T]}\colon (t,T]\times\Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_s)_{s\in(t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in (t,T]$ it holds a.s. that

$$
Z^x_{t,s}=\frac{1}{s-t}\int_t^s(\sigma(r,X^x_{t,r}))^{-1}\left(\frac{\partial}{\partial x}X^x_{t,r}\right)\mathrm{d}W_r.
\tag{1.140}
$$

Then it holds for all $t \in [0,T)$, $s \in (t,T]$, $x \in O$ that

$$
\limsup\nolimits_{[0,s)\times O\ni(u,y)\to(t,x)}\Big[\mathbb{E}\big[\|Z^y_{u,s}-Z^x_{t,s}\|\big]\Big]=0.
\tag{1.141}
$$

Proof of Lemma 1.2.4. Throughout this proof let $\mathfrak{m} \in [0, \infty)$ satisfy $\mathfrak{m} = \max_{s \in [0,T]} [\frac{1}{2}\|\mu(s,0)\|^2 + \|\sigma(s,0)\|_F^2]$ and for every $t \in [0,T]$, $x \in O$ let $Y_t^x = (Y_{t,s}^x)_{s \in [t,T]} \colon [t,T] \times \Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_s)_{s \in [t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in [t,T]$ it holds a.s. that

$$Y_{t,s}^x = \int_t^s (\sigma(r, X_{t,r}^x))^{-1} \left(\frac{\partial}{\partial x} X_{t,r}^x\right) \mathrm{d}W_r. \tag{1.142}$$

Next observe that Markov's inequality, items (i) and (iii) of Lemma 1.2.2 and the fact that for all $a, b \in \mathbb{R}$ it holds that $(a+b)^2 \leqslant 2(a^2 + b^2)$ show that for all $t \in [0,T]$, $x \in O$, and all stopping times $\tau \colon \Omega \to [t,T]$ it holds that

$$\begin{aligned} &\mathbb{P}\Big(\|X_{t,\tau}^x\| + \big\|\frac{\partial}{\partial x} X_{t,\tau}^x\big\|_F \geqslant k\Big) \leqslant \frac{1}{k^2}\,\mathbb{E}\Big[\big(\|X_{t,\tau}^x\| + \big\|\frac{\partial}{\partial x} X_{t,\tau}^x\big\|_F\big)^2\Big] \\ &\leqslant \frac{2}{k^2}\,\mathbb{E}\Big[\|X_{t,\tau}^x\|^2 + \big\|\frac{\partial}{\partial x} X_{t,\tau}^x\big\|_F^2\Big] \\ &\leqslant \frac{2}{k^2}\Big(\exp((4c+2)T)\Big(\|x\|^2 + \frac{\mathfrak{m}}{2c+1}\Big) + d\exp(2cT)\Big). \end{aligned} \tag{1.143}$$

This implies that for all non-empty, compact $\mathcal{K} \subseteq [0,T] \times O$ it holds that

$$\begin{aligned} &\inf_{k \in \mathbb{N}} \Big[\sup_{(t,x) \in \mathcal{K}} \Big(\sup_{\tau \colon \Omega \to [t,T] \text{ stopping time}} \mathbb{P}\Big(\|X_{t,\tau}^x\| + \big\|\frac{\partial}{\partial x} X_{t,\tau}^x\big\|_F \geqslant k\Big)\Big)\Big] \\ &\leqslant \inf_{k \in \mathbb{N}} \Big[\sup_{(t,x) \in \mathcal{K}} \Big(\frac{2}{k^2}\Big(\exp((4c+2)T)\Big(\|x\|^2 + \frac{\mathfrak{m}}{2c+1}\Big) + d\exp(2cT)\Big)\Big)\Big] \\ &= 0. \end{aligned} \tag{1.144}$$

Lemma 1.2.3 (applied with $d \curvearrowleft 2d$, $m \curvearrowleft d$, $O \curvearrowleft O \times \mathbb{R}^d$, $\mu \curvearrowleft ([0,T] \times \mathbb{R}^d \times \mathbb{R}^d \ni (t,x,z) \mapsto (\mu(t,x), (\frac{\partial \mu}{\partial x})(t,x)z) \in \mathbb{R}^d \times \mathbb{R}^d)$, $\sigma \curvearrowleft ([0,T] \times \mathbb{R}^d \times \mathbb{R}^d \ni (t,x,z) \mapsto (\sigma(t,x), \sum_{i=1}^d (\frac{\partial \sigma}{\partial x_i})(t,x)z_i) \in \mathbb{R}^{d \times d} \times \mathbb{R}^{d \times d})$, and for every $j \in \{1, 2, \ldots, d\}$, $t \in [0,T]$, $x \in O$ with $X_t^x \curvearrowleft ([t,T] \times \Omega \ni (s,\omega) \mapsto (X_{t,s}^x(\omega), \frac{\partial}{\partial x_j} X_{t,s}^x(\omega)) \in O \times \mathbb{R}^d)$ in the notation of Lemma 1.2.3) and (1.139) therefore demonstrate that for all $j \in \{1, 2, \ldots, d\}$, $\varepsilon \in (0, \infty)$, $s \in [0,T]$, and all $(t_n, x_n) \in [0,T] \times O$, $n \in \mathbb{N}_0$, with $\limsup_{n \to \infty} [|t_n - t_0| + \|x_n - x_0\|] = 0$ it holds that

$$\begin{aligned} \limsup_{n \to \infty} \Big[\mathbb{P}\Big(&\|X_{t_n, \max\{s,t_n\}}^{x_n} - X_{t_0, \max\{s,t_0\}}^{x_0}\| \\ &+ \big\|\frac{\partial}{\partial x_j} X_{t_n, \max\{s,t_n\}}^{x_n} - \frac{\partial}{\partial x_j} X_{t_0, \max\{s,t_0\}}^{x_0}\big\| \geqslant \varepsilon\Big)\Big] = 0. \end{aligned} \tag{1.145}$$

Furthermore, note that the assumption that σ is continuous implies that σ^{-1} is continuous. Combining this with (1.145) and the Continuous Mapping Theorem shows that for all $\varepsilon \in (0, \infty)$, $s \in [0,T]$, $(t_n, x_n) \in [0,T] \times O$, $n \in \mathbb{N}$, with $\limsup_{n \to \infty} [|t_n - t_0| + \|x_n - x_0\|] = 0$ it holds that

$$\begin{aligned} \limsup_{n \to \infty} [\mathbb{P}(\|(\sigma(\max\{s,t_n\}, X_{t_n, \max\{s,t_n\}}^{x_n}))^{-1} \\ - (\sigma(\max\{s,t_0\}, X_{t_0, \max\{s,t_0\}}^{x_0}))^{-1}\| \geqslant \varepsilon] = 0. \end{aligned} \tag{1.146}$$

This and (1.145) ensure that for all $\varepsilon \in (0, \infty)$, $s \in [0,T]$, $(t_n, x_n) \in [0,T] \times O$, $n \in \mathbb{N}$, with $\limsup_{n \to \infty} [|t_n - t_0| + \|x_n - x_0\|] = 0$ it holds that

$$\begin{aligned} \limsup_{n \to \infty} \Big[\mathbb{P}\Big(\Big\|&(\sigma(\max\{s,t_n\}, X_{t_n, \max\{s,t_n\}}^{x_n}))^{-1} \frac{\partial}{\partial x} X_{t_n, \max\{s,t_n\}}^{x_n} \\ &- (\sigma(\max\{s,t_0\}, X_{t_0, \max\{s,t_0\}}^{x_0}))^{-1} \frac{\partial}{\partial x} X_{t_0, \max\{s,t_0\}}^{x_0}\Big\| \geqslant \varepsilon\Big)\Big] = 0. \end{aligned} \tag{1.147}$$

Next observe that items (iii) and (iv) of Lemma 1.2.2 show that for all $t \in [0,T]$, $s \in (t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\mathbb{E}\left[\left\|(\sigma(s,X_{t,s}^x))^{-1}\Big(\frac{\partial}{\partial x}X_{t,s}^x\Big)\right\|_F^2\right] \leqslant \mathbb{E}\left[\|(\sigma(s,X_{t,s}^x))^{-1}\|_{L(\mathbb{R}^d)}^2\left\|\frac{\partial}{\partial x}X_{t,s}^x\right\|_F^2\right] \\ &\leqslant \frac{1}{\alpha}\mathbb{E}\left[\left\|\frac{\partial}{\partial x}X_{t,s}^x\right\|_F^2\right] = \frac{1}{\alpha}\mathbb{E}\left[\left\|\frac{\partial}{\partial x}X_{t,s}^x\right\|_F^2\right] \leqslant \frac{d}{\alpha}\exp(2cT). \end{aligned} \tag{1.148}$$

Combining this and (1.147) with [68, Corollary 18.13] (applied for every $n \in \mathbb{N}$, $t \in [0,T]$, $s \in (t,T]$, $x \in O$, $(t_m)_{m\in\mathbb{N}} \subseteq [0,s)$ which satisfies $\lim_{m\to\infty} t_m = t$, $(x_m)_{m\in\mathbb{N}} \subseteq O$ which satisfies $\lim_{m\to\infty} x_m = x$ with $X \curvearrowleft W$, $U \curvearrowleft (\frac{d}{\alpha})^{1/2}\exp(cT)$, $V \curvearrowleft (([0,T]\times\Omega \ni (s,\omega) \mapsto \sigma(s,X_{t,s}^x(\omega)))^{-1}(\frac{\partial}{\partial x}X_{t,s}^x(\omega) \in \mathbb{R}^{d\times d}))$, $V_n \curvearrowleft (\Omega\times[0,T] \ni (\omega,s) \mapsto (\sigma(s,X_{t_n,s}^{x_n}(\omega)))^{-1} \cdot(\frac{\partial}{\partial x}X_{t_n,s}^{x_n}(\omega)) \in \mathbb{R}^{d\times d})$ in the notation of [68, Corollary 18.13]) ensures that for all $\varepsilon \in (0,\infty)$, $s \in [0,T]$, and all $(t_n,x_n) \in [0,T]\times O$, $n \in \mathbb{N}_0$, with $\limsup_{n\to\infty}[|t_n-t_0| + \|x_n - x_0\|] = 0$ it holds that

$$\limsup\nolimits_{n\to\infty}\left[\mathbb{P}\Big(\|Y_{t_n,\max\{s,t_n\}}^{x_n} - Y_{t_0,\max\{s,t_0\}}^{x_0}\| \geqslant \varepsilon\Big)\right] = 0. \tag{1.149}$$

In addition, note that item (v) of Lemma 1.2.2 ensures that for all $t \in [0,T]$, $x \in O$, $s \in [t,T]$ it holds that

$$\mathbb{E}\big[\|Y_{t,s}^x\|^2\big] \leqslant \frac{dT}{\alpha}\exp(2cT). \tag{1.150}$$

Next observe that Hölder's inequality demonstrates that for all $\varepsilon \in (0,\infty)$, $s \in (0,T]$, $u,t \in [0,s)$, $x,y \in O$ it holds that

$$\begin{aligned} \mathbb{E}\big[\|Y_{u,s}^y - Y_{t,s}^x\|\big] &\leqslant \varepsilon + \mathbb{E}\big[\|Y_{u,s}^y - Y_{t,s}^x\|\mathbb{1}_{\{\|Y_{u,s}^y - Y_{t,s}^x\|>\varepsilon\}}\big] \\ &\leqslant \varepsilon + \Big(\mathbb{E}\big[\|Y_{u,s}^y - Y_{t,s}^x\|^2\big]\Big)^{\frac12}\Big(\mathbb{E}\big[\mathbb{1}_{\{\|Y_{u,s}^y - Y_{t,s}^x\|>\varepsilon\}}\big]\Big)^{\frac12} \\ &\leqslant \varepsilon + \Big(\mathbb{E}\big[\|Y_{u,s}^y - Y_{t,s}^x\|^2\big]\Big)^{\frac12}\Big(\mathbb{P}\big(\|Y_{u,s}^y - Y_{t,s}^x\| > \varepsilon\big)\Big)^{\frac12}. \end{aligned} \tag{1.151}$$

Combining this with (1.149) and (1.150) proves that for all $t \in [0,T)$, $s \in [t,T]$, $x \in O$ it holds that

$$\limsup\nolimits_{[0,s)\times O \ni (u,y)\to(t,x)}\Big[\mathbb{E}\big[\|Y_{u,s}^y - Y_{t,s}^x\|\big]\Big] = 0. \tag{1.152}$$

Throughout the rest of the proof let $(t_0,x_0) \in [0,T]\times O$, $s_0 \in (t_0,T]$, and $(t_n,x_n) \in [0,s_0)\times O$, $n \in \mathbb{N}$, satisfy that $\limsup_{n\to\infty}[|t_n - t_0| + \|x_n - x_0\|] = 0$. Observe that for the proof of (1.141) it is sufficient to show that

$$\limsup_{n\to\infty}\Big[\mathbb{E}\big[\|Z_{t_n,s_0}^{x_n} - Z_{t_0,s_0}^{x_0}\|\big]\Big] = 0. \tag{1.153}$$

Next note that the fact that $\limsup_{n\to\infty}[|t_n - t_0|] = 0$ ensures that there exists $r \in \mathbb{N}$ which satisfies that for all $n \in \mathbb{N}$ it holds that $t_n \in [0, s_0 - \frac{1}{r}]$. The triangle inequality hence demonstrates that for all $n \in \mathbb{N}$ it holds that

$$\begin{aligned} &\mathbb{E}\big[\|Z_{t_n,s_0}^{x_n} - Z_{t_0,s_0}^{x_0}\|\big] = \mathbb{E}\left[\left\|\frac{1}{s_0-t_n}Y_{t_n,s_0}^{x_n} - \frac{1}{s_0-t_0}Y_{t_0,s_0}^{x_0}\right\|\right] \\ &\leqslant \frac{1}{s_0-t_n}\mathbb{E}\big[\|Y_{t_n,s_0}^{x_n} - Y_{t_0,s_0}^{x_0}\|\big] + \left|\frac{1}{s_0-t_n} - \frac{1}{s_0-t_0}\right|\mathbb{E}\big[\|Y_{t_0,s_0}^{x_0}\|\big] \\ &\leqslant \sup_{u\in[0,T-\frac1r]}\left[\frac{1}{s_0-u}\right]\mathbb{E}\Big[\|Y_{t_n,s_0}^{x_n} - Y_{t_0,s_0}^{x_0}\|\Big] \\ &\quad + \left|\frac{1}{s_0-t_n} - \frac{1}{s_0-t_0}\right|\Big(\mathbb{E}\big[\|Y_{t_0,s_0}^{x_0}\|^2\big]\Big)^{\frac12}. \end{aligned} \tag{1.154}$$

Combining this with (1.150) and (1.152) demonstrates (1.153). The proof of Lemma 1.2.4 is thus complete. □

1.2.3 Existence and uniqueness properties of solutions of SFPEs associated with SDEs

In the following theorem we combine Theorem 1.1.11 with the results in Section 1.2.1-1.2.2 to obtain this section's main result, Theorem 1.2.5. Theorem 1.2.5 extends the abstract existence and uniqueness result in Theorem 1.1.11 to a class of SFPEs associated with certain SDEs. In this regard, Theorem 1.2.5 generalizes [13, Theorem 3.8] to the case of gradient-dependent nonlinearities under more restrictive assumptions.

Theorem 1.2.5. *Assume Setting 1.2.1, let $b, K, L \in (0,\infty)$, let $|||\cdot|||\colon \mathbb{R}^{d+1} \to [0,\infty)$ satisfy for all $x = (x_1,\ldots,x_{d+1}) \in \mathbb{R}^{d+1}$ that $|||x||| = (\sum_{i=1}^{d+1}|x_i|^2)^{1/2}$, for every $r \in (0,\infty)$ let $K_r \subseteq [0,T)$, $O_r \subseteq O$ satisfy $K_r = [0, \max\{T - \frac{1}{r}, 0\}]$ and $O_r = \{x \in O\colon \|x\| \leqslant r$ and $\{y \in \mathbb{R}^d\colon \|y - x\| < \frac{1}{r}\} \subseteq O\}$, assume for all $r \in (0,\infty)$, $j \in \{1,\ldots,d\}$ that*

$$\sup\left(\left\{\frac{\|\frac{\partial\mu}{\partial x}(t,x) - \frac{\partial\mu}{\partial x}(t,y)\|_F + \|\frac{\partial\sigma}{\partial x_j}(t,x) - \frac{\partial\sigma}{\partial x_j}(t,y)\|_F}{\|x-y\|}\colon t \in [0,T], x,y \in O_r, x \neq y\right\} \cup \{0\}\right) < \infty, \tag{1.155}$$

for every $t \in [0,T]$, $x \in O$ let $Z^x_t = (Z^x_{t,s})_{s\in(t,T]}\colon (t,T] \times \Omega \to \mathbb{R}^{d+1}$ be an $(\mathbb{F}_s)_{s\in(t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in (t,T]$ it holds a.s. that

$$Z^x_{t,s} = \begin{pmatrix} 1 \\ \frac{1}{s-t}\int_t^s (\sigma(r, X^x_{t,r}))^{-1}\left(\frac{\partial}{\partial x}X^x_{t,r}\right) dW_r, \end{pmatrix} \tag{1.156}$$

let $V \in C^{1,2}([0,T] \times O, (0,\infty))$ satisfy a.s. for all $t \in [0,T]$, $s \in [t,T]$, $x \in O$ that

$$\begin{split}(\tfrac{\partial V}{\partial s})(s, X^x_{t,s}) + \langle(\tfrac{\partial V}{\partial x})(s, X^x_{t,s}), \mu(s, X^x_{t,s})\rangle \\ + \tfrac{1}{2}\operatorname{Trace}(\sigma(s, X^x_{t,s})[\sigma(s, X^x_{t,s})]^*(\operatorname{Hess}_x V)(s, X^x_{t,s})) \leqslant KV(s, X^x_{t,s}) + b,\end{split} \tag{1.157}$$

let $f \in C([0,T) \times O \times \mathbb{R}^{d+1}, \mathbb{R})$, $g \in C(O, \mathbb{R})$ satisfy for all $t \in [0,T)$, $x \in O$, $v, w \in \mathbb{R}^{d+1}$ that $|f(t,x,v) - f(t,x,w)| \leqslant L|||v - w|||$, and assume that $\inf_{r\in(0,\infty)}[\sup_{t\in[0,T)\setminus K_r} \sup_{x\in O\setminus O_r}(\frac{|g(x)|^2}{V(T,x)}) + \frac{|f(t,x,0)|^2}{V(t,x)}(T-t)] = 0$, and $\liminf_{r\to\infty}[\inf_{t\in[0,T]}\inf_{x\in O\setminus O_r} V(t,x)] = \infty$. Then there exists a unique $v \in C([0,T) \times O, \mathbb{R}^{d+1})$ such that

(i) it holds that

$$\limsup_{r\to\infty}\left[\sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}\left(\frac{|||v(t,x)|||}{\sqrt{V(t,x)}}\sqrt{T-t}\right)\right] = 0, \tag{1.158}$$

(ii) it holds that

$$[0,T) \times O \ni (t,x) \mapsto \mathbb{E}\left[g(X^x_{t,T})Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, v(r, X^x_{t,r}))Z^x_{t,r}\, dr\right] \in \mathbb{R}^{d+1} \tag{1.159}$$

is well-defined and continuous, and

(iii) for all $t \in [0,T)$, $x \in O$ it holds that

$$v(t,x) = \mathbb{E}\left[g(X^x_{t,T}) Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, v(r, X^x_{t,r})) Z^x_{t,r} \, dr \right]. \tag{1.160}$$

Proof of Theorem 1.2.5. First note that the assumptions that $V \in C([0,T] \times O, (0,\infty))$ and that $\liminf_{r\to\infty}[\inf_{t\in[0,T]} \inf_{x\in O\setminus O_r} V(t,x)] = \infty$ ensure that

$$\inf_{t\in[0,T]} \inf_{x\in O} V(t,x) > 0. \tag{1.161}$$

Furthermore, observe that [50, Theorem 2.4] (applied for every $t \in [0,T]$, $x \in O$ with $H \curvearrowleft \mathbb{R}^d$, $U \curvearrowleft \mathbb{R}^d$, $T \curvearrowleft T-t$, $X \curvearrowleft ([0,T-t] \times \Omega \ni (s,\omega) \mapsto X^x_{t,t+s}(\omega) \in \mathbb{R}^d)$, $a \curvearrowleft ([0,T-t]\times\Omega \ni (s,\omega) \mapsto \mu(t+s,x) \in \mathbb{R}^d)$, $b \curvearrowleft ([0,T-t]\times\Omega \ni (s,\omega) \mapsto \sigma(t+s,x) \in \mathbb{R}^{d\times d})$, $p \curvearrowleft 1$, $\alpha \curvearrowleft ([0,T-t]\times\Omega \ni (s,\omega) \mapsto K \in [0,\infty))$, $\beta \curvearrowleft ([0,T-t]\times\Omega \ni (s,\omega) \mapsto b \in [0,\infty))$, $q_1 \curvearrowleft 1$, $q_2 \curvearrowleft \infty$ in the notation of [50, Theorem 2.4]) and (1.157) demonstrate that for all $t \in [0,T]$, $x \in \mathbb{R}^d$, and all stopping times $\tau \colon \Omega \to [t,T]$ it holds that

$$\begin{aligned} \mathbb{E}\big[V(\tau, X^x_{t,\tau})\big] &\leq \exp(KT)\left(\mathbb{E}[V(t,x)] + b\int_0^T e^{-Kr}\, dr\right) \\ &= \exp(KT)\, V(t,x) + \frac{b}{K}(\exp(KT) - 1). \end{aligned} \tag{1.162}$$

Moreover, note that item (vi) of Lemma 1.2.2 and the fact that for all $a,b \in \mathbb{R}$ it holds that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ show that for all $t \in [0,T]$, $s \in (t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} \left(\mathbb{E}\big[\|Z^x_{t,s}\|^2\big]\right)^{\frac{1}{2}} &\leq \left(1 + \frac{1}{\alpha(s-t)^2}\int_t^s d\exp(2(r-t)c)\, dr\right)^{\frac{1}{2}} \\ &\leq \left(1 + \frac{d}{\alpha(s-t)}\exp(2cT)\right)^{\frac{1}{2}} \leq 1 + \left(\frac{d}{\alpha(s-t)}\exp(2cT)\right)^{\frac{1}{2}}. \end{aligned} \tag{1.163}$$

The Cauchy-Schwarz inequality and (1.162) therefore show that for all $t \in [0,T]$, $s \in (t,T]$, $x \in O$ it holds that

$$\begin{aligned} &\mathbb{E}\left[\sqrt{V(s, X^x_{t,s})}\, \|Z^x_{t,s}\|\right] \leq \left(\mathbb{E}\big[V(s, X^x_{t,s})\big]\right)^{\frac{1}{2}} \left(\mathbb{E}\big[\|Z^x_{t,s}\|^2\big]\right)^{\frac{1}{2}} \\ &\leq \left(\exp(KT)\, V(t,x) + \frac{b}{K}(\exp(KT)-1)\right)^{\frac{1}{2}} \left[1 + \left(\frac{d}{\alpha(s-t)}\exp(2cT)\right)^{\frac{1}{2}}\right] \\ &\leq \frac{1}{\sqrt{s-t}}\left(\exp(KT)\, V(t,x) + \frac{b}{K}(\exp(KT)-1)\right)^{\frac{1}{2}} \left[\sqrt{T} + \left(\frac{d}{\alpha}\exp(2cT)\right)^{\frac{1}{2}}\right] \\ &\leq \frac{\sqrt{V(t,x)}}{\sqrt{s-t}}\left(\exp(KT) + \frac{b}{K\,[\inf_{u\in[0,T]}\inf_{y\in O} V(u,y)]}(\exp(KT)-1)\right)^{\frac{1}{2}} \\ &\qquad \cdot \left[\sqrt{T} + \left(\frac{d}{\alpha}\exp(2cT)\right)^{\frac{1}{2}}\right]. \end{aligned} \tag{1.164}$$

Furthermore, observe that for all $k \in \mathbb{N}$, $t \in [0,T]$, $x \in O$, and all stopping times $\tau \colon \Omega \to [t,T]$ it holds that

$$\begin{aligned} &\mathbb{E}[V(\tau, X^x_{t,\tau})] = \int_\Omega V(\tau(\omega), X^x_{t,\tau(\omega)}(\omega))\, \mathbb{P}(d\omega) \\ &\geq \int_{\{\tilde\omega\in\Omega\colon \|X^x_{t,\tau(\tilde\omega)}(\tilde\omega)\|\geq k\}} V(\tau(\omega), X^x_{t,\tau(\omega)}(\omega))\, \mathbb{P}(d\omega) \\ &\geq \left[\inf_{s\in[t,T]}\inf_{y\in O\setminus O_k} V(s,y)\right] \mathbb{P}(\|X^x_{t,\tau}\| \geq k) \geq \left[\inf_{s\in[0,T]}\inf_{y\in O\setminus O_k} V(s,y)\right] \mathbb{P}(\|X^x_{t,\tau}\| \geq k). \end{aligned} \tag{1.165}$$

Combining this with (1.162) implies that for all $k \in \mathbb{N}$, $t \in [0,T]$, $x \in O$, and all stopping times $\tau \colon \Omega \to [t,T]$ it holds that

$$\begin{aligned} \mathbb{P}(\|X^x_{t,\tau}\| \geqslant k) &\leqslant \frac{\mathbb{E}[V(\tau, X^x_{t,\tau})]}{\inf_{s\in[0,T]}\inf_{y\in O\setminus O_k} V(s,y)} \\ &\leqslant \frac{\exp(KT)\,V(t,x) + \frac{b}{K}(\exp(KT)-1)}{\inf_{s\in[0,T]}\inf_{y\in O\setminus O_k} V(s,y)}. \end{aligned} \tag{1.166}$$

The fact that $\sup_{r\in(0,\infty)}[\inf_{t\in[0,T]}\inf_{x\in O\setminus O_r} V(t,x)] = \infty$ therefore proves that for all non-empty, compact $\mathcal{K} \subseteq [0,T] \times O$ it holds that

$$\begin{aligned} &\inf_{k\in\mathbb{N}}\biggl[\sup_{(t,x)\in\mathcal{K}}\biggl(\sup_{\tau\colon\Omega\to[0,T]\text{ stopping time}} \mathbb{P}\bigl(\|X^x_{t,\tau}\| \geqslant k\bigr)\biggr)\biggr] \\ &\leqslant \inf_{k\in\mathbb{N}}\left[\frac{\exp(KT)[\sup_{(t,x)\in\mathcal{K}} V(t,x)] + \frac{b}{K}(\exp(KT)-1)}{\inf_{s\in[0,T]}\inf_{y\in O\setminus O_k} V(s,y)}\right] \\ &\leqslant \frac{\exp(KT)[\sup_{(t,x)\in\mathcal{K}} V(t,x)] + \frac{b}{K}(\exp(KT)-1)}{\sup_{k\in\mathbb{N}}[\inf_{s\in[0,T]}\inf_{y\in O\setminus O_k} V(s,y)]} = 0. \end{aligned} \tag{1.167}$$

Lemma 1.2.3 and the assumption that $\mu \in C^{0,1}([0,T] \times O, \mathbb{R}^d)$ and $\sigma \in C^{0,1}([0,T] \times O, \mathbb{R}^{d\times d})$ hence demonstrate that for all $\varepsilon \in (0,\infty)$, $s \in [0,T]$, $(t_n, x_n) \in [0,T] \times O$, $n \in \mathbb{N}$, with $\limsup_{n\to\infty}[|t_n - t_0| + \|x_n - x_0\|] = 0$ it holds that

$$\limsup_{n\to\infty}\bigl[\mathbb{P}(\|X^{x_n}_{t_n,\max\{s,t_n\}} - X^{x_0}_{t_0,\max\{s,t_0\}}\|) > \varepsilon\bigr] = 0. \tag{1.168}$$

In addition, observe that Lemma 1.2.4 ensures that for all $t \in [0,T)$, $s \in (t,T]$, $x \in O$ it holds that

$$\limsup_{[0,s)\times O \ni (u,y)\to(t,x)} \bigl[\mathbb{E}\bigl[\|\!|Z^y_{u,s} - Z^x_{t,s}|\!\|\bigr]\bigr] = 0. \tag{1.169}$$

Next note that (1.163) and the fact that for all $t \in [0,T]$ it holds that $\int_t^T \frac{1}{\sqrt{s-t}}\,ds = 2\sqrt{T-t}$ demonstrate that

$$\begin{aligned} &\sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}\left[\mathbb{E}\biggl[\int_t^T \|\!|Z^x_{t,s}|\!\|(ds + \delta_T(ds))\biggr]\sqrt{T-t}\right] \\ &\leqslant \sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}\left[\int_t^T \Bigl(\mathbb{E}\bigl[\|\!|Z^x_{t,s}|\!\|^2\bigr]\Bigr)^{\frac{1}{2}}(ds + \delta_T(ds))\sqrt{T-t}\right] \\ &\leqslant \sup_{t\in[0,T)\setminus K_r}\left[\int_t^T \biggl(1 + \sqrt{\frac{d}{\alpha(s-t)}}\exp(cT)\biggr)(ds + \delta_T(ds))\sqrt{T-t}\right] \\ &= \sup_{t\in[0,T)\setminus K_r}\left[\biggl(T - t + 1 + \sqrt{\frac{d}{\alpha}}\exp(cT)\biggl(\int_t^T \frac{1}{\sqrt{s-t}}ds + \frac{1}{\sqrt{T-t}}\biggr)\biggr)\sqrt{T-t}\right] \\ &= \sup_{t\in[0,T)\setminus K_r}\biggl(\sqrt{T-t}(T-t+1) + \sqrt{\frac{d}{\alpha}}\exp(cT)(2(T-t)+1)\biggr) \\ &\leqslant \sqrt{T}(T+1) + \sqrt{\frac{d}{\alpha}}\exp(cT)(2T+1). \end{aligned} \tag{1.170}$$

This and (1.161) ensure that

$$
\begin{aligned}
&\limsup_{r\to\infty}\left[\sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}\left(\frac{\mathbb{E}\Big[\int_t^T |||Z_{t,s}^x|||(\mathrm{d}s+\delta_T(\mathrm{d}s))\Big]\sqrt{T-t}}{\sqrt{V(t,x)}}\right)\right]\\
&\leqslant \limsup_{r\to\infty}\left[\frac{\sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}\left(\mathbb{E}\Big[\int_t^T |||Z_{t,s}^x|||(\mathrm{d}s+\delta_T(\mathrm{d}s))\Big]\sqrt{T-t}\right)}{\inf_{t\in[0,T)\setminus K_r}\inf_{x\in O\setminus O_r}\sqrt{V(t,x)}}\right]\\
&\leqslant \limsup_{r\to\infty}\left[\frac{\sqrt{T}(T+1)+\sqrt{\frac{d}{\alpha}}\exp(cT)(2T+1)}{\inf_{t\in[0,T)\setminus K_r}\inf_{x\in O\setminus O_r}\sqrt{V(t,x)}}\right]\\
&\leqslant \frac{\sqrt{T}(T+1)+\sqrt{\frac{d}{\alpha}}\exp(cT)(2T+1)}{\liminf_{r\to\infty}[\inf_{t\in[0,T]}\inf_{x\in O\setminus O_r}\sqrt{V(t,x)}]}=0.
\end{aligned}
\tag{1.171}
$$

Combining this, (1.164), (1.168), and (1.169) with Theorem 1.1.11 establishes item (i), (ii), and (iii). The proof of Theorem 1.2.5 is thus complete. □

The following corollary applies the results in Theorem 1.2.5 to the case of $V \curvearrowleft ([0,T]\times \mathbb{R}^d \ni (t,x)\mapsto 1+\|x\|^{q+1}\in(0,\infty))$ for sufficiently large $q\in[1,\infty)$.

Corollary 1.2.6. *Let $d\in\mathbb{N}$, $c\in[1,\infty)$, $T\in(0,\infty)$, let $\|\cdot\|\colon\mathbb{R}^d\to[0,\infty)$, $|||\cdot|||\colon\mathbb{R}^{d+1}\to[0,\infty)$, $\|\cdot\|_F\colon\mathbb{R}^{d\times d}\to[0,\infty)$ be norms, let $(\Omega,\mathcal{F},\mathbb{P},(\mathbb{F}_s)_{s\in[0,T]})$ be a filtered probability space, let $\mu\in C^2(\mathbb{R}^d,\mathbb{R}^d)$, $\sigma\in C^2(\mathbb{R}^d,\mathbb{R}^{d\times d})$ satisfy for all $x,y\in\mathbb{R}^d$ that $y^*\sigma(x)(\sigma(x))^*y\geqslant \frac{1}{c}\|y\|^2$ and*

$$
\max\left\{(\mu(x)-\mu(y))^*(x-y),\|\sigma(x)-\sigma(y)\|_F^2\right\}\leqslant c\|x-y\|^2, \tag{1.172}
$$

for every $t\in[0,T]$, $x\in\mathbb{R}^d$ let $X_t^x=(X_{t,s}^x)_{s\in[t,T]}\colon[t,T]\times\Omega\to\mathbb{R}^d$ be an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s\in[t,T]$ it holds a.s. that

$$
X_{t,s}^x = x+\int_t^s \mu(X_{t,r}^x)\,\mathrm{d}r+\int_t^s\sigma(X_{t,r}^x)\,\mathrm{d}W_r, \tag{1.173}
$$

assume for all $t\in[0,T]$, $\omega\in\Omega$ that $([t,T]\times\mathbb{R}^d\ni(s,x)\mapsto X_{t,s}^x(\omega)\in\mathbb{R}^d)\in C^{0,1}([t,T]\times\mathbb{R}^d,\mathbb{R}^d)$, for every $t\in[0,T]$, $x\in\mathbb{R}^d$ let $Z_t^x=(Z_{t,s}^x)_{s\in(t,T]}\colon(t,T]\times\Omega\to\mathbb{R}^{d+1}$ be an $(\mathbb{F}_s)_{s\in(t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s\in(t,T]$ it holds a.s. that

$$
Z_{t,s}^x=\begin{pmatrix} 1\\ \frac{1}{s-t}\int_t^s(\sigma(X_{t,r}^x))^{-1}\ \left(\frac{\partial}{\partial x}X_{t,r}^x\right)\ \mathrm{d}W_r\end{pmatrix}, \tag{1.174}
$$

let $f\in C([0,T)\times\mathbb{R}^d\times\mathbb{R}^{d+1},\mathbb{R})$, $g\in C(\mathbb{R}^d,\mathbb{R})$ satisfy for all $t\in[0,T)$, $x\in\mathbb{R}^d$, $v,w\in\mathbb{R}^{d+1}$ that $|f(t,x,v)-f(t,x,w)|\leqslant c|||v-w|||$ and $\max\{|g(x)|,|f(t,x,0)|\}\leqslant c(\|x\|^c+1)$. Then there exists a unique $v\in C([0,T)\times\mathbb{R}^d,\mathbb{R}^{d+1})$ such that $(v(t,x)\sqrt{T-t})_{t\in[0,T),x\in\mathbb{R}^d}$ grows at most polynomially and for all $t\in[0,T)$, $x\in\mathbb{R}^d$ it holds that

$$
v(t,x)=\mathbb{E}\left[g(X_{t,T}^x)Z_{t,T}^x+\int_t^T f(r,X_{t,r}^x,v(r,X_{t,r}^x))Z_{t,r}^x\,\mathrm{d}r\right]. \tag{1.175}
$$

Proof of Corollary 1.2.6. Throughout this proof let $V_q\colon[0,T]\times\mathbb{R}^d\to(0,\infty)$, $q\in[1,\infty)$, satisfy that for all $q\in[1,\infty)$, $t\in[0,T]$, $x\in\mathbb{R}^d$ it holds that $V_q(t,x)=1+\|x\|^{q+1}$. First note that for all $q\in[1,\infty)$ it holds that

$$
\liminf_{r\to\infty}\left[\inf_{t\in[0,T]}\inf_{x\in\mathbb{R}^d,\|x\|>r}V_q(t,x)\right]=\liminf_{r\to\infty}\left[\inf_{x\in\mathbb{R}^d,\|x\|>r}\left(1+\|x\|^{q+1}\right)\right]=\infty. \tag{1.176}
$$

In addition, observe that for all $q \in [1, \infty)$, $i, j \in \{1, 2, \ldots, d\}$, $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that $V_q \in C^{1,2}([0, T] \times \mathbb{R}^d, (0, \infty))$ with $\frac{\partial V_q}{\partial x_i}(t, x) = (q + 1)\|x\|^{q-1} x_i$ and

$$\frac{\partial V_q}{\partial x_i x_j}(t, x) = \mathbb{1}_{\mathbb{R}\setminus\{0\}}(x)(q + 1)(q - 1)\|x\|^{q-3} x_i x_j + \mathbb{1}_{\{j\}}(i)(q + 1)\|x\|^{q-1}. \tag{1.177}$$

This, the Cauchy-Schwarz inequality, and (1.172) imply that for all $q \in [1, \infty)$, $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&[\mu(t, x)]^*(\nabla_x V_q)(t, x) = [\mu(t, x)]^*(q + 1)\|x\|^{q-1} x \\
&\leqslant [\mu(t, x) - \mu(t, 0)]^*(q + 1)\|x\|^{q-1} x + [\mu(t, 0)]^*(q + 1)\|x\|^{q-1} x \\
&\leqslant c(q + 1)\|x\|^q + \|\mu(t, 0)\|(q + 1)\|x\|^q \leqslant (q + 1)(c + \|\mu(t, 0)\|)V_q(t, x).
\end{aligned} \tag{1.178}$$

Moreover, note that Young's inequality, (1.172), and (1.177) ensures that for all $q \in [1, \infty)$, $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\tfrac{1}{2} \operatorname{Tr}(\sigma(t, x)[\sigma(t, x)]^*(\operatorname{Hess}_x V_q)(t, x)) \\
&= \tfrac{1}{2} \operatorname{Tr}(\sigma(t, x)[\sigma(t, x)]^*(\mathbb{1}_{\mathbb{R}\setminus\{0\}}(x)(q + 1)(q - 1)\|x\|^{q-3} x_i x_j \\
&\quad + \mathbb{1}_{\{j\}}(i)(q + 1)\|x\|^{q-1})) \\
&\leqslant \frac{q(q + 1)}{2} \|x\|^{q-1} \|\sigma(t, x)\|_F^2 \\
&\leqslant \frac{q(q + 1)}{2} \|x\|^{q-1} \left(2\|\sigma(t, x) - \sigma(t, 0)\|_F^2 + 2\|\sigma(t, 0)\|_F^2\right) \\
&\leqslant cq(q + 1)\|x\|^q + q(q + 1)\|x\|^{q-1}\|\sigma(t, 0)\|_F^2 \\
&\leqslant cq(q + 1)\|x\|^q + (q - 1)(q + 1)\|x\|^q + (q + 1)\|\sigma(t, 0)\|_F^{2q} \\
&\leqslant 2cq(q + 1)V_q(t, x) + (q + 1)\|\sigma(t, 0)\|_F^{2q}.
\end{aligned} \tag{1.179}$$

Combining this with (1.178) and the fact that for all $q \in [1, \infty)$, $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that $\frac{\partial V_q}{\partial t}(t, x) = 0$ proves that for all $q \in [1, \infty)$, $t \in [0, T]$, $x \in O$ it holds that

$$\begin{aligned}
&(\tfrac{\partial V_q}{\partial s})(t, x) + [\mu(t, x)]^*(\nabla_x V_q)(t, x) \\
&\quad + \tfrac{1}{2} \operatorname{Tr}(\sigma(t, x)[\sigma(t, x)]^*(\operatorname{Hess}_x V_q)(t, x)) + \tfrac{1}{2} \tfrac{\|[(\nabla_x V_q)(t,x)]^* \sigma(t,x)\|^2}{V_q(t,x)} \\
&\leqslant (q + 1)\left(2cq + c + \|\mu(t, 0)\|\right) V_q(t, x) + (q + 1)\|\sigma(t, 0)\|_F^{2q}.
\end{aligned} \tag{1.180}$$

Furthermore, note that the assumption that for all $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that $\max\{|g(x)|, |f(t, x, 0)|\} \leqslant c(1 + \|x\|^c)$ ensures that there exists $\alpha \in [c^2, \infty)$ which satisfies

$$\begin{aligned}
&\inf_{r\in(0,\infty)} \left[\sup_{t\in[0,T)} \sup_{x\in\mathbb{R}^d, \|x\|>r} \left(\frac{|g(x)|^2}{V_\alpha(T, x)} + \frac{|f(t, x, 0)|^2}{V_\alpha(t, x)}(T - t)\right)\right] \\
&\leqslant \inf_{r\in(0,\infty)} \left[\sup_{x\in\mathbb{R}^d, \|x\|>r} \left(\frac{c^2(1 + \|x\|^c)^2}{1 + \|x\|^{\alpha+1}} + \frac{c^2 T(1 + \|x\|^c)^2}{1 + \|x\|^{\alpha+1}}\right)\right] = 0.
\end{aligned} \tag{1.181}$$

Combining this with (1.176), (1.180), and Theorem 1.2.5 (applied with $K \leftharpoondown (\alpha+1)(2c\alpha + c + [\sup_{t\in[0,T]}\|\mu(t, 0)\|])$, $b \leftharpoondown (\alpha+1)[\sup_{t\in[0,T]}\|\sigma(t, 0)\|^{2\alpha}]$, $O \leftharpoondown \mathbb{R}^d$, $V \leftharpoondown V_\alpha$, in the notion of Theorem 1.2.5) proves that there exists a unique $v \in C([0, T) \times \mathbb{R}^d, \mathbb{R}^{d+1})$ such that

(I) it holds that

$$\limsup_{r\to\infty} \left[\sup_{t\in[\max\{T-\frac{1}{r},0\},T)} \sup_{x\in\mathbb{R}^d, \|x\|>r} \left(\frac{\|v(t, x)\|}{\sqrt{1 + \|x\|^{\alpha+1}}} \sqrt{T - t}\right)\right] = 0, \tag{1.182}$$

(II) it holds that $[0,T) \times \mathbb{R}^d \ni (t,x) \mapsto \mathbb{E}\Big[g(X^x_{t,T})Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, v(r, X^x_{t,r}))Z^x_{t,r}\,\mathrm{d}r\Big] \in \mathbb{R}^{d+1}$ is well-defined and continuous, and

(III) for all $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$v(t,x) = \mathbb{E}\left[g(X^x_{t,T})Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, v(r, X^x_{t,r}))Z^x_{t,r}\,\mathrm{d}r\right]. \tag{1.183}$$

Let $w \in C([0,T) \times \mathbb{R}^d, \mathbb{R}^{d+1})$ be a function satisfying that $(w(t,x)\sqrt{T-t})_{t\in[0,T),x\in\mathbb{R}^d}$ grows at most polynomially, $[0,T) \times \mathbb{R}^d \ni (t,x) \mapsto \mathbb{E}[g(X^x_{t,T})Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, v(r, X^x_{t,r}))Z^x_{t,r}\,\mathrm{d}r] \in \mathbb{R}^{d+1}$ is well-defined, and for all $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$w(t,x) = \mathbb{E}\left[g(X^x_{t,T})Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, w(r, X^x_{t,r}))Z^x_{t,r}\,\mathrm{d}r\right]. \tag{1.184}$$

Observe that the assumption that $(w(t,x)\sqrt{T-t})_{t\in[0,T),x\in\mathbb{R}^d}$ grows at most polynomially guarantees that there exists $\beta \in [\alpha, \infty]$ which satisfies that

$$\limsup_{r\to\infty}\left[\sup_{t\in[\max\{T-\frac{1}{r},0\},T)}\ \sup_{x\in\mathbb{R}^d,\|x\|>r}\left(\frac{|||w(t,x)|||}{\sqrt{1+\|x\|^{\beta+1}}}\sqrt{T-t}\right)\right] = 0. \tag{1.185}$$

Theorem 1.2.5 (applied with $K \curvearrowleft (\beta+1)(2c\beta + c + [\sup_{t\in[0,T]}\|\mu(t,0)\|])$, $b \curvearrowleft (\beta + 1)[\sup_{t\in[0,T]}\|\sigma(t,0)\|^{2\beta}]$, $O \curvearrowleft \mathbb{R}^d$, $V \curvearrowleft V_\beta$ in the notation of Theorem 1.2.5), (1.176), (1.180), (1.181), and (1.184) hence show that $v = w$. The proof of Corollary 1.2.6 is thus complete. □

Chapter 2

Viscosity solutions

In Chapter 1 we established the existence of a unique solution to SFPEs of the form (1.160) in Theorem 1.2.5. The goal of this chapter is to prove that these SFPE solutions are also viscosity solution of their corresponding semilinear PDEs. As already mentioned above, in the case of linear Kolmogorov PDEs, the well-known Feynman-Kac formula establishes a strong connection between stochastic analysis and PDEs by offering stochastic representations for the solutions of these PDEs. With the goal to achieve a similar connection between semilinear PDEs and SFPEs, we establish in Theorem 1.1.11 a Bismut-Elworthy-Li type formula that holds under the same assumption we made in Theorem 1.2.5. The combination of this specialized Bismut-Elworthy-Li formula, the findings in Chapter 1, and extended results on the existence and uniqueness of viscosity solutions then leads to our key result of this chapter which proves that - under certain assumptions on the coefficients - the unique viscosity solution of a PDE of the form

$$\begin{aligned}&(\tfrac{\partial u}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x u)(t,x), \rangle \\ &+ \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x u)(t,x)) + f(t,x,u(t,x),(\nabla_x u)(t,x)) = 0\end{aligned} \tag{2.1}$$

with $u(T,x) = g(x)$ for $(t,x) \in (0,T) \times \mathbb{R}^d$ is exactly the unique solution of the SFPE in (1.86). Hence, the findings in this chapter generalize the results in [10] to the case of semilinear PDEs with gradient-dependent nonlinearities and therefore expand the class of PDEs for which we can provide a stochastic representation of the viscosity solution. Furthermore, this chapter's results are going to justify the construction of the MLP approximation scheme in Chapter 3 to gain numerical approximations of viscosity solutions of semilinear PDEs with gradient-dependent nonlinearities.

This chapter is structured as follows. In Section 2.1 we study existence and uniqueness properties of a certain class of Kolmogorov PDEs. For this, we recall in Subsection 2.1.1 several definitions which we use in Section 2.1. In Subsections 2.1.2 - 2.1.3 we establish results on existence and on uniqueness of viscosity solutions of semilinear PDEs, respectively. Section 2.2 provides a Bismut-Elworthy-Li formula that is fitted to our setting and that we need to derive the connection between SFPEs and PDEs. The main result of Section 2.2 is Theorem 2.2.3 which presents said Bismut-Elworty-Li type formula. In Section 2.3 we combine the results of Section 2.1 and Theorem 2.2.3 with the results from Chapter 1 to prove that the solutions of certain SFPEs, established in Theorem 1.1.11, are indeed viscosity solution of the connected PDE. The main result of Section 2.3 is Theorem 2.3.1 below. We illustrate the findings of Theorem 2.3.1 by presenting Corollary 2.3.3.

2.1 Existence and uniqueness results for viscosity solutions (VS) of Kolmogorov PDEs

In Subsection 2.1.1 we recall the definitions of elliptic functions, viscosity solutions, and parabolic superjets. These definitions are used in Subsections 2.1.2 and 2.1.3 to prove existence and uniqueness properties for a certain class of viscosity solutions, respectively.

2.1.1 Definitions

The concept of viscosity solutions as generalization of classical solutions to PDEs was introduced in [26] and further developed in [25]. The following definitions are from [10, Definitions 2.4-2.7] and [10, Definitions 2.11 and 2.13].

Definition 2.1.1 (Degenerate elliptic functions)**.** Let $d \in \mathbb{N}$, $T \in (0, \infty)$, let $O \subseteq \mathbb{R}^d$ be a non-empty open set, and let $\langle \cdot, \cdot \rangle \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the standard Euclidean scalar product on $\mathbb{R}^d$. Then G is degenerate elliptic on $(0,T) \times O \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d$ if and only if

(i) it holds that $G \colon (0,T) \times O \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ is a function and

(ii) it holds for all $t \in (0,T)$, $x \in O$, $r \in \mathbb{R}$, $p \in \mathbb{R}^d$, $A, B \in \mathbb{S}_d$ with $\forall y \in \mathbb{R}^d \colon \langle Ay, y \rangle \leqslant \langle By, y \rangle$ that $G(t,x,r,p,A) \leqslant G(t,x,r,p,B)$.

Definition 2.1.2 (Viscosity subsolution)**.** Let $d \in \mathbb{N}$, $T \in (0, \infty)$, let $O \subseteq \mathbb{R}^d$ be a non-empty open set, and let $G \colon (0,T) \times O \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ be degenerate elliptic (cf. Definition 2.1.1). Then we say that u is a viscosity solution of $(\frac{\partial}{\partial t} u)(t,x) + G(t,x,u(t,x),(\nabla_x u)(t,x),(\operatorname{Hess}_x u)(t,x)) \geqslant 0$ for $(t,x) \in (0,T) \times O$ (we say that u is a viscosity subsolution of $(\frac{\partial}{\partial t} u)(t,x) + G(t,x,u(t,x),(\nabla_x u)(t,x), (\operatorname{Hess}_x u)(t,x)) = 0$) if and only if there exists a set $A \subseteq \mathbb{R} \times \mathbb{R}^d$ such that

(i) it holds that $(0,T) \times O \subseteq A$,

(ii) it holds that $u \colon A \to \mathbb{R}$ is upper semi-continuous, and

(iii) for all $t \in (0,T)$, $x \in O$, $\phi \in C^{1,2}((0,T) \times O, \mathbb{R})$ with $\phi(t,x) = u(t,x)$ and $\phi \geqslant u$ it holds that

$$(\tfrac{\partial}{\partial t} \phi)(t,x) + G(t,x,\phi(t,x),(\nabla_x \phi)(t,x),(\operatorname{Hess}_x \phi)(t,x)) \geqslant 0. \tag{2.2}$$

Definition 2.1.3 (Viscosity supersolution)**.** Let $d \in \mathbb{N}$, $T \in (0, \infty)$, let $O \subseteq \mathbb{R}^d$ be a non-empty open set, and let $G \colon (0,T) \times O \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ be degenerate elliptic (cf. Definition 2.1.1). Then we say u is a viscosity solution of $(\frac{\partial}{\partial t} u)(t,x) + G(t,x,u(t,x),(\nabla_x u)(t,x),(\operatorname{Hess}_x u)(t,x)) \leqslant 0$ for $(t,x) \in (0,T) \times O$ (we say that u is a viscosity supersolution of $(\frac{\partial}{\partial t} u)(t,x) + G(t,x,u(t,x),(\nabla_x u)(t,x), (\operatorname{Hess}_x u)(t,x)) = 0$) if and only if there exists a set $A \subseteq \mathbb{R} \times \mathbb{R}^d$ such that

(i) it holds that $(0,T) \times O \subseteq A$,

(ii) it holds that $u \colon A \to \mathbb{R}$ is lower semi-continuous, and

(iii) for all $t \in (0,T)$, $x \in O$, $\phi \in C^{1,2}((0,T) \times O, \mathbb{R})$ with $\phi(t,x) = u(t,x)$ and $\phi \leqslant u$ it holds that

$$(\tfrac{\partial}{\partial t} \phi)(t,x) + G(t,x,\phi(t,x),(\nabla_x \phi)(t,x),(\operatorname{Hess}_x \phi)(t,x)) \leqslant 0. \tag{2.3}$$

Definition 2.1.4 (Viscosity solution)**.** Let $d \in \mathbb{N}$, $T \in (0,\infty)$, let $O \subseteq \mathbb{R}^d$ be a non-empty open set, and let $G\colon (0,T) \times O \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ be degenerate elliptic (cf. Definition 2.1.1). Then we say that u is a viscosity solution of $(\frac{\partial}{\partial t}u)(t,x)+G(t,x,u(t,x),(\nabla_x u)(t,x),(\operatorname{Hess}_x u)(t,x)) = 0$ for $(t,x) \in (0,T) \times O$ if and only if

(i) it holds that u is a viscosity subsolution of

$$(\tfrac{\partial}{\partial t}u)(t,x) + G(t,x,u(t,x),(\nabla_x u)(t,x),(\operatorname{Hess}_x u)(t,x)) = 0 \tag{2.4}$$

for $(t,x) \in (0,T) \times O$ and

(ii) it holds that u is a viscosity supersolution of

$$(\tfrac{\partial}{\partial t}u)(t,x) + G(t,x,u(t,x),(\nabla_x u)(t,x),(\operatorname{Hess}_x u)(t,x)) = 0 \tag{2.5}$$

for $(t,x) \in (0,T) \times O$

(cf. Definitions 2.1.2 and 2.1.3).

Definition 2.1.5 (Parabolic superjets)**.** Let $d \in \mathbb{N}$, $T \in (0,\infty)$, let $O \subseteq \mathbb{R}^d$ be a non-empty open set, let $t \in (0,T)$, $x \in O$, let $\langle\cdot,\cdot\rangle\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the standard Euclidean scalar product on $\mathbb{R}^d$, let $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, and let $u\colon (0,T) \times O \to \mathbb{R}$ be a function. Then we denote by $(\mathcal{P}^+u)(t,x)$ the set satisfying

$$\begin{aligned}(\mathcal{P}^+u)(t,x) = \Big\{&(b,p,A) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d\colon \\ &\limsup_{[(0,T)\times O]\setminus\{(t,x)\}\ni(s,y)\to(t,x)} \left[\tfrac{u(s,y)-u(t,x)-b(s-t)-\langle p,y-x\rangle-\frac{1}{2}\langle A(y-x),y-x\rangle}{|t-s|+\|x-y\|^2}\right] \leq 0\Big\}.\end{aligned} \tag{2.6}$$

Definition 2.1.6 (Generalized parabolic superjets)**.** Let $d \in \mathbb{N}$, $T \in (0,\infty)$, let $O \subseteq \mathbb{R}^d$ be a non-empty open set, let $t \in (0,T)$, $x \in O$, let $\langle\cdot,\cdot\rangle\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the standard Euclidean scalar product on $\mathbb{R}^d$, let $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, and let $u\colon (0,T) \times O \to \mathbb{R}$ be a function. Then we denote by $(\mathfrak{P}^+u)(t,x)$ the set satisfying

$$\begin{aligned}(\mathfrak{P}^+u)(t,x) = \Big\{&(b,p,A) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d\colon \\ &\Big(\exists (t_n,x_n,b_n,p_n,A_n)_{n\in\mathbb{N}} \subseteq (0,T) \times O \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d\colon \\ &\big(\forall n \in \mathbb{N}\colon (b_n,p_n,A_n) \in (\mathcal{P}^+u)(t_n,x_n)\big) \text{ and} \\ &\lim_{n\to\infty}(t_n,x_n,u(t_n,x_n),b_n,p_n,A_n) = (t,x,u(t,x),b,p,A)\Big)\Big\}\end{aligned} \tag{2.7}$$

(cf. Definition 2.1.5).

2.1.2 Existence result for viscosity solutions of linear inhomogeneous Kolmogorov PDEs

The following proposition establishes a Feynman-Kac representation for viscosity solutions of certain PDEs. Proposition 2.1.7 is a minor generalization of [10, Proposition 2.23] where we replace $[0,T]$ by $[0,T]\backslash K_r$.

Proposition 2.1.7. *Let $d,m \in \mathbb{N}$, $T \in (0,\infty)$, let $O \subseteq \mathbb{R}^d$ be a non-empty open set, let $\langle\cdot,\cdot\rangle\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the standard Euclidean scalar product on $\mathbb{R}^d$, let $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, let $\|\cdot\|_F\colon \mathbb{R}^{d\times m} \to [0,\infty)$ be the Frobenius norm* *on $\mathbb{R}^{d\times m}$, for every $r\in(0,\infty)$ let $K_r\subseteq[0,T)$, $O_r\subseteq O$ satisfy $K_r=[0,\max\{T-\frac{1}{r},0\}]$ and $O_r=\{x\in O\colon \|x\|\leqslant r$ and $\{y\in\mathbb{R}^d\colon \|y-x\|<\frac{1}{r}\}\subseteq O\}$, let $g\in C(O,\mathbb{R})$, $h\in C([0,T]\times O,\mathbb{R})$, $\mu\in C([0,T]\times O,\mathbb{R}^d)$, $\sigma\in C([0,T]\times O,\mathbb{R}^{d\times m})$, $V\in C^{1,2}([0,T]\times O,(0,\infty))$ satisfy for all $r\in(0,\infty)$ that*

$$\sup\left(\left\{\tfrac{\|\mu(t,x)-\mu(t,y)\|+\|\sigma(t,x)-\sigma(t,y)\|_F}{\|x-y\|}\colon t\in[0,T],x,y\in O_r,x\neq y\right\}\cup\{0\}\right)<\infty, \tag{2.8}$$

assume for all $t\in[0,T]$, $x\in O$ that

$$(\tfrac{\partial V}{\partial t})(t,x)+\langle\mu(t,x),(\nabla_x V)(t,x)\rangle+\tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x V)(t,x))\leqslant 0, \tag{2.9}$$

assume that $\sup_{r\in(0,\infty)}[\inf_{t\in[0,T]}\inf_{x\in O\setminus O_r}V(t,x)]=\infty$ and $\inf_{r\in(0,\infty)}[\sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}(\frac{|g(x)|}{V(T,x)}+\frac{|h(t,x)|}{V(t,x)}\sqrt{T-t})]=0$, let $(\Omega,\mathcal{F},\mathbb{P},(\mathbb{F}_t)_{t\in[0,T]})$ be a filtered probability space, let $W\colon[0,T]\times\Omega\to\mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion, for every $t\in[0,T]$, $x\in O$ let $X_t^x=(X_{t,s}^x)_{s\in[t,T]}\colon[t,T]\times\Omega\to O$ be an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s\in[t,T]$ it holds a.s. that

$$X_{t,s}^x=x+\int_t^s\mu(r,X_{t,r}^x)\,dr+\int_t^s\sigma(r,X_{t,r}^x)\,dW_r, \tag{2.10}$$

and let $u\colon[0,T]\times\mathbb{R}^d\to\mathbb{R}$ satisfy for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$u(t,x)=\mathbb{E}\Big[g(X_{t,T}^x)+\int_t^T h(s,X_{t,s}^x)\,ds\Big]. \tag{2.11}$$

Then u is a viscosity solution of

$$\begin{aligned}&(\tfrac{\partial u}{\partial t})(t,x)+\langle\mu(t,x),(\nabla_x u)(t,x)\rangle\\&\quad+\tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x u)(t,x))+h(t,x)=0\end{aligned} \tag{2.12}$$

with $u(T,x)=g(x)$ for $(t,x)\in(0,T)\times O$.

Proof of Proposition 2.1.7. Throughout this proof let $\mathfrak{g}_n\in C(\mathbb{R}^d,\mathbb{R})$, $n\in\mathbb{N}$, and $\mathfrak{h}_n\in C([0,T]\times\mathbb{R}^d,\mathbb{R})$, $n\in\mathbb{N}$, be compactly supported functions which satisfy that $[\bigcup_{n\in\mathbb{N}}\operatorname{supp}(\mathfrak{g}_n)]\subseteq O$, $[\bigcup_{n\in\mathbb{N}}\operatorname{supp}(\mathfrak{h}_n)]\subseteq[0,T]\times O$, and

$$\limsup_{n\to\infty}\left[\sup_{t\in[0,T)}\sup_{x\in O}\left(\frac{|\mathfrak{g}_n(x)-g(x)|}{V(T,x)}+\frac{|\mathfrak{h}_n(t,x)-h(t,x)|}{V(t,x)}\sqrt{T-t}\right)\right]=0 \tag{2.13}$$

(cf. Corollary 1.1.6), let $\mathfrak{m}_n\in C([0,T]\times\mathbb{R}^d,\mathbb{R}^d)$, $n\in\mathbb{N}$, and $\mathfrak{s}_n\in C([0,T]\times\mathbb{R}^d,\mathbb{R}^{d\times m})$, $n\in\mathbb{N}$, be compactly supported functions which satisfy that

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

$$\sup_{t\in[0,T]}\sup_{\substack{x,y\in\mathbb{R}^d\\x\neq y}}\left[\frac{\|\mathfrak{m}_n(t,x)-\mathfrak{m}_n(t,y)\|+\|\mathfrak{s}_n(t,x)-\mathfrak{s}_n(t,y)\|_F}{\|x-y\|}\right]<\infty, \tag{2.14}$$

(II) for all $n\in\mathbb{N}$, $t\in[0,T]$, $x\in O$ with $V(t,x)\leqslant n$ it holds that

$$[\|\mathfrak{m}_n(t,x)-\mu(t,x)\|+\|\mathfrak{s}_n(t,x)-\sigma(t,x)\|_F]=0, \tag{2.15}$$

and

(III) for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ with $V(t,x) \geqslant n+1$ it holds that $\|\mathfrak{m}_n(t,x)\| + \|\mathfrak{s}_n(t,x)\|_F = 0$

(cf., e.g., the proof of [13, Lemma 3.7]), for every $n \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ let $\mathfrak{X}^{x,n}_t = (\mathfrak{X}^{x,n}_{t,s})_{s\in[t,T]} \colon [t,T] \times \Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in [t,T]$ it holds a.s. that

$$\mathfrak{X}^{x,n}_{t,s} = x + \int_t^s \mathfrak{m}_n(r, \mathfrak{X}^{x,n}_{t,r})\,\mathrm{d}r + \int_t^s \mathfrak{s}_n(r, \mathfrak{X}^{x,n}_{t,r})\,\mathrm{d}W_r \tag{2.16}$$

(cf., e.g., [69, Theorem 5.2.9]), let $\mathfrak{u}^{n,k} \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $n \in \mathbb{N}_0$, $k \in \mathbb{N}$, satisfy for all $n,k \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\mathfrak{u}^{n,k}(t,x) = \mathbb{E}\biggl[\mathfrak{g}_k(\mathfrak{X}^{x,n}_{t,T}) + \int_t^T \mathfrak{h}_k(s, \mathfrak{X}^{x,n}_{t,s})\,\mathrm{d}s\biggr] \tag{2.17}$$

and

$$\mathfrak{u}^{0,k}(t,x) = \mathbb{E}\biggl[\mathfrak{g}_k(X^x_{t,T}) + \int_t^T \mathfrak{h}_k(s, X^x_{t,s})\,\mathrm{d}s\biggr], \tag{2.18}$$

and for every $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O$ let $\tau^{x,n}_t \colon \Omega \to [t,T]$ satisfy

$$\tau^{x,n}_t = \inf(\{s \in [t,T] \colon \max\{V(s,\mathfrak{X}^{x,n}_{t,s}), V(s,X^x_{t,s})\} \geqslant n\} \cup \{T\}). \tag{2.19}$$

Next note that [10, Lemma 2.22] (applied for every $n,k \in \mathbb{N}$ with $\mu \curvearrowleft \mathfrak{m}_n$, $\sigma \curvearrowleft \mathfrak{s}_n$, $g \curvearrowleft \mathfrak{g}_k$, $h \curvearrowleft \mathfrak{h}_k$ in the notation of [10, Lemma 2.22]), item (I), and the fact that for all $n \in \mathbb{N}$ it holds that $\mathfrak{m}_n$ and $\mathfrak{s}_n$ have compact support demonstrate that for all $n,k \in \mathbb{N}$ it holds that $\mathfrak{u}^{n,k}$ is a viscosity solution of

$$\begin{aligned}(\tfrac{\partial}{\partial t}\mathfrak{u}^{n,k})(t,x) + \langle \mathfrak{m}_n(t,x), (\nabla_x \mathfrak{u}^{n,k})(t,x)\rangle \\ + \tfrac{1}{2}\operatorname{Tr}(\mathfrak{s}_n(t,x)[\mathfrak{s}_n(t,x)]^*(\operatorname{Hess}_x \mathfrak{u}^{n,k})(t,x)) + \mathfrak{h}_k(t,x) = 0\end{aligned} \tag{2.20}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$. Furthermore, note that item (II), (2.8), (2.10), (2.16) and pathwise uniqueness ensure that for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O$ it holds that

$$\mathbb{P}(\forall s \in [t,T] \colon \mathbb{1}_{\{s \leqslant \tau^{x,n}_t\}} \mathfrak{X}^{x,n}_{t,s} = \mathbb{1}_{\{s \leqslant \tau^{x,n}_t\}} X^x_{t,s}) = 1. \tag{2.21}$$

Hence, we obtain for all $n,k \in \mathbb{N}$, $t \in [0,T]$, $x \in O$ that

$$\begin{aligned}\mathbb{E}\Bigl[|\mathfrak{g}_k(\mathfrak{X}^{x,n}_{t,T}) - \mathfrak{g}_k(X^x_{t,T})|\Bigr] &= \mathbb{E}\Bigl[\mathbb{1}_{\{\tau^{x,n}_t < T\}}|\mathfrak{g}_k(\mathfrak{X}^{x,n}_{t,T}) - \mathfrak{g}_k(X^x_{t,T})|\Bigr] \\ &\leqslant 2\Bigl[\sup\nolimits_{y\in O}|\mathfrak{g}_k(y)|\Bigr]\mathbb{P}(\tau^{x,n}_t < T)\end{aligned} \tag{2.22}$$

and

$$\begin{aligned}&\int_t^T \mathbb{E}\Bigl[|\mathfrak{h}_k(s,\mathfrak{X}^{x,n}_{t,s}) - \mathfrak{h}_k(s,X^x_{t,s})|\Bigr]\,\mathrm{d}s \\ &= \int_t^T \mathbb{E}\Bigl[\mathbb{1}_{\{\tau^{x,n}_t < T\}}|\mathfrak{h}_k(s,\mathfrak{X}^{x,n}_{t,s}) - \mathfrak{h}_k(s,X^x_{t,s})|\Bigr]\,\mathrm{d}s \\ &\leqslant 2T\Bigl[\sup\nolimits_{s\in[0,T]}\sup\nolimits_{y\in O}|\mathfrak{h}_k(s,y)|\Bigr]\mathbb{P}(\tau^{x,n}_t < T).\end{aligned} \tag{2.23}$$

In addition, observe that [13, Lemma 3.1] and (2.9) ensure that for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O$ it holds that

$$\mathbb{E}[V(\tau^{x,n}_t, X^x_{t,\tau^{x,n}_t})] \leqslant V(t,x). \tag{2.24}$$

Markov's inequality, (2.22), and (2.23) therefore imply that for all $n, k \in \mathbb{N}$, $t \in [0,T]$, $x \in O$ it holds that

$$\begin{aligned}
|\mathfrak{u}^{n,k}(t,x) - \mathfrak{u}^{0,k}(t,x)| &\leqslant 2\bigg[\sup_{y\in O}|\mathfrak{g}_k(y)| + T \sup_{s\in[0,T]}\sup_{y\in O}|\mathfrak{h}_k(s,y)|\bigg]\mathbb{P}(\tau_t^{x,n} < T) \\
&\leqslant 2\bigg[\sup_{y\in O}|\mathfrak{g}_k(y)| + T \sup_{s\in[0,T]}\sup_{y\in O}|\mathfrak{h}_k(s,y)|\bigg]\mathbb{P}(V(\tau_t^{x,n}, X^x_{t,\tau_t^{x,n}}) \geqslant n) \\
&\leqslant \frac{2}{n}\bigg[\sup_{y\in O}|\mathfrak{g}_k(y)| + T \sup_{s\in[0,T]}\sup_{y\in O}|\mathfrak{h}_k(s,y)|\bigg]\mathbb{E}[V(\tau_t^{x,n}, X^x_{t,\tau_t^{x,n}})] \\
&\leqslant \frac{2}{n}\bigg[\sup_{y\in O}|\mathfrak{g}_k(y)| + T \sup_{s\in[0,T]}\sup_{y\in O}|\mathfrak{h}_k(s,y)|\bigg]V(t,x).
\end{aligned} \tag{2.25}$$

This shows that for all $k \in \mathbb{N}$ and all non-empty compact $\mathcal{K} \subseteq (0,T) \times O$ it holds that

$$\limsup_{n\to\infty}\bigg[\sup_{(t,x)\in\mathcal{K}}|\mathfrak{u}^{n,k}(t,x) - \mathfrak{u}^{0,k}(t,x)|\bigg] = 0. \tag{2.26}$$

Moreover, observe that item (II) and the assumption that $\sup_{r\in(0,\infty)}[\inf_{t\in[0,T)\setminus K_r}\inf_{x\in O\setminus O_r} V(t,x)] = \infty$ imply that for all non-empty compact $\mathcal{K} \subseteq [0,T] \times O$ it holds that

$$\limsup_{n\to\infty}\bigg[\sup_{(t,x)\in\mathcal{K}}\Big(\|\mathfrak{m}_n(t,x) - \mu(t,x)\| + \|\mathfrak{s}_n(t,x) - \sigma(t,x)\|\Big)\bigg] = 0. \tag{2.27}$$

Combining [10, Corollary 2.20], (2.20), and (2.26) hence demonstrates that for all $k \in \mathbb{N}$ it holds that $\mathfrak{u}^{0,k}$ is a viscosity solution of

$$\begin{aligned}
&(\tfrac{\partial}{\partial t}\mathfrak{u}^{0,k})(t,x) + \langle \mu(t,x), (\nabla_x \mathfrak{u}^{0,k})(t,x)\rangle \\
&\quad + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x \mathfrak{u}^{0,k})(t,x)) + \mathfrak{h}_k(t,x) = 0
\end{aligned} \tag{2.28}$$

for $(t,x) \in (0,T) \times O$. Next observe that [13, Lemma 3.1] and (2.9) ensure that for all $t \in [0,T]$, $s \in [t,T]$, $x \in O$ it holds that $\mathbb{E}[V(s, X^x_{t,s})] \leqslant V(t,x)$. This demonstrates that for all $k \in \mathbb{N}$, $t \in (0,T)$, $x \in O$ it holds that

$$
\begin{aligned}
&|\mathfrak{u}^{0,k}(t,x) - u(t,x)| \\
&= \left| \mathbb{E}[\mathfrak{g}_k(X^x_{t,T}) - g(X^x_{t,T})] + \int_t^T \mathbb{E}[\mathfrak{h}_k(s,X^x_{t,s}) - h(s,X^x_{t,s})] \, ds \right| \\
&\leqslant \mathbb{E}\left[\frac{|\mathfrak{g}_k(X^x_{t,T}) - g(X^x_{t,T})| V(T,X^x_{t,T})}{V(T,X^x_{t,T})} \right] \\
&\quad + \int_t^T \mathbb{E}\left[\frac{|\mathfrak{h}_k(s,X^x_{t,s}) - h(s,X^x_{t,s})| V(s,X^x_{t,s}) \sqrt{T-s}}{V(s,X^x_{t,s})\sqrt{T-s}} \right] ds \\
&\leqslant \left[\sup_{y\in O} \frac{|\mathfrak{g}_k(y) - g(y)|}{V(T,y)} \right] \mathbb{E}\big[V(T,X^x_{t,T})\big] \\
&\quad + \left[\sup_{r\in[0,T)} \sup_{y\in O} \frac{|\mathfrak{h}_k(r,y) - h(r,y)|}{V(r,y)} \sqrt{T-r} \right] \int_t^T \mathbb{E}\left[\frac{V(s,X^x_{t,s})}{\sqrt{T-s}} \right] ds \\
&\leqslant \left[\sup_{y\in O} \frac{|\mathfrak{g}_k(y) - g(y)|}{V(T,y)} \right] V(T,x) \\
&\quad + \left[\sup_{r\in[0,T)} \sup_{y\in O} \frac{|\mathfrak{h}_k(r,y) - h(r,y)|}{V(r,y)} \sqrt{T-r} \right] \int_t^T \frac{V(t,x)}{\sqrt{T-s}} \, ds \\
&\leqslant \left[\sup_{y\in O} \frac{|\mathfrak{g}_k(y) - g(y)|}{V(T,y)} \right] V(T,x) \\
&\quad + \left[\sup_{r\in[0,T)} \sup_{y\in O} \frac{|\mathfrak{h}_k(r,y) - h(r,y)|}{V(r,y)} \sqrt{T-r} \right] 2\sqrt{T} \, V(t,x).
\end{aligned}
\tag{2.29}
$$

Combining this with (2.13) shows that for all non-empty compact $\mathcal{K} \subseteq (0,T) \times O$ it holds that

$$
\limsup_{k\to\infty} \left[\sup_{(t,x)\in\mathcal{K}} |\mathfrak{u}^{0,k}(t,x) - u(t,x)| \right] = 0. \tag{2.30}
$$

This, [10, Corollary 2.20], (2.13), and (2.28) imply that u is a viscosity solution of

$$
\begin{aligned}
&(\tfrac{\partial u}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x u)(t,x) \rangle \\
&\quad + \tfrac{1}{2} \operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^* (\operatorname{Hess}_x u)(t,x)) + h(t,x) = 0
\end{aligned}
\tag{2.31}
$$

for $(t,x) \in (0,T) \times O$. In addition, observe that (2.11) ensures that for all $x \in \mathbb{R}^d$ it holds that $u(T,x) = g(x)$. This and (2.31) establish (2.12). The proof of Proposition 2.1.7 is thus complete. □

2.1.3 Uniqueness results for viscosity solutions of semilinear Kolmogorov PDEs

The following proposition proves that certain semilinear PDEs with Lipschitz continuous, gradient-dependent nonlinearities admit at most one viscosity solution satisfying a specific growth condition. Proposition 2.1.8 generalizes [10, Proposition 3.5] to the case of semilinear PDEs with gradient-dependent nonlinearities.

Proposition 2.1.8. *Let $d, m \in \mathbb{N}$, $L_1, L_2, T \in (0,\infty)$, let $\langle \cdot,\cdot \rangle \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the standard Euclidean scalar product on $\mathbb{R}^d$, let $\|\cdot\| \colon \mathbb{R}^d \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, let $|||\cdot||| \colon \mathbb{R}^{d+1} \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^{d+1}$, let*

$\|\cdot\|_F\colon (\bigcup_{a,b=1}^\infty \mathbb{R}^{a\times b}) \to [0,\infty)$ *satisfy for all* $a,b\in\mathbb{N}$, $A=(A_{ij})_{(i,j)\in\{1,2,\dots,a\}\times\{1,2,\dots,b\}}\in\mathbb{R}^{a\times b}$ *that* $\|A\|_F = [\sum_{i=1}^a\sum_{j=1}^b |A_{ij}|^2]^{\frac{1}{2}}$, *let* $O\subseteq\mathbb{R}^d$ *be a non-empty open set, for every* $r\in(0,\infty)$ *let* $K_r\subseteq[0,T)$, $O_r\subseteq O$ *satisfy* $K_r=[0,\max\{T-\frac{1}{r},0\}]$ *and* $O_r=\{x\in O\colon \|x\|\leqslant r \text{ and } \{y\in\mathbb{R}^d\colon \|y-x\|<\frac{1}{r}\}\subseteq O\}$, *let* $g\in C(O,\mathbb{R})$, $f\in C([0,T]\times O\times\mathbb{R}\times\mathbb{R}^d,\mathbb{R})$, $\mu\in C([0,T]\times O,\mathbb{R}^d)$, $\sigma\in C([0,T]\times O,\mathbb{R}^{d\times m})$, $V\in C^{1,2}([0,T]\times O,[1,\infty))$ *satisfy for all* $r\in(0,\infty)$ *that*

$$\sup\left(\left\{\tfrac{\|\mu(t,x)-\mu(t,y)\|+\|\sigma(t,x)-\sigma(t,y)\|_F}{\|x-y\|}\colon t\in[0,T], x,y\in O_r, x\neq y\right\}\cup\{0\}\right)<\infty, \tag{2.32}$$

for all $t\in[0,T]$, $x\in O$, $r,s\in\mathbb{R}$, $v,w\in\mathbb{R}^d$ *that* $(f(t,x,r,v)-f(t,x,s,v))(r-s)\leqslant L_1|r-s|^2$ *and* $|f(t,x,r,v)-f(t,x,r,w)|\leqslant L_2\|v-w\|$, *and*

$$\begin{aligned}&(\tfrac{\partial V}{\partial t})(t,x)+\langle\mu(t,x),(\nabla_x V)(t,x)\rangle+\tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x V)(t,x))\\&\quad+L_2\|(\nabla_x V)(t,x)\|\leqslant 0,\end{aligned}\tag{2.33}$$

and let $u_1,u_2\in C([0,T]\times O,\mathbb{R})$ *satisfy for all* $i\in\{1,2\}$ *that* $\limsup_{r\to\infty}[\sup_{t\in[0,T]}\sup_{x\in O\setminus O_r}\frac{|u_i(t,x)|}{V(t,x)}]=0$ *and that* u_i *is a viscosity solution of*

$$\begin{aligned}&(\tfrac{\partial u_i}{\partial t})(t,x)+\langle\mu(t,x),(\nabla_x u_i)(t,x)\rangle+\tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x u_i)(t,x))\\&\quad+f(t,x,u_i(t,x),(\nabla_x u_i)(t,x))=0\end{aligned}\tag{2.34}$$

with $u_i(T,x)=g(x)$ *for* $(t,x)\in(0,T)\times O$. *Then it holds for all* $t\in[0,T]$, $x\in O$ *that* $u_1(t,x)=u_2(t,x)$.

Proof of Proposition 2.1.8. Throughout this proof let $\mathbb{V}\colon[0,T]\times O\to(0,\infty)$ satisfy for all $t\in[0,T]$, $x\in O$ that $\mathbb{V}(t,x)=e^{-L_1 t}V(t,x)$, let $v_i\colon[0,T]\times O\to\mathbb{R}$, $i\in\{1,2\}$, satisfy for all $i\in\{1,2\}$, $t\in[0,T]$, $x\in O$ that $v_i(t,x)=\frac{u_i(t,x)}{\mathbb{V}(t,x)}$, let $G\colon(0,T)\times O\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{S}_d\to\mathbb{R}$ satisfy for all $t\in(0,T)$, $x\in O$, $r\in\mathbb{R}$, $p\in\mathbb{R}^d$, $A\in\mathbb{S}_d$ that

$$G(t,x,r,p,A)=\langle\mu(t,x),p\rangle+\tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*A)+f(t,x,r,p),\tag{2.35}$$

and let $H\colon(0,T)\times O\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{S}_d\to\mathbb{R}$ satisfy for all $t\in(0,T)$, $x\in O$, $r\in\mathbb{R}$, $p\in\mathbb{R}^d$, $A\in\mathbb{S}_d$ that

$$\begin{aligned}H(t,x,r,p,A)&=\tfrac{r}{\mathbb{V}(t,x)}(\tfrac{\partial}{\partial t}\mathbb{V})(t,x)\\&\quad+\tfrac{1}{\mathbb{V}(t,x)}G\Big(t,x,r\mathbb{V}(t,x),\mathbb{V}(t,x)p+r(\nabla_x\mathbb{V})(t,x),\\&\qquad\mathbb{V}(t,x)A+p[(\nabla_x\mathbb{V})(t,x)]^*+(\nabla_x\mathbb{V})(t,x)p^*+r(\operatorname{Hess}_x\mathbb{V})(t,x)\Big).\end{aligned}\tag{2.36}$$

Observe that (2.33) and the assumption that $V\in C^{1,2}([0,T]\times O,(0,\infty))$ ensure that for all $t\in[0,T]$, $x\in O$ it holds that $\mathbb{V}\in C^{1,2}([0,T]\times O,(0,\infty))$ and

$$\begin{aligned}&(\tfrac{\partial}{\partial t}\mathbb{V})(t,x)+\langle\mu(t,x),(\nabla_x\mathbb{V})(t,x)\rangle+\tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x\mathbb{V})(t,x))\\&\quad+L_1\mathbb{V}(t,x)+L_2\|(\nabla_x\mathbb{V})(t,x)\|\leqslant 0.\end{aligned}\tag{2.37}$$

Next note that (2.35) implies that $G\in C((0,T)\times O\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{S}_d,\mathbb{R})$ is degenerate elliptic. Combining this with (2.36) shows that $H\in C((0,T)\times O\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{S}_d,\mathbb{R})$ is degenerate elliptic. In the next step observe that the assumption that for all $i\in\{1,2\}$, $x\in O$ it holds that $u_i(T,x)=g(x)$ implies that for all $x\in O$ it holds that

$$v_1(T,x)\leqslant v_2(T,x)\leqslant v_1(T,x).\tag{2.38}$$

Furthermore, note that the hypothesis that $\limsup_{r\to\infty}[\sup_{t\in[0,T]}\sup_{x\in O\backslash O_r}(\frac{|u_1(t,x)|+|u_2(t,x)|}{V(t,x)})] = 0$ shows that

$$\limsup_{r\to\infty}\left[\sup_{t\in[0,T]}\sup_{x\in O\backslash O_r}\Big(|v_1(t,x)-v_2(t,x)|\Big)\right] = 0. \tag{2.39}$$

In addition, observe that [45, Lemma 4.12] (applied for every $i \in \{1,2\}$ with $\tilde{G} \curvearrowleft H$, $V \curvearrowleft \mathbb{V}$, $\tilde{u} \curvearrowleft (v_i(T-t,x))_{t\in[0,T],x\in O}$ in the notation of [45, Lemma 4.12]), (2.34), and (2.36) demonstrate that for all $i \in \{1,2\}$ it holds that v_i is a viscosity solution of

$$(\tfrac{\partial}{\partial t}v_i)(t,x) + H(t,x,v_i(t,x),(\nabla_x v_i)(t,x),(\operatorname{Hess}_x v_i)(t,x)) = 0 \tag{2.40}$$

for $(t,x) \in (0,T)\times O$. Throughout the rest of the proof let $\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_m \in \mathbb{R}^m$ satisfy $\mathbf{e}_1 = (1,0,\ldots,0)$, $\mathbf{e}_2 = (0,1,0,\ldots,0)$, $\ldots$, $\mathbf{e}_m = (0,\ldots,0,1)$, and let $(t_n,x_n,r_n,A_n) \in (0,T)\times O\times\mathbb{R}\times\mathbb{S}_d$, $n\in\mathbb{N}_0$, and $(\mathfrak{x}_n,\mathfrak{r}_n,\mathfrak{A}_n) \in O\times\mathbb{R}\times\mathbb{S}_d$, $n\in\mathbb{N}_0$, satisfy $\limsup_{n\to\infty}[|t_n - t_0| + \|x_n - x_0\| + \sqrt{n}\|x_n-\mathfrak{x}_n\|] = 0 < r_0 = \liminf_{n\to\infty}|r_n-\mathfrak{r}_n| = \limsup_{n\to\infty}|r_n-\mathfrak{r}_n| \leqslant \sup_{n\in\mathbb{N}}(|r_n|+|\mathfrak{r}_n|) < \infty$ and for all $n\in\mathbb{N}$, $y,z\in\mathbb{R}^d$ that $\langle y, A_n y\rangle - \langle z,\mathfrak{A}_n z\rangle \leqslant 5\|y-z\|^2$. Note that (2.32) and the fact that $\limsup_{n\to\infty}[\sqrt{n}\|x_n-\mathfrak{x}_n\|] = 0$ ensure that

$$\limsup_{n\to\infty}\left[n\|\sigma(t_n,x_n)-\sigma(t_n,\mathfrak{x}_n)\|_F^2\right] = 0. \tag{2.41}$$

The fact that for all $B\in\mathbb{S}_d$, $C\in\mathbb{R}^{d\times m}$ it holds that $\operatorname{Tr}(CC^*B) = \sum_{i=1}^m\langle C\mathbf{e}_i, BC\mathbf{e}_i\rangle$ and the assumption that for all $n\in\mathbb{N}$, $y,z\in\mathbb{R}^d$ it holds that $\langle y,A_n y\rangle - \langle z,\mathfrak{A}_n z\rangle \leqslant 5\|y-z\|^2$ therefore imply that

$$\begin{aligned}
&\limsup_{n\to\infty}\left[\tfrac{1}{2}\operatorname{Tr}\left(\tfrac{\sigma(t_n,x_n)[\sigma(t_n,x_n)]^*}{\mathbb{V}(t_n,x_n)}\mathbb{V}(t_n,x_n)nA_n - \tfrac{\sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*}{\mathbb{V}(t_n,\mathfrak{x}_n)}\mathbb{V}(t_n,\mathfrak{x}_n)n\mathfrak{A}_n\right)\right]\\
&= \limsup_{n\to\infty}\left[\tfrac{n}{2}\operatorname{Tr}\left(\sigma(t_n,x_n)[\sigma(t_n,x_n)]^*A_n - \sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*\mathfrak{A}_n\right)\right]\\
&= \limsup_{n\to\infty}\left[\tfrac{n}{2}\sum_{i=1}^m\Big(\langle\sigma(t_n,x_n)\mathbf{e}_i, A_n\sigma(t_n,x_n)\mathbf{e}_i\rangle - \langle\sigma(t_n,\mathfrak{x}_n)\mathbf{e}_i,\mathfrak{A}_n\sigma(t_n,\mathfrak{x}_n)\mathbf{e}_i\rangle\Big)\right]\\
&\leqslant \limsup_{n\to\infty}\left[\sum_{i=1}^m\tfrac{5n}{2}\|\sigma(t_n,x_n)\mathbf{e}_i - \sigma(t_n,\mathfrak{x}_n)\mathbf{e}_i\|^2\right]\\
&= \tfrac{5}{2}\limsup_{n\to\infty}\left[n\|\sigma(t_n,x_n)-\sigma(t_n,\mathfrak{x}_n)\|_F^2\right] = 0.
\end{aligned} \tag{2.42}$$

Furthermore, note that (2.32) and the fact that $\mathbb{V}\in C^{1,2}([0,T]\times O,(0,\infty))$ show that for all compact $\mathcal{K}\subseteq O$ there exists $c\in\mathbb{R}$ which satisfies for all $s\in[0,T]$, $y_1,y_2\in\mathcal{K}$ that

$$\begin{aligned}
&\left\|\tfrac{\sigma(s,y_1)[\sigma(s,y_1)]^*}{\mathbb{V}(s,y_1)} - \tfrac{\sigma(s,y_2)[\sigma(s,y_2)]^*}{\mathbb{V}(s,y_2)}\right\|_F + \|(\nabla_x\mathbb{V})(s,y_1) - (\nabla_x\mathbb{V})(s,y_2)\|\\
&\leqslant c\|y_1-y_2\|.
\end{aligned} \tag{2.43}$$

This, the fact that $\limsup_{n\to\infty}[|t_n-t_0|+\|x_n-x_0\|] = 0$ and $\limsup_{n\to\infty}[\sqrt{n}\|x_n-\mathfrak{x}_n\|] = 0$ demonstrate that

$$\limsup_{n\to\infty}\left[n\|x-\mathfrak{x}_n\|\left\|\frac{\sigma(t_n,x_n)[\sigma(t_n,x_n)]^*}{\mathbb{V}(t_n,x_n)} - \frac{\sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*}{\mathbb{V}(t_n,\mathfrak{x}_n)}\right\|_F\right] = 0 \tag{2.44}$$

and

$$\limsup_{n\to\infty}[n\|x_n-\mathfrak{x}_n\|\,\|(\nabla_x\mathbb{V})(t_n,x_n) - (\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n)\|] = 0. \tag{2.45}$$

In addition, note that for all $B \in \mathbb{S}_d$, $v, w \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\operatorname{Tr}(Bvw^*) = \operatorname{Tr}(w^*Bv) = w^*Bv = \langle w, Bv\rangle = \langle Bw, v\rangle \\ &= \langle v, Bw\rangle = v^*Bw = \operatorname{Tr}(v^*Bw) = \operatorname{Tr}(Bwv^*). \end{aligned} \tag{2.46}$$

This, the Cauchy-Schwarz inequality, (2.44), and (2.45) demonstrate that

$$\begin{aligned} &\limsup_{n\to\infty} \Bigg[\tfrac{1}{2} \operatorname{Tr}\Bigg(\tfrac{\sigma(t_n,x_n)[\sigma(t_n,x_n)]^*}{\mathbb{V}(t_n,x_n)} \Big(n(x_n - \mathfrak{x}_n)[(\nabla_x \mathbb{V})(t_n,x_n)]^* \\ &\quad + (\nabla_x \mathbb{V})(t_n,x_n) n(x_n - \mathfrak{x}_n)^* \Big) \\ &\quad - \tfrac{\sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*}{\mathbb{V}(t_n,\mathfrak{x}_n)} \Big(n(x_n - \mathfrak{x}_n)[(\nabla_x \mathbb{V})(t_n,\mathfrak{x}_n)]^* \\ &\quad + (\nabla_x \mathbb{V})(t_n,\mathfrak{x}_n) n(x_n - \mathfrak{x}_n)^* \Big) \Bigg) \Bigg] \\ &= \limsup_{n\to\infty} \Bigg[\Big\langle \tfrac{\sigma(t_n,x_n)[\sigma(t_n,x_n)]^*}{\mathbb{V}(t_n,x_n)} n(x_n - \mathfrak{x}_n), (\nabla_x \mathbb{V})(t_n,x_n) \Big\rangle \\ &\quad - \Big\langle \tfrac{\sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*}{\mathbb{V}(t_n,\mathfrak{x}_n)} n(x_n - \mathfrak{x}_n), (\nabla_x \mathbb{V})(t_n,\mathfrak{x}_n) \Big\rangle \Bigg] \\ &= \limsup_{n\to\infty} \Bigg[\Big\langle \Big(\tfrac{\sigma(t_n,x_n)[\sigma(t_n,x_n)]^*}{\mathbb{V}(t_n,x_n)} - \tfrac{\sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*}{\mathbb{V}(t_n,\mathfrak{x}_n)} \Big) n(x_n - \mathfrak{x}_n), (\nabla_x \mathbb{V})(t_n,x_n) \Big\rangle \\ &\quad + \Big\langle \tfrac{\sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*}{\mathbb{V}(t_n,\mathfrak{x}_n)} n(x_n - \mathfrak{x}_n), (\nabla_x \mathbb{V})(t_n,x_n) - (\nabla_x \mathbb{V})(t_n,\mathfrak{x}_n) \Big\rangle \Bigg] \\ &\leqslant \limsup_{n\to\infty} \Bigg[\Big\| \tfrac{\sigma(t_n,x_n)[\sigma(t_n,x_n)]^*}{\mathbb{V}(t_n,x_n)} - \tfrac{\sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*}{\mathbb{V}(t_n,\mathfrak{x}_n)} \Big\|_F n\|x_n - \mathfrak{x}_n\| \, \|(\nabla_x \mathbb{V})(t_n,x_n)\| \\ &\quad + \Big\| \tfrac{\sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*}{\mathbb{V}(t_n,\mathfrak{x}_n)} \Big\|_F n\|x_n - \mathfrak{x}_n\| \, \|(\nabla_x \mathbb{V})(t_n,x_n) - (\nabla_x \mathbb{V})(t_n,\mathfrak{x}_n)\| \Bigg] \\ &= 0. \end{aligned} \tag{2.47}$$

Next observe that the fact that $(0,T) \times O \ni (s,y) \mapsto \frac{\sigma(s,y)[\sigma(s,y)]^*}{\mathbb{V}(s,y)} (\operatorname{Hess}_x \mathbb{V})(s,y) \in \mathbb{R}^{d\times d}$ is continuous and the assumption that $\limsup_{n\to\infty}[|t_n - t_0| + \|x_n - x_0\|] = 0$ and $\limsup_{n\to\infty}[\sqrt{n}\|x_n - \mathfrak{x}_n\|] = 0$ show that

$$\begin{aligned} &\limsup_{n\to\infty} \Big| \operatorname{Tr}\Big(\tfrac{\sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*}{\mathbb{V}(t_n,\mathfrak{x}_n)} (\operatorname{Hess}_x \mathbb{V})(t_n,\mathfrak{x}_n) \\ &\quad - \tfrac{\sigma(t_0,x_0)[\sigma(t_0,x_0)]^*}{\mathbb{V}(t_0,x_0)} (\operatorname{Hess}_x \mathbb{V})(t_0,x_0) \Big) \Big| \\ &= \limsup_{n\to\infty} \Big| \operatorname{Tr}\Big(\tfrac{\sigma(t_n,x_n)[\sigma(t_n,x_n)]^*}{\mathbb{V}(t_n,x_n)} (\operatorname{Hess}_x \mathbb{V})(t_n,x_n) \\ &\quad - \tfrac{\sigma(t_0,x_0)[\sigma(t_0,x_0)]^*}{\mathbb{V}(t_0,x_0)} (\operatorname{Hess}_x \mathbb{V})(t_0,x_0) \Big) \Big| = 0. \end{aligned} \tag{2.48}$$

The fact that $0 < r_0 = \liminf_{n\to\infty} |r_n - \mathfrak{r}_n| = \limsup_{n\to\infty} |r_n - \mathfrak{r}_n| \leqslant \sup_{n\in\mathbb{N}}(|r_n| + |\mathfrak{r}_n|) < \infty$

therefore ensures that

$$\begin{aligned}
&\limsup_{n\to\infty}\bigg[\tfrac{1}{2}\operatorname{Tr}\bigg(\tfrac{\sigma(t_n,x_n)[\sigma(t_n,x_n)]^*}{\mathbb{V}(t_n,x_n)}r_n(\operatorname{Hess}_x\mathbb{V})(t_n,x_n)\\
&\qquad-\tfrac{\sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*}{\mathbb{V}(t_n,\mathfrak{x}_n)}\mathfrak{r}_n(\operatorname{Hess}_x\mathbb{V})(t_n,\mathfrak{x}_n)\bigg)\bigg]\\
&=\tfrac{1}{2}\limsup_{n\to\infty}\bigg[(r_n-\mathfrak{r}_n)\operatorname{Tr}\bigg(\tfrac{\sigma(t_n,x_n)[\sigma(t_n,x_n)]^*}{\mathbb{V}(t_n,x_n)}(\operatorname{Hess}_x\mathbb{V})(t_n,x_n)\bigg)\\
&\qquad+\mathfrak{r}_n\operatorname{Tr}\bigg(\tfrac{\sigma(t_n,x_n)[\sigma(t_n,x_n)]^*}{\mathbb{V}(t_n,x_n)}(\operatorname{Hess}_x\mathbb{V})(t_n,x_n)\\
&\qquad-\tfrac{\sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*}{\mathbb{V}(t_n,\mathfrak{x}_n)}(\operatorname{Hess}_x\mathbb{V})(t_n,\mathfrak{x}_n)\bigg)\bigg]\\
&=\tfrac{r_0}{2\mathbb{V}(t_0,x_0)}\operatorname{Tr}\big(\sigma(t_0,x_0)[\sigma(t_0,x_0)]^*(\operatorname{Hess}_x\mathbb{V})(t_0,x_0)\big).
\end{aligned}\tag{2.49}$$

Combining this with (2.42) and (2.47) ensures that

$$\begin{aligned}
&\limsup_{n\to\infty}\bigg[\tfrac{1}{2}\operatorname{Tr}\bigg(\tfrac{\sigma(t_n,x_n)[\sigma(t_n,x_n)]^*}{\mathbb{V}(t_n,x_n)}\Big(\mathbb{V}(t_n,x_n)nA_n+n(x_n-\mathfrak{x}_n)[(\nabla_x\mathbb{V})(t_n,x_n)]^*\Big)\\
&\quad+(\nabla_x\mathbb{V})(t_n,x_n)n(x_n-\mathfrak{x}_n)^*+r_n(\operatorname{Hess}_x\mathbb{V})(t_n,x_n)\bigg)\\
&\quad-\tfrac{1}{2}\operatorname{Tr}\bigg(\tfrac{\sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*}{\mathbb{V}(t_n,\mathfrak{x}_n)}\Big(\mathbb{V}(t_n,\mathfrak{x}_n)n\mathfrak{A}_n+n(x_n-\mathfrak{x}_n)[(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n)]^*\Big)\\
&\quad+(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n)n(x_n-\mathfrak{x}_n)^*+\mathfrak{r}_n(\operatorname{Hess}_x\mathbb{V})(t_n,\mathfrak{x}_n)\bigg)\bigg]\\
&\leqslant\tfrac{r_0}{2\mathbb{V}(t_0,x_0)}\operatorname{Tr}\big(\sigma(t_0,x_0)[\sigma(t_0,x_0)]^*(\operatorname{Hess}_x\mathbb{V})(t_0,x_0)\big).
\end{aligned}\tag{2.50}$$

Next note that the fact that $(0,T)\times O\ni(s,y)\mapsto\frac{1}{\mathbb{V}(s,y)}(\frac{\partial}{\partial t}\mathbb{V})(s,y)\in\mathbb{R}$ is continuous and the assumption that $0<r_0=\liminf_{n\to\infty}(r_n-\mathfrak{r}_n)=\limsup_{n\to\infty}(r_n-\mathfrak{r}_n)\leqslant\sup_{n\in\mathbb{N}}(|r_n|+|\mathfrak{r}_n|)<\infty$ prove that

$$\begin{aligned}
&\limsup_{n\to\infty}\Big[\tfrac{r_n}{\mathbb{V}(t_n,x_n)}(\tfrac{\partial}{\partial t}\mathbb{V})(t_n,x_n)-\tfrac{\mathfrak{r}_n}{\mathbb{V}(t_n,\mathfrak{x}_n)}(\tfrac{\partial}{\partial t}\mathbb{V})(t_n,\mathfrak{x}_n)\Big]\\
&=\limsup_{n\to\infty}\bigg[\tfrac{r_n-\mathfrak{r}_n}{\mathbb{V}(t_n,x_n)}(\tfrac{\partial}{\partial t}\mathbb{V})(t_n,x_n)\\
&\qquad+\mathfrak{r}_n\Big(\tfrac{1}{\mathbb{V}(t_n,x_n)}(\tfrac{\partial}{\partial t}\mathbb{V})(t_n,x_n)-\tfrac{1}{\mathbb{V}(t_n,\mathfrak{x}_n)}(\tfrac{\partial}{\partial t}\mathbb{V})(t_n,\mathfrak{x}_n)\Big)\bigg]\\
&=\tfrac{r_0}{\mathbb{V}(t_0,x_0)}(\tfrac{\partial}{\partial t}\mathbb{V})(t_0,x_0).
\end{aligned}\tag{2.51}$$

Furthermore, observe that (2.32) and the fact that $\limsup_{n\to\infty}[|t_n-t_0|+\|x_n-x_0\|]=0=\limsup_{n\to\infty}[\sqrt{n}\|x_n-\mathfrak{x}_n\|]$ ensure that

$$\limsup_{n\to\infty}\Big[n\|\mu(t_n,x_n)-\mu(t_n,\mathfrak{x}_n)\|\,\|x_n-\mathfrak{x}_n\|\Big]=0.\tag{2.52}$$

Combining this with the Cauchy-Schwarz inequality, the fact that $(0,T)\times O\ni(s,y)\mapsto\langle\frac{\mu(s,y)}{\mathbb{V}(s,y)},(\nabla_x\mathbb{V})(s,y)\rangle\in\mathbb{R}$ is continuous, and the assumption that $0<r_0=\liminf_{n\to\infty}(r_n-$

$\mathfrak{r}_n) = \limsup_{n\to\infty}(r_n - \mathfrak{r}_n) \leqslant \sup_{n\in\mathbb{N}}(|r_n| + |\mathfrak{r}_n|) < \infty$ implies that

$$
\begin{aligned}
&\limsup_{n\to\infty}\bigg[\frac{1}{\mathbb{V}(t_n,x_n)}\langle\mu(t_n,x_n), \mathbb{V}(t_n,x_n)n(x_n-\mathfrak{x}_n) + r_n(\nabla_x\mathbb{V})(t_n,x_n)\rangle \\
&\qquad - \frac{1}{\mathbb{V}(t_n,\mathfrak{x}_n)}\langle\mu(t_n,\mathfrak{x}_n), \mathbb{V}(t_n,\mathfrak{x}_n)n(x_n-\mathfrak{x}_n) + \mathfrak{r}_n(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n)\rangle\bigg] \\
&= \limsup_{n\to\infty}\bigg[\Big\langle \mu(t_n,x_n) - \mu(t_n,\mathfrak{x}_n), n(x_n-\mathfrak{x}_n)\Big\rangle \\
&\qquad + r_n\Big\langle\frac{\mu(t_n,x_n)}{\mathbb{V}(t_n,x_n)}, (\nabla_x\mathbb{V})(t_n,x_n)\Big\rangle - \mathfrak{r}_n\Big\langle\frac{\mu(t_n,\mathfrak{x}_n)}{\mathbb{V}(t_n,\mathfrak{x}_n)}, (\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n)\Big\rangle\bigg] \\
&\leqslant \limsup_{n\to\infty}\big[\|\mu(t_n,x_n) - \mu(t_n,\mathfrak{x}_n)\|n\|x_n-\mathfrak{x}_n\|\big] \qquad (2.53)\\
&\qquad + \limsup_{n\to\infty}\bigg[(r_n - \mathfrak{r}_n)\Big\langle\frac{\mu(t_n,x_n)}{\mathbb{V}(t_n,x_n)}, (\nabla_x\mathbb{V})(t_n,x_n)\Big\rangle\bigg] \\
&\qquad + \limsup_{n\to\infty}\bigg[\mathfrak{r}_n\bigg(\Big\langle\frac{\mu(t_n,x_n)}{\mathbb{V}(t_n,x_n)}, (\nabla_x\mathbb{V})(t_n,x_n)\Big\rangle \\
&\qquad - \Big\langle\frac{\mu(t_n,\mathfrak{x}_n)}{\mathbb{V}(t_n,\mathfrak{x}_n)}, (\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n)\Big\rangle\bigg)\bigg] \\
&= \frac{r_0}{\mathbb{V}(t_0,x_0)}\big\langle\mu(t_0,x_0), (\nabla_x\mathbb{V})(t_0,x_0)\big\rangle.
\end{aligned}
$$

Next note that the assumption that $f \in C([0,T] \times O \times \mathbb{R} \times \mathbb{R}^d, \mathbb{R})$ shows that for all compact $\mathcal{K} \subseteq [0,T] \times O \times \mathbb{R} \times \mathbb{R}^d$ it holds that

$$
\begin{aligned}
\limsup_{(0,\infty)\ni\varepsilon\to 0}\Big[&\sup\big(\{|f(s_1,y_1,a_1,w_1) - f(s_2,y_2,a_2,w_2)|\colon \\
&(s_1,y_1,a_1,w_1),(s_2,y_2,a_2,w_2) \in \mathcal{K}, |s_1 - s_2| \leqslant \varepsilon, \qquad (2.54)\\
&\|y_1 - y_2\| \leqslant \varepsilon, |a_1 - a_2| \leqslant \varepsilon, \|w_1 - w_2\| \leqslant \varepsilon\} \cup \{0\}\big)\Big] = 0.
\end{aligned}
$$

Moreover, observe that the assumption that for all $t \in [0,T]$, $x \in O$, $r, s \in \mathbb{R}$, $v, w \in \mathbb{R}^d$ it holds that $(f(t,x,r,v) - f(t,x,s,v))(r-s) \leqslant L_1|r-s|^2$ and $|f(t,x,r,v) - f(t,x,r,w)| \leqslant L_2\|v-w\|$ ensures that for all $n \in \mathbb{N}$ it holds that

$$
\begin{aligned}
&\tfrac{f(t_n,x_n,r_n\mathbb{V}(t_n,x_n),\mathbb{V}(t_n,x_n)n(x_n-\mathfrak{x}_n)+r_n(\nabla_x\mathbb{V})(t_n,x_n))}{V(t_n,x_n)} \\
&\qquad - \tfrac{f(t_n,\mathfrak{x}_n,\mathfrak{r}_n\mathbb{V}(t_n,\mathfrak{x}_n),\mathbb{V}(t_n,\mathfrak{x}_n)n(x_n-\mathfrak{x}_n)+\mathfrak{r}_n(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n))}{V(t_n,\mathfrak{x}_n)} \\
&\leqslant \tfrac{f(t_n,x_n,r_n\mathbb{V}(t_n,x_n),\mathbb{V}(t_n,x_n)n(x_n-\mathfrak{x}_n)+r_n(\nabla_x\mathbb{V})(t_n,x_n))}{\mathbb{V}(t_n,x_n)} \\
&\qquad - \tfrac{f(t_n,\mathfrak{x}_n,r_n\mathbb{V}(t_n,\mathfrak{x}_n),\mathbb{V}(t_n,\mathfrak{x}_n)n(x_n-\mathfrak{x}_n)+r_n(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n))}{\mathbb{V}(t_n,\mathfrak{x}_n)} \\
&\quad + \tfrac{f(t_n,\mathfrak{x}_n,r_n\mathbb{V}(t_n,\mathfrak{x}_n),\mathbb{V}(t_n,\mathfrak{x}_n)n(x_n-\mathfrak{x}_n)+r_n(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n))}{\mathbb{V}(t_n,\mathfrak{x}_n)} \\
&\qquad - \tfrac{f(t_n,\mathfrak{x}_n,\mathfrak{r}_n\mathbb{V}(t_n,\mathfrak{x}_n),\mathbb{V}(t_n,\mathfrak{x}_n)n(x_n-\mathfrak{x}_n)+r_n(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n))}{\mathbb{V}(t_n,\mathfrak{x}_n)} \qquad (2.55)\\
&\quad + \tfrac{f(t_n,\mathfrak{x}_n,\mathfrak{r}_n\mathbb{V}(t_n,\mathfrak{x}_n),\mathbb{V}(t_n,\mathfrak{x}_n)n(x_n-\mathfrak{x}_n)+r_n(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n))}{\mathbb{V}(t_n,\mathfrak{x}_n)} \\
&\qquad - \tfrac{f(t_n,\mathfrak{x}_n,\mathfrak{r}_n\mathbb{V}(t_n,\mathfrak{x}_n),\mathbb{V}(t_n,\mathfrak{x}_n)n(x_n-\mathfrak{x}_n)+\mathfrak{r}_n(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n))}{\mathbb{V}(t_n,\mathfrak{x}_n)} \\
&\leqslant \tfrac{f(t_n,x_n,r_n\mathbb{V}(t_n,x_n),\mathbb{V}(t_n,x_n)n(x_n-\mathfrak{x}_n)+r_n(\nabla_x\mathbb{V})(t_n,x_n))}{\mathbb{V}(t_n,x_n)} \\
&\qquad - \tfrac{f(t_n,\mathfrak{x}_n,r_n\mathbb{V}(t_n,\mathfrak{x}_n),\mathbb{V}(t_n,\mathfrak{x}_n)n(x_n-\mathfrak{x}_n)+r_n(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n))}{\mathbb{V}(t_n,\mathfrak{x}_n)} \\
&\quad + \tfrac{L_1(r_n-\mathfrak{r}_n)\mathbb{V}(t_n,\mathfrak{x}_n)}{\mathbb{V}(t_n,\mathfrak{x}_n)} + \tfrac{L_2\|(r_n-\mathfrak{r}_n)(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n)\|}{\mathbb{V}(t_n,\mathfrak{x}_n)}.
\end{aligned}
$$

This and (2.54) demonstrate that

$$\begin{aligned}
&\limsup_{n\to\infty}\Bigg[\tfrac{f(t_n,x_n,r_n\mathbb{V}(t_n,x_n),\mathbb{V}(t_n,x_n)n(x_n-\mathfrak{x}_n)+r_n(\nabla_x\mathbb{V})(t_n,x_n))}{\mathbb{V}(t_n,x_n)}\\
&\qquad-\tfrac{f(t_n,\mathfrak{x}_n,\mathfrak{r}_n\mathbb{V}(t_n,\mathfrak{x}_n),\mathbb{V}(t_n,\mathfrak{x}_n)n(x_n-\mathfrak{x}_n)+\mathfrak{r}_n(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n))}{\mathbb{V}(t_n,\mathfrak{x}_n)}\Bigg]\\
&\leqslant\limsup_{n\to\infty}\left[L_1(r_n-\mathfrak{r}_n)+\tfrac{L_2|r_n-\mathfrak{r}_n|\|(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n)\|}{\mathbb{V}(t_n,\mathfrak{x}_n)}\right]=L_1r_0+L_2r_0\tfrac{\|(\nabla_x\mathbb{V})(t_0,x_0)\|}{\mathbb{V}(t_0,x_0)}.
\end{aligned}\tag{2.56}$$

Combing this with (2.35), (2.36), (2.37) (2.50), (2.51), and (2.53) proves that

$$\begin{aligned}
&\limsup_{n\to\infty}[H(t_n,x_n,r_n,n(x_n-\mathfrak{x}_n),nA_n)-H(t_n,\mathfrak{x}_n,\mathfrak{r}_n,n(x_n-\mathfrak{x}_n),n\mathfrak{A}_n)]\\
&=\limsup_{n\to\infty}\Bigg[\tfrac{r_n}{\mathbb{V}(t_n,x_n)}(\tfrac{\partial}{\partial t}\mathbb{V})(t_n,x_n)-\tfrac{\mathfrak{r}_n}{\mathbb{V}(t_n,\mathfrak{x}_n)}(\tfrac{\partial}{\partial t}\mathbb{V})(t_n,\mathfrak{x}_n)\\
&\quad+\tfrac{1}{2}\operatorname{Tr}\Bigg(\tfrac{\sigma(t_n,x_n)[\sigma(t_n,x_n)]^*}{\mathbb{V}(t_n,x_n)}\Big(\mathbb{V}(t_n,x_n)nA_n+n(x_n-\mathfrak{x}_n)[(\nabla_x\mathbb{V})(t_n,x_n)]^*\Big)\\
&\quad+(\nabla_x\mathbb{V})(t_n,x_n)n(x_n-\mathfrak{x}_n)^*+r_n(\operatorname{Hess}_x\mathbb{V})(t_n,x_n)\Bigg)\\
&\quad-\tfrac{1}{2}\operatorname{Tr}\Bigg(\tfrac{\sigma(t_n,\mathfrak{x}_n)[\sigma(t_n,\mathfrak{x}_n)]^*}{\mathbb{V}(t_n,\mathfrak{x}_n)}\Big(\mathbb{V}(t_n,\mathfrak{x}_n)n\mathfrak{A}_n+n(x_n-\mathfrak{x}_n)[(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n)]^*\Big)\\
&\quad+(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n)n(x_n-\mathfrak{x}_n)^*+\mathfrak{r}_n(\operatorname{Hess}_x\mathbb{V})(t_n,\mathfrak{x}_n)\Bigg)\\
&\quad+\tfrac{1}{\mathbb{V}(t_n,x_n)}\langle\mu(t_n,x_n),\mathbb{V}(t_n,x_n)n(x_n-\mathfrak{x}_n)+r_n(\nabla_x\mathbb{V})(t_n,x_n)\rangle\\
&\quad-\tfrac{1}{\mathbb{V}(t_n,\mathfrak{x}_n)}\langle\mu(t_n,\mathfrak{x}_n),\mathbb{V}(t_n,\mathfrak{x}_n)n(x_n-\mathfrak{x}_n)+\mathfrak{r}_n(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n)\rangle\\
&\quad+\tfrac{f(t_n,x_n,r_n\mathbb{V}(t_n,x_n),\mathbb{V}(t_n,x_n)n(x_n-\mathfrak{x}_n)+r_n(\nabla_x\mathbb{V})(t_n,x_n))}{\mathbb{V}(t_n,x_n)}\\
&\quad-\tfrac{f(t_n,\mathfrak{x}_n,\mathfrak{r}_n\mathbb{V}(t_n,\mathfrak{x}_n),\mathbb{V}(t_n,\mathfrak{x}_n)n(x_n-\mathfrak{x}_n)+\mathfrak{r}_n(\nabla_x\mathbb{V})(t_n,\mathfrak{x}_n))}{\mathbb{V}(t_n,\mathfrak{x}_n)}\Bigg]\\
&\leqslant\tfrac{r_0}{\mathbb{V}(t_0,x_0)}\Bigg[(\tfrac{\partial}{\partial t}\mathbb{V})(t_0,x_0)+\tfrac{1}{2}\operatorname{Tr}(\sigma(t_0,x_0)[\sigma(t_0,x_0)]^*(\operatorname{Hess}_x\mathbb{V})(t_0,x_0)\\
&\quad+\langle\mu(t_0,x_0),(\nabla_x\mathbb{V})(t_0,x_0)\rangle+L_1\mathbb{V}(t_0,x_0)+L_2\|(\nabla_x\mathbb{V})(t_0,x_0)\|\Bigg]\leqslant 0.
\end{aligned}\tag{2.57}$$

This, [10, Corollary 3.4], (2.38), and (2.39) demonstrate that $v_1\leqslant v_2$ and $v_2\leqslant v_1$. This implies $v_1=v_2$. Hence, we obtain that $u_1=u_2$. The proof of Proposition 2.1.8 is thus complete. □

2.2 Bismut-Elworthy-Li type formula

In this section we derive in Theorem 2.2.3 a Bismut-Elworthy-Li type formula that holds under the assumptions we used in Theorem 1.2.5. To achieve this, we use findings from Malliavin calculus to establish the representation in (2.71) for the derivative of the PDE solution. For this, we follow the notation of [86] and denote the Malliavin derivative of a random variable $X\in\mathbb{D}^{1,2}$ by $\{D_tX\colon t\in[0,T]\}$ where $\mathbb{D}^{1,2}\subseteq L^2(\Omega;\mathbb{R}^d)$ denotes the space of Malliavin differentiable random variables. For the Skorohod integral of a Skorohod integrable stochastic process $u\in L^2([0,T]\times\Omega;\mathbb{R}^d)$ we write $\int_0^T u_r\,\delta W_r$. To prove the Bismut-Elworthy-Li type formula in Theorem 2.2.3 we need the following results, Lemma 2.2.1 and Lemma 2.2.2. Lemma 2.2.1 establishes a representation for the Malliavin derivative of a solution of the SDE in (2.59) below under the global monotonicity assumption. The proof of this lemma is based on the ideas in [86, ğ2.3.1].

Lemma 2.2.1. *Let $d, m \in \mathbb{N}$, $c, T \in (0,\infty)$, let $O \subseteq \mathbb{R}^d$ be an open set, let $\langle \cdot,\cdot \rangle \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the standard Euclidean scalar product on $\mathbb{R}^d$, let $\|\cdot\| \colon \mathbb{R}^d \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, let $\|\cdot\|_F \colon \mathbb{R}^{d\times m} \to [0,\infty)$ be the Frobenius norm on $\mathbb{R}^{d\times m}$, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_s)_{s\in[0,T]})$ be a filtered probability space satisfying the usual conditions, let $W = (W^1, W^2, \dots, W^m) \colon [0,T]\times\Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_s)_{s\in[0,T]}$-Brownian motion, let $\mu = (\mu_1, \mu_2, \dots, \mu_d) \in C^{0,1}([0,T]\times O, \mathbb{R}^d)$ and $\sigma = (\sigma_{ij})_{i\in\{1,2,\dots,d\}, j\in\{1,2,\dots,m\}} \in C^{0,1}([0,T]\times O, \mathbb{R}^{d\times m})$ satisfy for all $s \in [0,T]$, $x, y \in O$ that*

$$\max\{\langle x-y, \mu(s,x)-\mu(s,y)\rangle, \tfrac{1}{2}\|\sigma(s,x)-\sigma(s,y)\|_F^2\} \leqslant \tfrac{c}{2}\|x-y\|^2, \tag{2.58}$$

for every $x \in O$ let $X^x = ((X_s^{x,(1)}, X_s^{x,(2)}, \dots, X_s^{x,(d)}))_{s\in[0,T]} \colon [0,T]\times\Omega \to O$ be an adapted stochastic process with continuous sample paths satisfying that for all $s \in [0,T]$ it holds a.s. that

$$X_s = x + \int_0^s \mu(r, X_r)\,\mathrm{d}r + \int_0^s \sigma(r, X_r)\,\mathrm{d}W_r, \tag{2.59}$$

and assume for all $\omega \in \Omega$ that $([0,T]\times O \ni (s,x) \mapsto X_s^x(\omega) \in O) \in C^{0,1}([t,T]\times O, O)$. Then it holds a.s. for all $t \in [0,T]$, $x \in O$ that $\frac{\partial}{\partial x}X_t^x$ is invertible and it holds a.s. for all $x \in O$ and Lebesgue almost all $t \in [0,T]$, $s \in [t,T]$ that

$$D_t X_s^x = \left(\frac{\partial}{\partial x}X_s^x\right)\left(\frac{\partial}{\partial x}X_t^x\right)^{-1}\sigma(t, X_t^x). \tag{2.60}$$

Proof of Lemma 2.2.1. Throughout this proof let $I_d \in \mathbb{R}^{d\times d}$ be the unity matrix and for every $x \in O$ let $Z^x = (Z_s^{x,(i,j)})_{i,j\in\{1,2,\dots,d\}, s\in[0,T]} \colon [0,T]\times\Omega \to \mathbb{R}^{d\times d}$ be an adapted stochastic process with continuous sample paths satisfying that for all $i, j \in \{1,2,\dots,d\}$, $s \in [0,T]$ it holds a.s. that

$$\begin{aligned} Z_s^{x,(i,j)} = \delta_{ij} - \int_0^s \sum_{\alpha=1}^d Z_r^{x,(i,\alpha)} \Bigg[\left(\frac{\partial \mu_\alpha}{\partial x_j}\right)(r, X_r^x) \\ - \sum_{n=1}^m \sum_{p=1}^d \left(\frac{\partial \sigma_{\alpha n}}{\partial x_p}\right)(r, X_r^x)\left(\frac{\partial \sigma_{pn}}{\partial x_j}\right)(r, X_r^x)\Bigg]\mathrm{d}r \\ - \sum_{l=1}^m \int_0^s \sum_{\alpha=1}^d Z_r^{x,(i,\alpha)}\left(\frac{\partial \sigma_{\alpha l}}{\partial x_j}\right)(r, X_r^x)\mathrm{d}W_r^l. \end{aligned} \tag{2.61}$$

First note that (2.59), the Leibniz integral rule, the chain rule, the assumption that for all $\omega \in \Omega$ it holds that $([0,T]\times O \ni (s,x) \mapsto X_s^x(\omega) \in O) \in C^{0,1}([t,T]\times O, O)$, and the assumption that $\mu \in C^{0,1}([0,T]\times O, \mathbb{R}^d)$ and $\sigma \in C^{0,1}([0,T]\times O, \mathbb{R}^{d\times m})$ imply that for all $i, j \in \{1,2,\dots,d\}$, $s \in [0,T]$, $x \in O$ it holds a.s. that

$$\begin{aligned} \frac{\partial}{\partial x_j}X_s^{x,(i)} = \delta_{ij} + \int_0^s \sum_{k=1}^d \left[\left(\frac{\partial \mu_i}{\partial x_k}\right)(r, X_r^x)\left(\frac{\partial}{\partial x_j}X_s^{x,(k)}\right)\right]\mathrm{d}r \\ + \sum_{l=1}^m \left[\int_0^s \sum_{k=1}^d \left[\left(\frac{\partial \sigma_{il}}{\partial x_k}\right)(r, X_r^x)\left(\frac{\partial}{\partial x_j}X_s^{x,(k)}\right)\right]\mathrm{d}W_r^l\right]. \end{aligned} \tag{2.62}$$

Itô's lemma and (2.61) therefore ensure that for all $i, k \in \{1, 2, \ldots, d\}$, $s \in [0, T]$, $x \in O$ it holds a.s. that

$$\begin{aligned}
&\sum_{j=1}^{d} Z_s^{x,(i,j)} \left(\frac{\partial}{\partial x_k} X_s^{x,(j)} \right) \\
&= \sum_{j=1}^{d} \Bigg(\delta_{ij}\delta_{jk} + \int_0^s Z_r^{x,(i,j)} \sum_{\alpha=1}^{d} \left[\left(\frac{\partial \mu_j}{\partial x_\alpha} \right)(r, X_r^x) \left(\frac{\partial}{\partial x_k} X_r^{x,(\alpha)} \right) \right] \mathrm{d}r \\
&\quad + \sum_{l=1}^{m} \left[\int_0^s Z_r^{x,(i,j)} \sum_{\alpha=1}^{d} \left[\left(\frac{\partial \sigma_{jl}}{\partial x_\alpha} \right)(r, X_r^x) \left(\frac{\partial}{\partial x_k} X_r^{x,(\alpha)} \right) \right] \mathrm{d}W_r^l \right] \\
&\quad - \int_0^s \sum_{\alpha=1}^{d} \Bigg[Z_r^{x,(i,\alpha)} \Bigg(\left(\frac{\partial \mu_\alpha}{\partial x_j} \right)(r, X_r^x) - \sum_{n=1}^{m} \sum_{p=1}^{d} \left[\left(\frac{\partial \sigma_{\alpha n}}{\partial x_p} \right)(r, X_r^x) \left(\frac{\partial \sigma_{pn}}{\partial x_j} \right)(r, X_r^x) \right] \Bigg) \\
&\qquad \cdot \left(\frac{\partial}{\partial x_k} X_r^{x,(j)} \right) \Bigg] \mathrm{d}r \\
&\quad - \sum_{l=1}^{m} \left[\int_0^s \sum_{\alpha=1}^{d} \left[Z_r^{x,(i,\alpha)} \left(\frac{\partial \sigma_{\alpha l}}{\partial x_j} \right)(r, X_r^x) \left(\frac{\partial}{\partial x_k} X_r^{x,(j)} \right) \right] \mathrm{d}W_r^l \right] \\
&\quad - \int_0^s \sum_{\alpha=1}^{d} \left[Z_r^{x,(i,\alpha)} \sum_{n=1}^{m} \sum_{p=1}^{d} \left[\left(\frac{\partial \sigma_{\alpha n}}{\partial x_j} \right)(r, X_r^x) \left(\frac{\partial \sigma_{jn}}{\partial x_p} \right)(r, X_r^x) \left(\frac{\partial}{\partial x_k} X_r^{x,(p)} \right) \right] \right] \mathrm{d}r \Bigg) = \delta_{ik}.
\end{aligned} \tag{2.63}$$

Combining this with, e.g., [20, Proposition 4.1] and [20, Proposition 4.4] (applied for every $s \in [0, T]$ $x \in O$, with $n \curvearrowleft d$, $A \curvearrowleft (\frac{\partial}{\partial x} X_s^x)$, $B \curvearrowleft Z_s^x$ in the notation of [20, Proposition 4.4]) implies that for all $s \in [0, T]$ it holds a.s. that $(\frac{\partial}{\partial x} X_s^x) Z_s^x = I_d = Z_s^x (\frac{\partial}{\partial x} X_s^x)$. This together with path continuity implies that a.s. it holds for all $s \in [0, T]$, $x \in O$ that $(\frac{\partial}{\partial x} X_s^x)$ is invertible. Next note that [60, Corollary 3.5] (applied for every $x \in O$ with $p \curvearrowleft 2$, $\theta \curvearrowleft (\Omega \ni \omega \mapsto x \in \mathbb{R}^d)$, $b \curvearrowleft ([0, T] \times \Omega \times \mathbb{R}^d \ni (s, \omega, x) \mapsto \mu(s, x) \in \mathbb{R}^d)$, $\sigma \curvearrowleft ([0, T] \times \Omega \times \mathbb{R}^d \ni (s, \omega, x) \mapsto \sigma(s, x) \in \mathbb{R}^{d \times d})$ in the notation of [60, Corollary 3.5]), the assumption that $\mu \in C^{0,1}([0, T] \times \mathbb{R}^d, \mathbb{R}^d)$, $\sigma \in C^{0,1}([0, T] \times \mathbb{R}^d, \mathbb{R}^{d \times m})$, and (2.58) demonstrate that for all $i \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$, $t \in [0, T]$, $s \in [t, T]$, $x \in O$ it holds that $X^x \in \mathbb{D}^{1,2}$ and it holds a.s. that

$$\begin{aligned}
D_t^j X_s^{x,(i)} =& \sigma_{ij}(t, X_t^x) + \int_t^s \sum_{k=1}^{d} \left(\frac{\partial \mu_i}{\partial x_k} \right)(r, X_r^x) D_t^j X_r^{x,(k)} \, \mathrm{d}r \\
&+ \sum_{l=1}^{m} \int_t^s \sum_{k=1}^{d} \left(\frac{\partial \sigma_{il}}{\partial x_k} \right)(r, X_r^x) D_t^j X_r^{x,(k)} \, \mathrm{d}W_r^l.
\end{aligned} \tag{2.64}$$

Moreover, observe that (2.62) and the fact that for all $s \in [0, T]$, $x \in O$ it holds a.s. that $(\frac{\partial}{\partial x} X_s^x) Z_s^x = I_d = Z_s^x (\frac{\partial}{\partial x} X_s^x)$ imply that for all $i \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$,

$t \in [0,T]$, $s \in [t,T]$, $x \in O$ it holds a.s. that

$$
\begin{aligned}
&\sigma_{ij}(t, X_t^x) + \int_t^s \sum_{k=1}^d \left(\frac{\partial \mu_i}{\partial x_k}\right)(r, X_r^x) \sum_{n=1}^d \sum_{p=1}^d \left(\frac{\partial}{\partial_n} X_r^{x,(k)}\right) Z_t^{x,(n,p)} \sigma_{pj}(t, X_t^x) \,\mathrm{d}r \\
&\quad + \sum_{l=1}^m \int_t^s \sum_{k=1}^d \left(\frac{\partial \sigma_{il}}{\partial x_k}\right)(r, X_r^x) \sum_{n=1}^d \sum_{p=1}^d \left(\frac{\partial}{\partial_n} X_r^{x,(k)}\right) Z_t^{x,(n,p)} \sigma_{pj}(t, X_t^x) \,\mathrm{d}W_r^l \\
&= \sigma_{ij}(t, X_t^x) + \sum_{n=1}^d \sum_{p=1}^d \left[\left(\frac{\partial}{\partial_n} X_s^{x,(i)}\right) - \left(\frac{\partial}{\partial_n} X_t^{x,(i)}\right)\right] Z_t^{x,(n,p)} \sigma_{pj}(t, X_t^x) \\
&= \sum_{n=1}^d \sum_{p=1}^d \left(\frac{\partial}{\partial_n} X_t^{x,(i)}\right) Z_t^{x,(n,p)} \sigma_{pj}(t, X_t^x) \\
&\quad + \sum_{n=1}^d \sum_{p=1}^d \left[\left(\frac{\partial}{\partial_n} X_s^{x,(i)}\right) - \left(\frac{\partial}{\partial_n} X_t^{x,(i)}\right)\right] Z_t^{x,(n,p)} \sigma_{pj}(t, X_t^x) \\
&= \sum_{n=1}^d \sum_{p=1}^d \left(\frac{\partial}{\partial_n} X_s^{x,(i)}\right) Z_t^{x,(n,p)} \sigma_{pj}(t, X_t^x).
\end{aligned}
\tag{2.65}
$$

This, (2.64), and the fact that linear SDEs are pathwise unique establish (2.60). The proof of Lemma 2.2.1 is thus complete. □

The following Lemma is a well-known result on the uniform convergence of the derivatives of an approximating sequence, generalized to d dimensions. Lemma 2.2.2 follows from the coordinate wise application of [91, Theorem 7.17].

Lemma 2.2.2. *Let $d \in \mathbb{N}$, let $O \subseteq \mathbb{R}^d$ be open, let $f_0 \colon O \to \mathbb{R}$, $f_n \in C^1(O, \mathbb{R})$, $n \in \mathbb{N}$, and $g \colon O \to \mathbb{R}$ satisfy that $(f_n)_{n\in\mathbb{N}}$ converges pointwise to f_0 on O and $(\nabla f_n)_{n\in\mathbb{N}}$ converges uniformly to g on every compact $K \subseteq O$. Then it holds that $f_0 \in C^1(O, \mathbb{R})$ and $\nabla f_0 = g$.*

Making use of Lemma Lemma 2.2.1 and Lemma 2.2.2 leads to the following theorem, Theorem 2.2.3, which establishes a Bismut-Elworthy-Li type formula under global monotonicity assumption on the coefficients of the considered SDE.

Theorem 2.2.3. *Let $d \in \mathbb{N}$, $c \in [0, \infty)$, $\alpha, T \in (0, \infty)$, $t \in [0, T)$, let $O \subseteq \mathbb{R}^d$ be an open set, let $\langle \cdot, \cdot \rangle \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the standard Euclidean scalar product on $\mathbb{R}^d$, let $\|\cdot\| \colon \mathbb{R}^d \to [0, \infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, let $\|\cdot\|_F \colon \mathbb{R}^{d\times d} \to [0, \infty)$ be the Frobenius norm on $\mathbb{R}^{d\times d}$, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_s)_{s\in[0,T]})$ be a filtered probability space satisfying the usual conditions, let $W \colon [0, T] \times \Omega \to \mathbb{R}^d$ be a standard $(\mathbb{F}_s)_{s\in[0,T]}$-Brownian motion, let $\mu \in C^{0,1}([0, T] \times O, \mathbb{R}^d)$, $\sigma \in C^{0,1}([0, T] \times O, \mathbb{R}^{d\times d})$ satisfy for all $s \in [t, T]$, $x, y \in O$, $v \in \mathbb{R}^d$ that*

$$
\max\{\langle x - y, \mu(s, x) - \mu(s, y)\rangle, \tfrac{1}{2}\|\sigma(s, x) - \sigma(s, y)\|_F^2\} \leqslant \tfrac{c}{2}\|x - y\|^2 \tag{2.66}
$$

and $v^*\sigma(s, x)(\sigma(s, x))^* v \geqslant \alpha\|v\|^2$, for every $x \in O$ let $X^x = (X_s^x)_{s\in[t,T]} \colon [t, T] \times \Omega \to O$ be an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in [t, T]$ it holds a.s. that

$$
X_s^x = x + \int_t^s \mu(r, X_r^x) \,\mathrm{d}r + \int_t^s \sigma(r, X_r^x) \,\mathrm{d}W_r, \tag{2.67}
$$

assume for all $\omega \in \Omega$ *that* $\big([t,T]\times O \ni (s,x) \mapsto X_s^x(\omega) \in \mathbb{R}^d\big) \in C^{0,1}([t,T]\times O, O)$*, let* $f \in C(O,\mathbb{R})$ *satisfy for all* $x \in O$ *that* $\mathbb{E}[|f(X_T^x)|^2] < \infty$*, let* $u\colon O \to \mathbb{R}$ *satisfy for all* $x \in O$ *that*

$$u(x) = \mathbb{E}[f(X_T^x)], \tag{2.68}$$

and for every $x \in O$ *let* $Z^x = (Z_s^x)_{s\in(t,T]}\colon (t,T]\times\Omega \to \mathbb{R}^d$ *be an* $(\mathbb{F}_s)_{s\in(t,T]}$*-adapted stochastic process with continuous sample paths satisfying that for all* $s \in (t,T]$ *it holds a.s. that*

$$Z_s^x = \frac{1}{s-t}\int_t^s (\sigma(r,X_r^x))^{-1}\left(\frac{\partial}{\partial x}X_r^x\right)\mathrm{d}W_r. \tag{2.69}$$

Then

(i) for all $x \in O$ *it holds that*

$$\mathbb{E}\left[\|f(X_T^x)Z_T^x\|\right] < \infty, \tag{2.70}$$

(ii) it holds that $u \in C^1(O,\mathbb{R})$*, and*

(iii) for all $x \in O$ *it holds that*

$$\big(\nabla u\big)(x) = \mathbb{E}[f(X_T^x)Z_T^x]\,. \tag{2.71}$$

Proof of Theorem 2.2.3. Throughout this proof let $\mathbf{e}_1,\mathbf{e}_2,\dots,\mathbf{e}_d \in \mathbb{R}^d$ satisfy that $\mathbf{e}_1 = (1,0,\dots,0)$, $\mathbf{e}_2 = (0,1,0,\dots,0)$, $\dots$, $\mathbf{e}_d = (0,\dots,0,1)$. W.l.o.g. we assume that O is non-empty. First note that item (vi) of Lemma 1.2.2 proves that for all $x \in O$ it holds that

$$\mathbb{E}\Big[\|Z_T^x\|^2\Big] \leqslant \frac{d}{\alpha(T-t)^2}\int_t^T \exp(2(r-t)c)\,\mathrm{d}r. \tag{2.72}$$

The Cauchy-Schwarz inequality and the assumption that for all $x \in O$ it holds that $\mathbb{E}[|f(X_T^x)|^2] < \infty$ hence show that for all $x \in O$ it holds that

$$\mathbb{E}\left[\|f(X_T^x)Z_T^x\|\right] \leqslant \mathbb{E}\left[|f(X_T^x)|\|Z_T^x\|\right] \leqslant \left(\mathbb{E}\left[|f(X_T^x)|^2\right]\right)^{\frac{1}{2}}\left(\mathbb{E}\left[\|Z_T^x\|^2\right]\right)^{\frac{1}{2}} < \infty. \tag{2.73}$$

This establishes item (i). Next we prove items (ii) and (iii) in three steps.
Step 1: In addition to the assumptions of Theorem 2.2.3 we assume in step 1 that $f \in C_c^\infty(O,\mathbb{R})$. Observe that the assumption that $f \in C_c^\infty(O,\mathbb{R})$ ensures that there exists $L \in (0,\infty)$ such that for all $x \in O$ it holds that $\|(\nabla f)(x)\| \leqslant L$. The chain rule, the fundamental theorem of calculus, Jensen's inequality, Fubini's theorem, and item (ii) of Lemma 1.2.2 hence demonstrate that for all $h \in \mathbb{R}\backslash\{0\}$, $j \in \{1,2,\dots,d\}$, $x \in O$ it holds that

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{f(X_T^{x+h\mathbf{e}_j}) - f(X_T^x)}{h}\right|^2\right] = \mathbb{E}\left[\left|\int_0^1 (\nabla f)(X_T^{x+\lambda h\mathbf{e}_j})\Big(\frac{\partial}{\partial x_j}X_T^{x+\lambda h\mathbf{e}_j}\Big)\,\mathrm{d}\lambda\right|^2\right]\\
&\leqslant \mathbb{E}\left[\int_0^1 \left|(\nabla f)(X_T^{x+\lambda h\mathbf{e}_j})\Big(\frac{\partial}{\partial x_j}X_T^{x+\lambda h\mathbf{e}_j}\Big)\right|^2\mathrm{d}\lambda\right]\\
&= \int_0^1 \mathbb{E}\left[\left|(\nabla f)(X_T^{x+\lambda h\mathbf{e}_j})\Big(\frac{\partial}{\partial x_j}X_T^{x+\lambda h\mathbf{e}_j}\Big)\right|^2\right]\mathrm{d}\lambda\\
&\leqslant \int_0^1 \mathbb{E}\left[\left\|(\nabla f)(X_T^{x+\lambda h\mathbf{e}_j})\right\|^2\left\|\frac{\partial}{\partial x_j}X_T^{x+\lambda h\mathbf{e}_j}\right\|^2\right]\mathrm{d}\lambda \leqslant L^2\int_0^1 \mathbb{E}\left[\left\|\frac{\partial}{\partial x_j}X_T^{x+\lambda h\mathbf{e}_j}\right\|^2\right]\mathrm{d}\lambda\\
&\leqslant L^2\int_0^1 \exp\big(2c(T-t)\big)\,\mathrm{d}\lambda = L^2\exp(2c(T-t)).
\end{aligned} \tag{2.74}$$

In addition, observe that the chain rule, the assumption that $f \in C^\infty(O, \mathbb{R})$, and the fact that for all $\omega \in \Omega$, $s \in [t, T]$ it holds that $(O \ni x \mapsto X_s^x(\omega) \in O) \in C^1(O, O)$ ensure that for all $j \in \{1, 2, \dots, d\}$, $x \in O$, $\omega \in \Omega$ it holds that

$$\begin{aligned} \lim_{\mathbb{R}\backslash\{0\}\ni h \to 0} \frac{f(X_T^{x+h\mathbf{e}_j}(\omega)) - f(X_T^x(\omega))}{h} &= \frac{\partial}{\partial x_j}(f(X_T^x(\omega))) \\ &= (\nabla f)(X_T^x(\omega))\Big(\frac{\partial}{\partial x_j}(X_T^x(\omega))\Big). \end{aligned} \tag{2.75}$$

This, (2.74), and the Vitali convergence theorem demonstrate that for all $j \in \{1, 2, \dots, d\}$, $x \in O$ it holds that

$$\begin{aligned} 0 &= \lim_{\mathbb{R}\backslash\{0\}\ni h \to 0} \mathbb{E}\bigg[\bigg|\frac{f(X_T^{x+h\mathbf{e}_j}) - f(X_T^x)}{h} - (\nabla f)(X_T^x)\Big(\frac{\partial}{\partial x_j}X_T^x\Big)\bigg|\bigg] \\ &\geqslant \limsup_{\mathbb{R}\backslash\{0\}\ni h \to 0} \bigg|\mathbb{E}\bigg[\frac{f(X_T^{x+h\mathbf{e}_j}) - f(X_T^x)}{h} - (\nabla f)(X_T^x)\Big(\frac{\partial}{\partial x_j}X_T^x\Big)\bigg]\bigg| \\ &= \limsup_{\mathbb{R}\backslash\{0\}\ni h \to 0} \bigg|\frac{u(x + h\mathbf{e}_j) - u(x)}{h} - \mathbb{E}\Big[(\nabla f)(X_T^x)\Big(\frac{\partial}{\partial x_j}X_T^x\Big)\Big]\bigg| \geqslant 0. \end{aligned} \tag{2.76}$$

This proves that for all $j \in \{1, 2, \dots, d\}$, $x \in O$ it holds that

$$\lim_{\mathbb{R}\backslash\{0\}\ni h \to 0} \frac{u(x + h\mathbf{e}_j) - u(x)}{h} = \mathbb{E}\Big[(\nabla f)(X_T^x)\Big(\frac{\partial}{\partial x_j}X_T^x\Big)\Big]. \tag{2.77}$$

In addition, note that item (ii) of Lemma 1.2.2 and the fact that for all $x \in O$ it holds that $\|(\nabla f)(x)\| \leqslant L$ demonstrate that for all $h \in \mathbb{R}$, $j \in \{1, 2, \dots, d\}$, $x \in O$ it holds that

$$\begin{aligned} &\mathbb{E}\bigg[\Big\|(\nabla f)(X_T^{x+h\mathbf{e}_j})\Big(\frac{\partial}{\partial x_j}X_T^{x+h\mathbf{e}_j}\Big)\Big\|^2\bigg] \leqslant \mathbb{E}\bigg[\big\|(\nabla f)(X_T^{x+h\mathbf{e}_j})\big\|^2\Big\|\frac{\partial}{\partial x_j}X_T^{x+h\mathbf{e}_j}\Big\|^2\bigg] \\ &\leqslant L^2\mathbb{E}\bigg[\Big\|\frac{\partial}{\partial x_j}X_T^{x+h\mathbf{e}_j}\Big\|^2\bigg] \leqslant L^2\exp(2c(T-t)). \end{aligned} \tag{2.78}$$

Moreover, observe that the assumption that $f \in C^\infty(O, \mathbb{R})$ and the fact that for all $s \in [t, T]$, $\omega \in \Omega$ it holds that $(O \ni x \mapsto X_s^x(\omega) \in O) \in C^1(O, O)$ show that for all $j \in \{1, 2, \dots, d\}$, $x \in O$, $\omega \in \Omega$ it holds that

$$\lim_{\mathbb{R}\ni h \to 0} \Big[(\nabla f)(X_T^{x+h\mathbf{e}_j}(\omega))\Big(\frac{\partial}{\partial x_j}X_T^{x+h\mathbf{e}_j}(\omega)\Big)\Big] = (\nabla f)(X_T^x(\omega))\Big(\frac{\partial}{\partial x_j}X_T^x(\omega)\Big). \tag{2.79}$$

Combining this, (2.78), and the Vitali convergence theorem shows that for all $x \in O$, $j \in \{1, 2, \dots, d\}$ it holds that

$$\begin{aligned} 0 &= \lim_{\mathbb{R}\ni h \to 0} \mathbb{E}\Big[\Big|(\nabla f)(X_T^{x+h\mathbf{e}_j})\Big(\frac{\partial}{\partial x_j}X_T^{x+h\mathbf{e}_j}\Big) - (\nabla f)(X_T^x)\Big(\frac{\partial}{\partial x_j}X_T^x\Big)\Big|\Big] \\ &\geqslant \limsup_{\mathbb{R}\ni h \to 0} \Big|\mathbb{E}\Big[(\nabla f)(X_T^{x+h\mathbf{e}_j})\Big(\frac{\partial}{\partial x_j}X_T^{x+h\mathbf{e}_j}\Big) - (\nabla f)(X_T^x)\Big(\frac{\partial}{\partial x_j}X_T^x\Big)\Big]\Big| \\ &= \limsup_{\mathbb{R}\ni h \to 0} \Big|\mathbb{E}\Big[(\nabla f)(X_T^{x+h\mathbf{e}_j})\Big(\frac{\partial}{\partial x_j}X_T^{x+h\mathbf{e}_j}\Big)\Big] - \mathbb{E}\Big[(\nabla f)(X_T^x)\Big(\frac{\partial}{\partial x_j}X_T^x\Big)\Big]\Big| \geqslant 0. \end{aligned} \tag{2.80}$$

This proves that for all $j \in \{1, 2, \dots, d\}$ it holds that $(O \ni x \mapsto (\nabla f)(X_T^x)(\frac{\partial}{\partial x_j}X_T^x) \in L^1(\mathbb{P}; \mathbb{R})) \in C^0(O, L^1(\mathbb{P}; \mathbb{R}))$. This and (2.77) demonstrate that for all $j \in \{1, 2, \dots, d\}$, $x \in O$ it holds that $u \in C^1(O, \mathbb{R})$ and

$$\frac{\partial u}{\partial x_j}(x) = \mathbb{E}\Big[(\nabla f)(X_T^x)\Big(\frac{\partial}{\partial x_j}X_T^x\Big)\Big]. \tag{2.81}$$

Next observe that [60, Corollary 3.5] (applied with $T \curvearrowleft T-t$, $p \curvearrowleft 2$, $m \curvearrowleft d$, $\theta \curvearrowleft (\Omega \ni \omega \mapsto x \in O)$, $b \curvearrowleft ([0,T-t]\times\Omega\times O \ni (s,\omega,x) \mapsto \mu(t+s,x) \in \mathbb{R}^d)$, $\sigma \curvearrowleft ([0,T-t]\times\Omega\times O \ni (s,\omega,x) \mapsto \sigma(t+s,x) \in \mathbb{R}^{d\times d})$ in the notation of [60, Corollary 3.5]), the assumption that $\mu \in C^{0,1}([0,T]\times O,\mathbb{R}^d)$, $\sigma \in C^{0,1}([0,T]\times O,\mathbb{R}^{d\times d})$, and (2.66) ensure that for all $x \in O$ it holds that $X^x \in \mathbb{D}^{1,2}$. In addition, note that Lemma 2.2.1 (applied for every $x \in O$ with $m \curvearrowleft d$, $T \curvearrowleft T-t$, $X^x \curvearrowleft ([0,T-t]\times\Omega \ni (s,\omega) \mapsto X^x_{t+s} \in O)$, in the notation of Lemma 2.2.1) shows that it holds a.s. for all $t \in [0,T]$, $x \in O$ that $\frac{\partial}{\partial x}X^x_t$ is invertible and it holds a.s. for all $x \in O$ and Lebesgue almost all $t \in [0,T]$, $s \in [t,T]$ that

$$D_t X^x_s = \left(\frac{\partial}{\partial x}X^x_s\right)\left(\frac{\partial}{\partial x}X^x_t\right)^{-1}\sigma(t,X^x_t). \tag{2.82}$$

This, [86, Proposition 1.2.2] (applied for every $x \in O$ with $m \curvearrowleft d$, $\varphi \curvearrowleft f$, $p \curvearrowleft 2$, $F \curvearrowleft X^x$ in the notation of [86, Proposition 1.2.2]), and the assumption that $f \in C^\infty_c(O,\mathbb{R})$ demonstrate that for all $x \in O$ it holds that $f(X^x) \in \mathbb{D}^{1,2}$ and it holds a.s. for Lebesgue almost all $r \in [t,T]$, $s \in [t,r]$ that

$$D_s\left(f(X^x_r)\right) = (\nabla f)(X^x_r)D_sX^x_r = (\nabla f)(X^x_r)\Big(\frac{\partial}{\partial x}X^x_r\Big)\Big(\frac{\partial}{\partial x}X^x_s\Big)^{-1}\sigma(s,X^x_s). \tag{2.83}$$

Next note that the assumption that for all $s \in [t,T]$, $x \in O$, $v \in \mathbb{R}^d$ it holds that $v^*\sigma(s,x)(\sigma(s,x))^*v \geqslant \alpha\|v\|^2$ ensures that for all $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that $\sigma(s,x)(\sigma(s,x))^*$ is positive-definite and therefore invertible. In addition, observe that the fact that $\sigma \in C([0,T]\times O,\mathbb{R}^{d\times d})$ implies that $\sigma^{-1} \in C([0,T]\times O,\mathbb{R}^{d\times d})$. Integrating both sides of (2.83) hence implies that for all $r \in (t,T]$, $x \in O$ it holds a.s. that

$$(\nabla f)(X^x_r)\Big(\frac{\partial}{\partial x}X^x_r\Big) = \frac{1}{r-t}\int_t^r D_s\left(f(X^x_r)\right)(\sigma(s,X^x_s))^{-1}\Big(\frac{\partial}{\partial x}X^x_s\Big)\,ds. \tag{2.84}$$

The assumption that for every $x \in O$ it holds that X^x is an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted process therefore shows that for all $x \in O$ it holds that $((\sigma(s,X^x_s))^{-1}(\frac{\partial}{\partial x}X^x_s))_{s\in[t,T]}$ is an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted process. Combining this with [86, Proposition 1.3.4] (applied for every $j \in \{1,2,\dots,d\}$ with $u \curvearrowleft ([0,T]\times\Omega \ni (s,\omega) \mapsto (\sigma(t+s(T-t),X^x_{t+s(T-t)}))^{-1}(\frac{\partial}{\partial x}X^x_{t+s(T-t)})\mathbf{e}_j \in \mathbb{R}^d)$ in the notation of [86, Proposition 1.3.4]) and (2.72) demonstrates that for all $r \in [t,T]$, $x \in O$ it holds a.s. that

$$\int_t^r (\sigma(s,X^x_s))^{-1}\Big(\frac{\partial}{\partial x}X^x_s\Big)\,\delta W_s = \int_t^r (\sigma(s,X^x_s))^{-1}\Big(\frac{\partial}{\partial x}X^x_s\Big)\,dW_s. \tag{2.85}$$

The duality property of the Skorohod integral (cf., e.g., [86, Definition 1.3.1 (ii)] (applied with $T \curvearrowleft T-t$, $F \curvearrowleft (\Omega \ni \omega \mapsto f(X^x_T(\omega)) \in \mathbb{R}^d)$, $u \curvearrowleft ([0,T-t]\times\Omega \ni (s,\omega) \mapsto (\sigma(t+s,X^x_{t+s}))^{-1}(\frac{\partial}{\partial x}X^x_{t+s})))$) therefore shows that for all $x \in O$ it holds that

$$\begin{aligned}\mathbb{E}&\left[\int_t^T D_s\left(f(X^x_T)\right)(\sigma(s,X^x_s))^{-1}\Big(\frac{\partial}{\partial x}X^x_s\Big)\,ds\right]\\ &= \mathbb{E}\left[f(X^x_T)\int_t^T (\sigma(s,X^x_s))^{-1}\Big(\frac{\partial}{\partial x}X^x_s\Big)\,dW_s\right].\end{aligned} \tag{2.86}$$

This, (2.69), and (2.84) ensure that for all $x \in O$ it holds that

$$
\begin{aligned}
&\mathbb{E}\left[(\nabla f)(X_T^x)\left(\frac{\partial}{\partial x}X_T^x\right)\right] \\
&= \frac{1}{T-t}\mathbb{E}\left[\int_t^T D_s\left(f(X_T^x)\right)\left(\sigma(s,X_s^x)\right)^{-1}\left(\frac{\partial}{\partial x}X_s^x\right)\mathrm{d}s\right] \\
&= \frac{1}{T-t}\mathbb{E}\left[f(X_T^x)\int_t^T(\sigma(s,X_s^x))^{-1}\left(\frac{\partial}{\partial x}X_s^x\right)\mathrm{d}W_s\right] = \mathbb{E}\left[f(X_T^x)Z_T^x\right].
\end{aligned}
\tag{2.87}
$$

Combining this with (2.81) proves that $u \in C^1(O,\mathbb{R})$ and for all $x \in O$ it holds that $(\nabla u)(x) = \mathbb{E}[f(X_T^x)Z_T^x]$.

Step 2: In addition to the assumptions of Theorem 2.2.3 we assume in step 2 that $f \in C_c(O,\mathbb{R})$ with $\operatorname{supp}(f) \subsetneq O$. First note that the assumption that $f \in C_c(O,\mathbb{R})$ ensures that $f \in L^1(O;\mathbb{R})$. Let $\tilde{f} \in C(\mathbb{R}^d,\mathbb{R})$ satisfy for all $x \in O$ that $\tilde{f}(x) = f(x)$ and for all $x \in \mathbb{R}^d\backslash O$ that $f(x) = 0$, for every $\varepsilon \in (0,\infty)$ let $\varphi_\varepsilon \in C_c^\infty(\mathbb{R}^d,[0,\infty))$ satisfy $\operatorname{supp}(\varphi_\varepsilon) = \{x \in \mathbb{R}^d \colon \|x\| \leqslant \varepsilon\}$ and $\int_{\mathbb{R}^d}\varphi_\varepsilon(x)\,\mathrm{d}x = 1$ (cf. [96, Section 1.1]), for every $\varepsilon \in (0,\infty)$ let $f_\varepsilon \colon \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$ that

$$
f_\varepsilon(x) = \int_{\mathbb{R}^d}\tilde{f}(x-y)\varphi_\varepsilon(y)\,\mathrm{d}y,
\tag{2.88}
$$

and for every $\varepsilon \in (0,\infty)$ let $u_\varepsilon \colon O \to \mathbb{R}$ satisfy for all $x \in O$ that $u_\varepsilon(x) = \mathbb{E}[f_\varepsilon(X_T^x)]$. Combining this with [96, Lemma II.1] and the fact that $f \in L^1(O;\mathbb{R})$ implies that for all $\varepsilon \in (0,\infty)$ it holds that $f_\varepsilon \in C_c^\infty(\mathbb{R}^d,\mathbb{R})$. Step 1 therefore demonstrates that for all $\varepsilon \in (0,\infty)$, $x \in O$ it holds that $u_\varepsilon \in C^1(O,\mathbb{R})$ and $(\nabla u_\varepsilon)(x) = \mathbb{E}[f_\varepsilon(X_T^x)Z_T^x]$. Moreover, note that (2.88), the assumption that for all $\varepsilon \in (0,\infty)$ it holds that $\int_{\mathbb{R}^d}\varphi_\varepsilon(x)\,\mathrm{d}x = 1$, and the assumption that $f \in C_c(O,\mathbb{R})$ guarantuee that for all $\varepsilon \in (0,\infty)$, $x \in O$ it holds that

$$
|f_\varepsilon(X_T^x)| \leqslant \int_{\mathbb{R}^d}|\tilde{f}(X_T^x - y)|\varphi_\varepsilon(y)\,\mathrm{d}y \leqslant \sup_{y\in O}|f(y)| < \infty.
\tag{2.89}
$$

The triangle inequality, the dominated convergence theorem, [96, Lemma II.2], and the assumption that $f \in C_c(O,\mathbb{R})$ therefore show that for all $x \in O$ it holds that

$$
\begin{aligned}
&\limsup_{\varepsilon\to 0}|u_\varepsilon(x) - u(x)| = \limsup_{\varepsilon\to 0}\left|\mathbb{E}[f_\varepsilon(X_T^x)] - \mathbb{E}[f(X_T^x)]\right| \\
&\leqslant \limsup_{\varepsilon\to 0}\mathbb{E}[|f_\varepsilon(X_T^x) - f(X_T^x)|] = \mathbb{E}\left[\limsup_{\varepsilon\to 0}|f_\varepsilon(X_T^x) - f(X_T^x)|\right] = 0.
\end{aligned}
\tag{2.90}
$$

In addition, observe that the triangle inequality, the Cauchy-Schwarz inequality, and the fact that for all $\varepsilon \in (0,\infty)$, $x \in O$ it holds that $(\nabla u_\varepsilon)(x) = \mathbb{E}[f_\varepsilon(X_T^x)Z_T^x]$ prove that for all $\varepsilon \in (0,\infty)$, $x \in O$ it holds that

$$
\begin{aligned}
&\left\|(\nabla u_\varepsilon)(x) - \mathbb{E}[f(X_T^x)Z_T^x]\right\| = \left\|\mathbb{E}[f_\varepsilon(X_T^x)Z_T^x] - \mathbb{E}[f(X_T^x)Z_T^x]\right\| \\
&\leqslant \mathbb{E}[|f_\varepsilon(X_T^x) - f(X_T^x)|\|Z_T^x\|] \leqslant \left(\mathbb{E}[|f_\varepsilon(X_T^x) - f(X_T^x)|^2]\right)^{\frac{1}{2}}\left(\mathbb{E}[\|Z_T^x\|^2]\right)^{\frac{1}{2}}.
\end{aligned}
\tag{2.91}
$$

This, [96, Lemma II.2], (2.72), the assumption that $f \in C(O,\mathbb{R})$, and the fact that $(f_\varepsilon)_{\varepsilon\in(0,\infty)} \subseteq C^\infty(O,\mathbb{R})$ ensure that for all compact $K \subseteq O$ it holds that

$$
\limsup_{\varepsilon\to 0}\sup_{x\in K}\left\|(\nabla u_\varepsilon)(x) - \mathbb{E}[f(X_T^x)Z_T^x]\right\| = 0.
\tag{2.92}
$$

Hence, we obtain that u_ε converges uniformly to $(O \ni x \mapsto \mathbb{E}[f(X_T^x)Z_T^x] \in \mathbb{R}^d)$ on compact subsets of O for $\varepsilon \to 0$. Combining this with (2.90) and Lemma 2.2.2 proves that for all $x \in O$ it holds that $u \in C^1(O,\mathbb{R})$ and

$$(\nabla u)(x) = \lim_{\varepsilon\to 0}(\nabla u_\varepsilon)(x) = \lim_{\varepsilon\to 0}\mathbb{E}[f_\varepsilon(X_T^x)Z_T^x] = \mathbb{E}[f(X_T^x)Z_T^x]. \tag{2.93}$$

Step 3: Throughout the third step let $O_n \subseteq O$, $n \in \mathbb{N}$, satisfy for all $n \in \mathbb{N}$ that $O_n = \{x \in O\colon \|x\| \leqslant n \text{ and } \{y \in \mathbb{R}^d\colon \|y-x\| < \frac{1}{n}\} \subseteq O\}$ and let $U_n \subseteq O$ satisfy for all $n \in \mathbb{N}$ that $U_n = \{x \in O\colon (\exists y \in O_n\colon \|x-y\| \leqslant \frac{1}{2n})\}$. Observe that for all $n \in \mathbb{N}$ it holds that $O_n \subseteq O$ is compact, $U_n \subseteq O$ is open, and $O_n \subsetneq U_n \subseteq O$. Combining this with Urysohn's lemma (cf., e.g., [91, Lemma 2.12]) demonstrates that for every $n \in \mathbb{N}$ there exists $\psi_n \in C_c(O,\mathbb{R})$ which satisfies for all $x \in O$ that $\mathbb{1}_{O_n}(x) \leqslant \psi_n(x) \leqslant \mathbb{1}_{U_n}(x)$. For every $n \in \mathbb{N}$ let $\mathfrak{f}_n\colon O \to \mathbb{R}$, $\mathfrak{u}_n\colon O \to \mathbb{R}$ satisfy for all $x \in O$ that $\mathfrak{f}_n(x) = f(x)\psi_n(x)$ and $\mathfrak{u}_n(x) = \mathbb{E}[\mathfrak{f}_n(X_T^x)]$. Note that the fact that for all $n \in \mathbb{N}$ it holds that $\psi_n \in C_c(O,\mathbb{R})$ and $f \in C(O,\mathbb{R})$ ensure that for all $n \in \mathbb{N}$ it holds that $\mathfrak{f}_n \in C_c(O,\mathbb{R})$. Moreover, observe that for all $n \in \mathbb{N}$, $x \in O$ it holds that $|\mathfrak{f}_n(X_T^x)| = |f(X_T^x)\psi_n(X_T^x)| \leqslant |f(X_T^x)\mathbb{1}_{U_n}(X_T^x)| \leqslant |f(X_T^x)|$. The dominated convergence theorem and the assumption that for all $x \in O$ it holds that $\mathbb{E}[|f(X_T^x)|^2] < \infty$ therefore demonstrate that for all $x \in O$ it holds that

$$\begin{aligned}
&\limsup_{n\to\infty}|\mathfrak{u}_n(x) - u(x)| \leqslant \limsup_{n\to\infty}\mathbb{E}\left[|\mathfrak{f}_n(X_T^x) - f(X_T^x)|\right]\\
&= \mathbb{E}\left[\limsup_{n\to\infty}|\mathfrak{f}_n(X_T^x) - f(X_T^x)|\right] \leqslant \mathbb{E}\left[\limsup_{n\to\infty}|\mathbb{1}_{U_n}(X_T^x) - 1||f(X_T^x)|\right] = 0.
\end{aligned} \tag{2.94}$$

Furthermore, observe that step 2 and the fact that for every $n \in \mathbb{N}$ it holds that $\mathfrak{f}_n \in C_c(O,\mathbb{R})$ and $\operatorname{supp}(\mathfrak{f}_n) \subsetneq O$ imply that for all $n \in \mathbb{N}$, $x \in O$ it holds that $\mathfrak{u}_n \in C^1(O,\mathbb{R})$ and $(\nabla \mathfrak{u}_n)(x) = \mathbb{E}[\mathfrak{f}_n(X_T^x)Z_T^x]$. Combining this with the Cauchy-Schwarz inequality implies that for all $n \in \mathbb{N}$, $x \in O$ it holds that

$$\begin{aligned}
&\left\|(\nabla \mathfrak{u}_n)(x) - \mathbb{E}[f(X_T^x)Z_T^x]\right\| = \left\|\mathbb{E}[\mathfrak{f}_n(X_T^x)Z_T^x] - \mathbb{E}[f(X_T^x)Z_T^x]\right\|\\
&\leqslant \mathbb{E}\left[|\mathbb{1}_{U_n}(X_T^x) - 1||f(X_T^x)|\|Z_T^x\|\right]\\
&\leqslant \left(\mathbb{E}\left[|\mathbb{1}_{U_n}(X_T^x) - 1||f(X_T^x)|^2\right]\right)^{\frac{1}{2}}\left(\mathbb{E}\left[\|Z_T^x\|^2\right]\right)^{\frac{1}{2}}.
\end{aligned} \tag{2.95}$$

This, Dini's theorem, Lebesgue's dominated convergence theorem, (2.72), and the fact that for all $x \in O$ it holds that $\mathbb{E}[|f(X_T^x)|^2] < \infty$ show that for all $K \subseteq O$ compact it holds that

$$\limsup_{n\to\infty}\left[\sup_{x\in K}\left\|(\nabla \mathfrak{u}_n)(x) - \mathbb{E}[f(X_T^x)Z_T^x]\right\|\right] = 0. \tag{2.96}$$

Hence, we obtain that $\mathfrak{u}_n$ converges uniformly to $(O \ni x \mapsto \mathbb{E}[f(X_T^x)Z_T^x] \in \mathbb{R}^d)$ on compact subsets of O for $n \to \infty$. Lemma 2.2.2, (2.94), and (2.96) therefore prove that $u \in C^1(O,\mathbb{R})$ and for all $x \in O$ it holds that

$$(\nabla u)(x) = \mathbb{E}[f(X_T^x)Z_T^x]. \tag{2.97}$$

This establishes items (ii) and (iii). The proof of Theorem 2.2.3 is thus complete. □

2.3 Existence and uniqueness result for viscosity solutions of semilinear PDEs with gradient-dependent nonlinearities

In this section we use the results from Section 2.1 and 2.2 to show that the unique viscosity solution of semilinear PDEs and the unique solution of their connected SFPEs coincide.

Our main theorem, Theorem 2.3.1, proves exactly this connection under differentiability and global monotonicity assumptions on μ and σ and a Lipschitz and continuity assumption on f. We conclude this section with two corollaries of Theorem 2.3.1. Corollary 2.3.2 considers the case where the Lyapunov function does not depend on the time variable. Corollary 2.3.3 uses Corollary 2.3.2 to establish the findings in the case of a Lyapunov function of the form $\mathbb{R}^d \ni x \mapsto (1+\|x\|^2)^{\frac{p}{2}}$ where $p \in (0,\infty)$. The following theorem, Theorem 2.3.1, extends [10, Theorem 3.7] to PDEs with gradient-dependent nonlinearities under more restrictive assumptions.

Theorem 2.3.1. *Let $d \in \mathbb{N}$, $\alpha, c, L, T \in (0,\infty)$, let $\langle\cdot,\cdot\rangle\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the standard Euclidean scalar product on $\mathbb{R}^d$, let $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, let $|||\cdot|||\colon \mathbb{R}^{d+1} \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^{d+1}$, let $\|\cdot\|_F\colon \mathbb{R}^{d\times d} \to [0,\infty)$ be the Frobenius norm on $\mathbb{R}^{d\times d}$, let $O \subseteq \mathbb{R}^d$ be an open set, for every $r \in (0,\infty)$ let $K_r \subseteq [0,T)$, $O_r \subseteq O$ satisfy $K_r = [0,\max\{T-\frac{1}{r},0\}]$ and $O_r = \{x \in O\colon \|x\| \leqslant r$ and $\{y \in \mathbb{R}^d\colon \|y-x\| < \frac{1}{r}\} \subseteq O\}$, let $(\Omega,\mathcal{F},\mathbb{P},(\mathbb{F}_s)_{s\in[0,T]})$ be a filtered probability space satisfying the usual conditions, let $W\colon [0,T]\times\Omega \to \mathbb{R}^d$ be a standard $(\mathbb{F}_s)_{s\in[0,T]}$-Brownian motion, let $\mu \in C^{0,1}([0,T]\times O,\mathbb{R}^d)$, $\sigma \in C^{0,1}([0,T]\times O,\mathbb{R}^{d\times d})$ satisfy for all $s \in [0,T]$, $x,y \in O$, $v \in \mathbb{R}^d$ that*

$$\max\left\{\langle x-y,\mu(s,x)-\mu(s,y)\rangle, \tfrac{1}{2}\|\sigma(s,x)-\sigma(s,y)\|_F^2\right\} \leqslant \tfrac{c}{2}\|x-y\|^2 \tag{2.98}$$

and $v^*\sigma(s,x)(\sigma(s,x))^*v \geqslant \alpha\|v\|^2$, assume for all $r \in (0,\infty)$, $j \in \{1,2,\ldots,d\}$ that

$$\sup\left(\left\{\frac{\|\frac{\partial\mu}{\partial x}(t,x)-\frac{\partial\mu}{\partial x}(t,y)\|_F+\|\frac{\partial\sigma}{\partial x_j}(t,x)-\frac{\partial\sigma}{\partial x_j}(t,y)\|_F}{\|x-y\|}\colon\right.\right.$$
$$\left.\left. t\in[0,T], x,y\in O_r, x\neq y\right\}\cup\{0\}\right) < \infty, \tag{2.99}$$

for every $t \in [0,T]$, $x \in O$ let $X_t^x = (X_{t,s}^x)_{s\in[t,T]}\colon [t,T]\times\Omega \to O$ be an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in [t,T]$ it holds a.s. that

$$X_{t,s}^x = x + \int_t^s \mu(r,X_{t,r}^x)\,dr + \int_t^s \sigma(r,X_{t,r}^x)\,dW_r, \tag{2.100}$$

assume for all $t \in [0,T]$, $\omega \in \Omega$ that $\big([t,T]\times O \ni (s,x) \mapsto X_{t,s}^x(\omega) \in O\big) \in C^{0,1}([t,T]\times O,O)$, for every $t \in [0,T]$, $x \in O$ let $Z_t^x = (Z_{t,s}^x)_{s\in(t,T]}\colon (t,T]\times\Omega \to \mathbb{R}^{d+1}$ be an $(\mathbb{F}_s)_{s\in(t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in (t,T]$ it holds a.s. that

$$Z_{t,s}^x = \begin{pmatrix} 1 \\ \frac{1}{s-t}\int_t^s (\sigma(r,X_{t,r}^x))^{-1}\left(\frac{\partial}{\partial x}X_{t,r}^x\right)dW_r \end{pmatrix}, \tag{2.101}$$

let $V \in C^{1,2}([0,T]\times O,(0,\infty))$ satisfy that for all $t \in [0,T]$, $s \in [t,T]$, $x \in O$ it holds that

$$\begin{aligned}(\tfrac{\partial V}{\partial t})(t,x) + \langle\mu(t,x),(\nabla_x V)(t,x)\rangle + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x V)(t,x))\\ + L\|(\nabla_x V)(t,x)\| \leqslant 0,\end{aligned} \tag{2.102}$$

let $f \in C([0,T]\times O\times\mathbb{R}^{d+1},\mathbb{R})$, $g \in C(O,\mathbb{R})$ satisfy for all $t \in [0,T]$, $s \in [t,T]$, $x \in O$, $w_1,w_2 \in \mathbb{R}^{d+1}$ that $\mathbb{E}[|g(X_{t,T}^x)|^2] + \mathbb{E}[|f(s,X_{t,s}^x,w_1)|^2] < \infty$ and $|f(t,x,w_1)-f(t,x,w_2)| \leqslant L|||w_1-w_2|||$, and assume that $\limsup_{r\in(0,\infty)}[\sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}(\frac{|g(x)|}{\sqrt{V(T,x)}}+\frac{|f(t,x,0)|}{\sqrt{V(t,x)}}\sqrt{T-t})]$ $= 0$ and $\liminf_{r\to\infty}[\inf_{t\in[0,T]}\inf_{x\in O\setminus O_r} V(t,x)] = \infty$. Then

(i) *there exists a unique* $v \in C([0,T] \times O, \mathbb{R}) \cap C^{0,1}([0,T) \times O, \mathbb{R})$ *which satisfies for all* $t \in [0,T)$, $x \in O$ *that* $\limsup_{r\to\infty}[\sup_{s\in[0,T)\setminus K_r} \sup_{y\in O\setminus O_r}(\frac{\|(v,\nabla_x v)(s,y)\|}{\sqrt{V(s,y)}}\sqrt{T-s})] = 0$, $\mathbb{E}[|g(X^x_{t,T})| \, \|Z^x_{t,T}\| + \int_t^T |f(r, X^x_{t,r}, (v,\nabla_x v)(r, X^x_{t,r}))| \|Z^x_{t,r}\| \, dr] < \infty$, $v(T,x) = g(x)$, *and*

$$(v, \nabla_x v)(t,x) = \mathbb{E}\left[g(X^x_{t,T})Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, (v,\nabla_x v)(r, X^x_{t,r}))Z^x_{t,r}\, dr\right], \quad (2.103)$$

(ii) *there exists a unique viscosity solution* $u \in \{\mathbf{u} \in C([0,T] \times O, \mathbb{R}) \colon \limsup_{r\to\infty} [\sup_{t\in[0,T]} \sup_{x\in O\setminus O_r}(\frac{|\mathbf{u}(t,x)|}{V(t,x)})] = 0\}$ *of*

$$\begin{aligned}&(\tfrac{\partial u}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x u)(t,x)\rangle \\ &\quad + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x u)(t,x)) + f(t,x,(u,\nabla_x u)(t,x)) = 0\end{aligned} \quad (2.104)$$

with $u(T,x) = g(x)$ *for* $(t,x) \in (0,T) \times O$, *and*

(iii) *for all* $t \in [0,T]$, $x \in O$ *it holds that* $u(t,x) = v(t,x)$.

Proof of Theorem 2.3.1. First note that Theorem 1.2.5, (2.99), and (2.102) prove that there exists a unique $w = (w_1, w_2, \ldots, w_{d+1}) \in C([0,T) \times O, \mathbb{R}^{d+1})$ which satisfies

(I) that $\limsup_{r\to\infty} [\sup_{s\in[0,T)\setminus K_r} \sup_{y\in O\setminus O_r}(\frac{\|w(s,y)\|}{\sqrt{V(s,y)}}\sqrt{T-s})] = 0$,

(II) for all $t \in [0,T)$, $x \in O$ that

$$\mathbb{E}\left[|g(X^x_{t,T})| \, \|Z^x_{t,T}\| + \int_t^T |f(r, X^x_{t,r}, w(r, X^x_{t,r}))| \, \|Z^x_{t,r}\| \, dr\right] < \infty, \quad (2.105)$$

and

(III) for all $t \in [0,T)$, $x \in O$ it holds that

$$w(t,x) = \mathbb{E}\left[g(X^x_{t,T})Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, w(r, X^x_{t,r}))Z^x_{t,r}\, dr\right]. \quad (2.106)$$

Let $v \colon [0,T] \times O \to \mathbb{R}$ satisfy for all $t \in [0,T)$, $x \in O$ that $v(t,x) = w_1(t,x)$ and $v(T,x) = g(x)$, let $h \colon [0,T] \times O \to \mathbb{R}$ satisfy for all $t \in [0,T)$, $x \in O$ that $h(t,x) = f(t,x,(v,(\nabla_x v))(t,x))$, let $\mathfrak{g}_n \in C(O,\mathbb{R})$, $n \in \mathbb{N}$, and $\mathfrak{h}_n \in C([0,T) \times O, \mathbb{R})$, $n \in \mathbb{N}$, be compactly supported functions which satisfy

$$\limsup_{n\to\infty}\left[\sup_{t\in[0,T)} \sup_{x\in O}\left(\frac{|\mathfrak{g}_n(x) - g(x)|}{V(T,x)} + \frac{|\mathfrak{h}_n(t,x) - h(t,x)|}{V(t,x)}\sqrt{T-t}\right)\right] = 0 \quad (2.107)$$

(cf. Corollary 1.1.6) and let $\mathfrak{v}_n \colon [0,T] \times O \to \mathbb{R}$, $n \in \mathbb{N}$, satisfy for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O$ that

$$\mathfrak{v}_n(t,x) = \mathbb{E}\left[\mathfrak{g}_n(X^x_{t,T}) + \int_t^T \mathfrak{h}_n(r, X^x_{t,r})\, dr\right] \quad (2.108)$$

(cf. item (ii) of Lemma 1.1.4). Observe that the fact that $w \in C([0,T) \times O, \mathbb{R}^{d+1})$ implies that $v \in C([0,T) \times O, \mathbb{R})$. To prove that v is continuous in T let $(t_n)_{n\in\mathbb{N}} \subseteq [0,T)$ satisfy $\limsup_{n\to\infty}|t_n - T| = 0$. Note that [13, Lemma 3.1] and (2.102) imply that for all $t \in [0,T]$, $s \in [t,T]$, $x \in O$ it holds that

$$E[V(s, X^x_{t,s})] \leqslant V(t,x). \quad (2.109)$$

Combining this with Fubini's theorem and the assumption that for all $t \in [0,T]$, $x \in O$, $z_1, z_2 \in \mathbb{R}^{d+1}$ it holds that $|f(t,x,z_1) - f(t,x,z_2)| \leqslant L\|z_1 - z_2\|$ demonstrates that for all $n \in \mathbb{N}$, $x \in O$ it holds that

$$\begin{aligned}
&\mathbb{E}\bigg[\bigg|\int_{t_n}^T f(r, X^x_{t_n,r}, w(r, X^x_{t_n,r}))\,\mathrm{d}r\bigg|\bigg] \leqslant \mathbb{E}\bigg[\int_{t_n}^T |f(r, X^x_{t_n,r}, w(r, X^x_{t_n,r}))|\,\mathrm{d}r\bigg] \\
&\leqslant \mathbb{E}\bigg[\int_{t_n}^T \Big[|f(r, X^x_{t_n,r}, w(r, X^x_{t_n,r})) - f(r, X^x_{t_n,r}, 0)| + |f(r, X^x_{t_n,r}, 0)|\Big]\,\mathrm{d}r\bigg] \\
&\leqslant \mathbb{E}\bigg[\int_{t_n}^T \Big[L\big\|w(r, X^x_{t_n,r})\big\| + |f(r, X^x_{t_n,r}, 0)|\Big]\,\mathrm{d}r\bigg] \\
&= \mathbb{E}\bigg[\int_{t_n}^T \Big[\tfrac{L\|w(r,X^x_{t_n,r})\|+|f(r,X^x_{t_n,r},0,0)|}{V(r,X^x_{t_n,r})}\sqrt{T-r}\,\tfrac{V(r,X^x_{t_n,r})}{\sqrt{T-r}}\Big]\,\mathrm{d}r\bigg] \\
&\leqslant \bigg[\sup_{s\in[0,T)}\sup_{y\in O} \tfrac{L\|w(s,y)\|+|f(s,y,0)|}{V(s,y)}\sqrt{T-s}\bigg]\int_{t_n}^T \tfrac{E[V(r,X^x_{t_n,r})]}{\sqrt{T-r}}\,\mathrm{d}r \\
&\leqslant \bigg[\sup_{s\in[0,T)}\sup_{y\in O} \tfrac{L\|w(s,y)\|+|f(s,y,0)|}{V(s,y)}\sqrt{T-s}\bigg]\int_{t_n}^T \tfrac{V(t_n,x)}{\sqrt{T-r}}\,\mathrm{d}r \\
&= \bigg[\sup_{s\in[0,T)}\sup_{y\in O} \tfrac{L\|w(s,y)\|+|f(s,y,0)|}{V(s,y)}\sqrt{T-s}\bigg]\bigg[\sup_{s\in[0,T]} V(s,x)\bigg]2\sqrt{T-t_n}.
\end{aligned} \tag{2.110}$$

Item (I), the assumption that $\inf_{r\in(0,\infty)} [\sup_{t\in[0,T)\setminus K_r} \sup_{x\in O\setminus O_r}(\frac{|f(t,x,0)|}{\sqrt{V(t,x)}}\sqrt{T-t})] = 0$, and the fact that $V \in C^{1,2}([0,T]\times O,(0,\infty))$ therefore show that for all $x \in O$ it holds that

$$\limsup_{n\to\infty} \mathbb{E}\bigg[\bigg|\int_{t_n}^T f(r, X^x_{t_n,r}, w(r, X^x_{t_n,r}))\,\mathrm{d}r\bigg|\bigg] = 0. \tag{2.111}$$

In addition, note that the triangle inequality and (2.109) show that for all $k,n \in \mathbb{N}$, $x \in O$ it holds that

$$\begin{aligned}
&\mathbb{E}[|g(X^x_{t_n,T}) - g(x)|] \\
&\leqslant \mathbb{E}[|g(X^x_{t_n,T}) - \mathfrak{g}_k(X^x_{t_n,T})|] + \mathbb{E}[|\mathfrak{g}_k(X^x_{t_n,T}) - \mathfrak{g}_k(x)|] + \mathbb{E}[|\mathfrak{g}_k(x) - g(x)|] \\
&= \mathbb{E}\bigg[\frac{|g(X^x_{t_n,T}) - \mathfrak{g}_k(X^x_{t_n,T})|}{V(T, X^x_{t_n,T})}V(T, X^x_{t_n,T})\bigg] + \mathbb{E}[|\mathfrak{g}_k(X^x_{t_n,T}) - \mathfrak{g}_k(x)|] \\
&\quad + \mathbb{E}\bigg[\frac{|\mathfrak{g}_k(x) - g(x)|}{V(T,x)}V(T,x)\bigg] \\
&\leqslant \bigg[\sup_{y\in O}\frac{|g(y) - \mathfrak{g}_k(y)|}{V(T,y)}\bigg]\mathbb{E}[V(T, X^x_{t_n,T})] + \mathbb{E}[|\mathfrak{g}_k(X^x_{t_n,T}) - \mathfrak{g}_k(x)|] \\
&\quad + \bigg[\sup_{y\in O}\frac{|\mathfrak{g}_k(y) - g(y)|}{V(T,y)}\bigg]V(T,x) \\
&\leqslant \bigg[\sup_{y\in O}\frac{|g(y) - \mathfrak{g}_k(y)|}{V(T,y)}\bigg]V(t_n,x) + \mathbb{E}[|\mathfrak{g}_k(X^x_{t_n,T}) - \mathfrak{g}_k(x)|] \\
&\quad + \bigg[\sup_{y\in O}\frac{|\mathfrak{g}_k(y) - g(y)|}{V(T,y)}\bigg]V(T,x) \\
&\leqslant 2\bigg[\sup_{y\in O}\frac{|g(y) - \mathfrak{g}_k(y)|}{V(T,y)}\bigg]\bigg[\sup_{s\in[0,T]} V(s,x)\bigg] + \mathbb{E}[|\mathfrak{g}_k(X^x_{t_n,T}) - \mathfrak{g}_k(x)|].
\end{aligned} \tag{2.112}$$

This, [13, Lemma 3.7], the Portemonteau theorem, and (2.107) demonstrate for all $x \in O$ that

$$\limsup_{n\to\infty} \mathbb{E}[|g(X^x_{t_n,T}) - g(x)|] = 0. \tag{2.113}$$

Combining this with (2.111) proves that for all $x \in O$ it holds that

$$\limsup_{n\to\infty} |v(t_n, x) - g(x)| = 0. \tag{2.114}$$

The assumption that for all $x \in \mathbb{R}^d$ it holds that $v(T, x) = g(x)$ and the fact that $v \in C([0,T) \times O, \mathbb{R})$ hence demonstrate that $v \in C([0,T] \times O, \mathbb{R})$. Next note that items (ii) and (iii) of Theorem 2.2.3 (applied for all $t \in [0,T)$, $r \in (t,T]$ with $O \curvearrowleft \mathbb{R}^d$, $f \curvearrowleft g$ and $T \curvearrowleft r$, $f \curvearrowleft (\mathbb{R}^d \ni x \mapsto f(r, x, w(r,x)) \in \mathbb{R})$ in the notation of Theorem 2.2.3), Leibniz integral rule, Fubini's theorem, item (II), and the assumption that for all $t \in [0,T]$, $s \in [t,T]$, $x \in O$, $z \in \mathbb{R}^{d+1}$ it holds that $\mathbb{E}[|g(X^x_{t,T})|^2] + \mathbb{E}[|f(s, X^x_{t,s}, z)|^2] < \infty$ show that $v \in C^{0,1}([0,T) \times O, \mathbb{R})$ and for all $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
(\nabla_x v)(t,x) &= \nabla_x\Big(\mathbb{E}[g(X^x_{t,T})]\Big) + \nabla_x\bigg(\int_t^T \mathbb{E}[f(r, X^x_{t,r}, w(r, X^x_{t,r}))]\,\mathrm{d}r\bigg)\\
&= \nabla_x\Big(\mathbb{E}[g(X^x_{t,T})]\Big) + \int_t^T \nabla_x\Big(\mathbb{E}\left[f(r, X^x_{t,r}, w(r, X^x_{t,r}))\right]\Big)\,\mathrm{d}r\\
&= \mathbb{E}\left[g(X^x_{t,T})\frac{1}{T-t}\int_t^T (\sigma(s, X^x_{t,s}))^{-1}\left(\frac{\partial}{\partial x}X^x_{t,s}\right)\mathrm{d}W_s\right]\\
&\quad + \int_t^T \mathbb{E}\left[f(r, X^x_{t,r}, w(r, X^x_{t,r}))\frac{1}{r-t}\int_t^r (\sigma(s, X^x_{t,s}))^{-1}\left(\frac{\partial}{\partial x}X^x_{t,s}\right)\mathrm{d}W_s\right]\mathrm{d}r\\
&= \mathbb{E}[g(X^x_{t,T})\frac{1}{T-t}\int_t^T (\sigma(s, X^x_{t,s}))^{-1}\left(\frac{\partial}{\partial x}X^x_{t,s}\right)\mathrm{d}W_s]\\
&\quad + \mathbb{E}\bigg[\int_t^T f(r, X^x_{t,r}, w(r, X^x_{t,r}))\frac{1}{r-t}\int_t^r (\sigma(s, X^x_{t,s}))^{-1}\left(\frac{\partial}{\partial x}X^x_{t,s}\right)\mathrm{d}W_s\,\mathrm{d}r\bigg].
\end{aligned} \tag{2.115}$$

Item (III) therefore implies that for all $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that $(w_2, w_3, \ldots, w_{d+1})(t,x) = (\nabla_x v)(t,x)$. This, items (I)-(III), and the fact that $v \in C([0,T] \times O, \mathbb{R})$ establish item (i). Next we prove items (ii) and (iii). First note that item (I), the fact that for all $t \in [0,T]$, $x \in O$, $z_1, z_2 \in \mathbb{R}^{d+1}$ it holds that $|f(t,x,z_1) - f(t,x,z_2)| \leqslant L|\!|\!|z_1 - z_2|\!|\!|$, and the fact that $\limsup_{r\in(0,\infty)}[\sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}(\frac{|f(t,x,0)|}{V(t,x)}\sqrt{T-t})] = 0$ imply that $h \in C([0,T) \times O, \mathbb{R})$ and

$$\begin{aligned}
&\limsup_{r\to\infty}\left[\sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}\left(\tfrac{|h(t,x)|}{V(t,x)}\sqrt{T-t}\right)\right]\\
&= \limsup_{r\to\infty}\left[\sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}\left(\tfrac{|f(t,x,(v,\nabla_x v)(t,x))|}{V(t,x)}\sqrt{T-t}\right)\right]\\
&\leqslant \limsup_{r\to\infty}\left[\sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}\left(\tfrac{|f(t,x,0)|}{V(t,x)}\sqrt{T-t} + \tfrac{|f(t,x,(v,\nabla_x v)(t,x)) - f(t,x,0)|}{V(t,x)}\sqrt{T-t}\right)\right]\\
&\leqslant \limsup_{r\to\infty}\left[\sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}\left(\tfrac{|f(t,x,0)| + L|\!|\!|(v,\nabla_x v)(t,x)|\!|\!|}{V(t,x)}\sqrt{T-t}\right)\right] = 0.
\end{aligned} \tag{2.116}$$

Next note that (2.109) demonstrates that for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in O$ it holds that

$$\begin{aligned}
\mathbb{E}\left[|\mathfrak{g}_n(X^x_{t,T}) - g(X^x_{t,T})|\right] &\leqslant \left[\sup_{y\in O}\frac{|\mathfrak{g}_n(y) - g(y)|}{V(T,y)}\right]\mathbb{E}[V(T, X^x_{t,T})]\\
&\leqslant \left[\sup_{y\in O}\frac{|\mathfrak{g}_n(y) - g(y)|}{V(T,y)}\right]V(t,x).
\end{aligned} \tag{2.117}$$

Combining this with (2.107) implies that

$$\limsup_{n\to\infty}\left[\sup_{t\in[0,T]}\sup_{x\in O}\left(\frac{|\mathbb{E}[\mathfrak{g}_n(X^x_{t,T})]-\mathbb{E}[g(X^x_{t,T})]|}{V(t,x)}\right)\right]=0. \tag{2.118}$$

Moreover, observe that (2.109) shows that for all $n\in\mathbb{N}$, $t\in[0,T]$, $x\in O$ it holds that

$$\begin{aligned}&\mathbb{E}\left[\int_t^T|\mathfrak{h}_n(r,X^x_{t,r})-h(r,X^x_{t,r})|\,\mathrm{d}r\right]\\&\leqslant\left[\sup_{s\in[0,T)}\sup_{y\in O}\left(\frac{|\mathfrak{h}_n(s,y)-h(s,y)|}{V(s,y)}\sqrt{T-s}\right)\right]\int_t^T\frac{\mathbb{E}[V(r,X^x_{t,r})]}{\sqrt{T-r}}\,\mathrm{d}r\\&\leqslant\left[\sup_{s\in[0,T)}\sup_{y\in O}\left(\frac{|\mathfrak{h}_n(s,y)-h(s,y)|}{V(s,y)}\sqrt{T-s}\right)\right]2\sqrt{T-t}\,V(t,x).\end{aligned} \tag{2.119}$$

This and (2.107) imply that

$$\limsup_{n\to\infty}\left[\sup_{t\in[0,T]}\sup_{x\in O}\left(\frac{|\mathbb{E}[\int_t^T\mathfrak{h}_n(r,X^x_{t,r})\,\mathrm{d}r]-\mathbb{E}[\int_t^T h(r,X^x_{t,r})]\,\mathrm{d}r|}{V(t,x)}\right)\right]=0. \tag{2.120}$$

The triangle inequality and (2.118) therefore demonstrates that

$$\limsup_{n\to\infty}\left[\sup_{t\in[0,T]}\sup_{x\in O}\left(\frac{|\mathfrak{v}_n(t,x)-v(t,x)|}{V(t,x)}\right)\right]=0. \tag{2.121}$$

Furthermore, note that (2.108), the assumption that $\liminf_{r\to\infty}[\inf_{t\in[0,T]}\inf_{x\in O\backslash O_r}V(t,x)]=\infty$, and the fact that for all $n\in\mathbb{N}$ it holds that $\mathfrak{g}_n$ and $\mathfrak{h}_n$ are compactly supported ensure for all $n\in\mathbb{N}$ that

$$\begin{aligned}&\limsup_{r\to\infty}\left[\sup_{t\in[0,T]}\sup_{x\in O\backslash O_r}\left(\frac{|\mathfrak{v}_n(t,x)|}{V(t,x)}\right)\right]\\&\leqslant\limsup_{r\to\infty}\left(\left[\left(\sup_{x\in O}|\mathfrak{g}_n(x)|\right)\right.\right.\\&\qquad\left.\left.+T\left(\sup_{t\in[0,T]}\sup_{x\in O}|\mathfrak{h}_n(t,x)|\right)\right]\left[\sup_{t\in[0,T]}\sup_{x\in O\backslash O_r}\frac{1}{V(t,x)}\right]\right)=0.\end{aligned} \tag{2.122}$$

Combining this with the triangle inequality and (2.121) proves that

$$\begin{aligned}&\limsup_{r\to\infty}\left[\sup_{t\in[0,T]}\sup_{x\in O\backslash O_r}\frac{|v(t,x)|}{V(t,x)}\right]\\&\leqslant\inf_{n\in\mathbb{N}}\left(\limsup_{r\to\infty}\left[\sup_{t\in[0,T]}\sup_{x\in O\backslash O_r}\frac{|v(t,x)-\mathfrak{v}(t,x)|}{V(t,x)}\right]\right.\\&\qquad\left.+\limsup_{r\to\infty}\left[\sup_{t\in[0,T]}\sup_{x\in O\backslash O_r}\frac{|\mathfrak{v}(t,x)|}{V(t,x)}\right]\right)\\&\leqslant\inf_{n\in\mathbb{N}}\left(\sup_{t\in[0,T]}\sup_{x\in O}\frac{|v(t,x)-\mathfrak{v}(t,x)|}{V(t,x)}\right)\\&\leqslant\limsup_{n\to\infty}\left(\sup_{t\in[0,T]}\sup_{x\in O}\frac{|v(t,x)-\mathfrak{v}(t,x)|}{V(t,x)}\right)=0.\end{aligned} \tag{2.123}$$

Next note that Proposition 2.1.7, (2.99), (2.102), (2.103), and (2.116) demonstrate that v is a viscosity solution of

$$\begin{aligned}(\tfrac{\partial v}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x v)(t,x)\rangle \\ + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x v)(t,x)) + h(t,x) = 0\end{aligned} \tag{2.124}$$

for $(t,x) \in (0,T)\times O$. This ensures that for all $t \in (0,T)$, $x \in O$, $\phi \in C^{1,2}((0,T)\times O, \mathbb{R})$ with $\phi \geqslant v$ and $\phi(t,x) = v(t,x)$ it holds that

$$\begin{aligned}&(\tfrac{\partial \phi}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x \phi)(t,x)\rangle \\ &\quad + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x \phi)(t,x)) + f(t,x,\phi(t,x),(\nabla_x\phi)(t,x)) \\ &= (\tfrac{\partial \phi}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x \phi)(t,x)\rangle \\ &\quad + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x \phi)(t,x)) + h(t,x) \geqslant 0.\end{aligned} \tag{2.125}$$

Moreover, observe that (2.124) shows that for all $t \in (0,T)$, $x \in O$, $\phi \in C^{1,2}((0,T)\times O, \mathbb{R})$ with $\phi \leqslant v$ and $\phi(t,x) = v(t,x)$ it holds that

$$\begin{aligned}&(\tfrac{\partial \phi}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x \phi)(t,x)\rangle \\ &\quad + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x \phi)(t,x)) + f(t,x,\phi(t,x),(\nabla_x\phi)(t,x)) \\ &= (\tfrac{\partial \phi}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x \phi)(t,x)\rangle \\ &\quad + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x \phi)(t,x)) + h(t,x) \leqslant 0.\end{aligned} \tag{2.126}$$

Combining this with (2.125) proves that v is a viscosity solution of

$$\begin{aligned}&(\tfrac{\partial v}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x v)(t,x)\rangle \\ &\quad + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x v)(t,x)) + f(t,x,(v,\nabla_x v)(t,x)) = 0\end{aligned} \tag{2.127}$$

for $(t,x) \in (0,T)\times O$. Proposition 2.1.8, (applied with $u_1 \curvearrowleft v$ in the notation of Proposition 2.1.8), (2.102), and (2.123) therefore establish items (ii) and (iii). The proof of Theorem 2.3.1 is thus complete. □

The following corollary applies the results of Theorem 2.3.1 to a function V that is independent of the time component. The proof of the Corollary 2.3.2 is similar to the one of [10, Corollary 3.8].

Corollary 2.3.2. *Let $d \in \mathbb{N}$, $\alpha, c, L, T \in (0,\infty)$, $\rho \in \mathbb{R}$, let $\langle\cdot,\cdot\rangle\colon \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}$ be the standard Euclidean scalar product on $\mathbb{R}^d$, let $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, let $|||\cdot|||\colon \mathbb{R}^{d+1} \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^{d+1}$, let $\|\cdot\|_F\colon \mathbb{R}^{d\times d} \to [0,\infty)$ be the Frobenius norm on $\mathbb{R}^{d\times d}$, let $O \subseteq \mathbb{R}^d$ be an open set, for every $r \in (0,\infty)$ let $K_r \subseteq [0,T)$, $O_r \subseteq O$ satisfy $K_r = [0,\max\{T-\frac{1}{r},0\}]$ and $O_r = \{x \in O\colon \|x\| \leqslant r \text{ and } \{y\in\mathbb{R}^d\colon \|y-x\| < \frac{1}{r}\} \subseteq O\}$, let $(\Omega,\mathcal{F},\mathbb{P},(\mathbb{F}_s)_{s\in[0,T]})$ be a filtered probability space satisfying the usual conditions, let $W\colon [0,T]\times\Omega \to \mathbb{R}^d$ be a standard $(\mathbb{F}_s)_{s\in[0,T]}$-Brownian motion, let $\mu \in C^{0,1}([0,T]\times O, \mathbb{R}^d)$, $\sigma \in C^{0,1}([0,T]\times O,\mathbb{R}^{d\times d})$ satisfy for all $s \in [0,T]$, $x, y \in O$, $v \in \mathbb{R}^d$ that*

$$\max\Big\{\langle x-y, \mu(s,x)-\mu(s,y)\rangle, \tfrac{1}{2}\|\sigma(s,x)-\sigma(s,y)\|_F^2\Big\} \leqslant \tfrac{c}{2}\|x-y\|^2 \tag{2.128}$$

and $v^*\sigma(s,x)(\sigma(s,x))^*v \geqslant \alpha\|v\|^2$, assume for all $r \in (0,\infty)$, $j \in \{1,2,\dots,d\}$ that

$$\sup\left(\left\{\frac{\|\frac{\partial\mu}{\partial x}(t,x)-\frac{\partial\mu}{\partial x}(t,y)\|_F + \|\frac{\partial\sigma}{\partial x_j}(t,x)-\frac{\partial\sigma}{\partial x_j}(t,y)\|_F}{\|x-y\|}\colon t\in[0,T], x,y\in O_r, x\neq y\right\}\cup\{0\}\right) < \infty, \tag{2.129}$$

for every $t \in [0,T]$, $x \in O$ *let* $X^x_t = (X^x_{t,s})_{s\in[t,T]} \colon [t,T] \times \Omega \to O$ *be an* $(\mathbb{F}_s)_{s\in[t,T]}$*-adapted stochastic process with continuous sample paths satisfying that for all* $s \in [t,T]$ *it holds a.s. that*

$$X^x_{t,s} = x + \int_t^s \mu(r, X^x_{t,r})\,\mathrm{d}r + \int_t^s \sigma(r, X^x_{t,r})\,\mathrm{d}W_r, \tag{2.130}$$

assume for all $t \in [0,T]$, $\omega \in \Omega$ *that* $\big([t,T] \times O \ni (s,x) \mapsto X^x_{t,s}(\omega) \in O\big) \in C^{0,1}([t,T] \times O, O)$, *for every* $t \in [0,T]$, $x \in O$ *let* $Z^x_t = (Z^x_{t,s})_{s\in(t,T]} \colon (t,T] \times \Omega \to \mathbb{R}^{d+1}$ *be an* $(\mathbb{F}_s)_{s\in(t,T]}$*-adapted stochastic process with continuous sample paths satisfying that for all* $s \in (t,T]$ *it holds a.s. that*

$$Z^x_{t,s} = \begin{pmatrix} 1 \\ \frac{1}{s-t}\int_t^s (\sigma(r, X^x_{t,r}))^{-1}\left(\frac{\partial}{\partial x} X^x_{t,r}\right)\mathrm{d}W_r \end{pmatrix}, \tag{2.131}$$

let $V \in C^2(O,(0,\infty))$ *satisfy for all* $t \in [0,T]$, $x \in O$ *that*

$$\begin{aligned} &\langle \mu(t,x), (\nabla V)(x)\rangle + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess} V)(x)) \\ &\quad + L\|(\nabla V)(x)\| \leqslant \rho V(x), \end{aligned} \tag{2.132}$$

let $f \in C([0,T] \times O \times \mathbb{R}^{d+1}, \mathbb{R})$, $g \in C(O,\mathbb{R})$ *satisfy for all* $t \in [0,T]$, $s \in [t,T]$, $x \in O$, $w_1, w_2 \in \mathbb{R}^{d+1}$ *that* $\mathbb{E}[|g(X^x_{t,T})|^2] + \mathbb{E}[|f(s, X^x_{t,s}, w_1)|^2] < \infty$ *and* $|f(t,x,w_1) - f(t,x,w_2)| \leqslant L|\!|\!|w_1 - w_2|\!|\!|$, *and assume that* $\inf_{r\in(0,\infty)}[\sup_{t\in[0,T)\setminus K_r}\sup_{x\in O\setminus O_r}(\frac{|g(x)|+|f(t,x,0)|\sqrt{T-t}}{\sqrt{V(x)}})] = 0$ *and* $\liminf_{r\to\infty}[\inf_{x\in O\setminus O_r} V(x)] = \infty$. *Then*

(i) *there exists a unique* $v \in C([0,T] \times O,\mathbb{R}) \cap C^{0,1}([0,T) \times O,\mathbb{R})$ *which satisfies for all* $t \in [0,T)$, $x \in O$ *that* $\limsup_{r\to\infty}[\sup_{s\in[0,T)\setminus K_r}\sup_{y\in O\setminus O_r}(\frac{|\!|\!|(v,\nabla_x v)(s,y)|\!|\!|}{\sqrt{V(y)}}\sqrt{T-s})] = 0$, $\mathbb{E}[|g(X^x_{t,T})||\!|\!|Z^x_{t,T}|\!|\!| + \int_t^T |f(r, X^x_{t,r}, (v,\nabla_x v)(r, X^x_{t,r}))||\!|\!|Z^x_{t,r}|\!|\!|\,\mathrm{d}r] < \infty$, *and*

$$(v,\nabla_x v)(t,x) = \mathbb{E}\left[g(X^x_{t,T})Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, (v,\nabla_x v)(r, X^x_{t,r}))Z^x_{t,r}\,\mathrm{d}r\right], \tag{2.133}$$

(ii) *there exists a unique viscosity solution* $u \in \{\mathbf{u} \in C([0,T] \times O,\mathbb{R}) \colon \limsup_{r\to\infty}[\sup_{t\in[0,T]}\sup_{x\in O\setminus O_r}(\frac{|\mathbf{u}(t,x)|}{V(x)})] = 0\}$ *of*

$$\begin{aligned} &(\tfrac{\partial u}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x u)(t,x)\rangle \\ &\quad + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x u)(t,x)) + f(t,x,(v,\nabla_x u)(t,x)) = 0 \end{aligned} \tag{2.134}$$

with $u(T,x) = g(x)$ *for* $(t,x) \in (0,T) \times O$, *and*

(iii) *for all* $t \in [0,T]$, $x \in O$ *it holds that* $u(t,x) = v(t,x)$.

Proof of Corollary 2.3.2. Throughout this proof let $\mathbb{V} \colon [0,T] \times O \to \mathbb{R}$ satisfy for all $t \in [0,T]$, $x \in O$ that $\mathbb{V}(t,x) = e^{-\rho t}V(x)$. Note that the product rule and the fact that $V \in C^2(O,(0,\infty))$ ensure that $\mathbb{V} \in C^{1,2}([0,T] \times O,(0,\infty))$. This and (2.132) show that for all $t \in [0,T]$, $x \in O$ it holds that

$$\begin{aligned} &(\tfrac{\partial \mathbb{V}}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x \mathbb{V})(t,x)\rangle + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x \mathbb{V})(t,x)) \\ &\qquad + L\|(\nabla_x \mathbb{V})(x)\| \\ &= e^{-\rho t}\Big(-\rho V(x) + \langle \mu(t,x), (\nabla V)(x)\rangle \\ &\qquad + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess} V)(x)) + L\|(\nabla V)(x)\|\Big) \leqslant 0. \end{aligned} \tag{2.135}$$

Furthermore, note that the fact that $\inf_{r\in(0,\infty)}[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}(\frac{|g(x)|}{\sqrt{V(x)}}+\frac{|f(t,x,0,0)|}{\sqrt{V(x)}}\cdot\sqrt{T-t})]=0$ ensures that

$$\begin{aligned}&\inf_{r\in(0,\infty)}\left[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}\left(\frac{|g(x)|}{\sqrt{\mathbb{V}(T,x)}}+\frac{|f(t,x,0,0)|}{\sqrt{\mathbb{V}(t,x)}}\sqrt{T-t}\right)\right]\\&=\inf_{r\in(0,\infty)}\left[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}\left(\frac{|g(x)|}{\sqrt{e^{-\rho T}V(x)}}+\frac{|f(t,x,0,0)|}{\sqrt{e^{-\rho t}V(x)}}\sqrt{T-t}\right)\right]\\&\leqslant\sqrt{e^{\rho T}}\inf_{r\in(0,\infty)}\left[\sup_{t\in[0,T)\backslash K_r}\sup_{x\in O\backslash O_r}\left(\frac{|g(x)|}{\sqrt{V(x)}}+\frac{|f(t,x,0,0)|}{\sqrt{V(x)}}\sqrt{T-t}\right)\right]=0.\end{aligned}\tag{2.136}$$

In addition, observe that the assumption that $\liminf_{r\to\infty}[\inf_{x\in O\backslash O_r}V(x)]=\infty$ guarantees that

$$\liminf_{r\to\infty}[\inf_{t\in[0,T]}\inf_{x\in O\backslash O_r}\mathbb{V}(t,x)]=\infty.\tag{2.137}$$

Combining this with (2.135)-(2.136) and Theorem 2.3.1 (applied with $V\curvearrowleft\mathbb{V}$ in the notation of Theorem 2.3.1) establishes items (i)-(iii). The proof of Corollary 2.3.2 is thus complete. □

In the following corollary we show that the function $\mathbb{R}^d\ni x\mapsto(1+\|x\|^2)^{\frac{p}{2}}\in\mathbb{R}$, $p\in(0,\infty)$, satisfies the condition (2.102) in Theorem 2.3.1 and can therefore - under the right assumptions on the coefficients μ and σ- ensure that there exists a solution to the SFPE in (2.103) which is also a viscosity solution to the corresponding PDE. The proof of the following corollary, Corollary 2.3.3, uses some ideas of the proof of [10, Corollary 3.9].

Corollary 2.3.3. *Let $d\in\mathbb{N}$, $\alpha,c,L,T\in(0,\infty)$, let $\langle\cdot,\cdot\rangle\colon\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ be the standard Euclidean scalar product on $\mathbb{R}^d$, let $\|\cdot\|\colon\mathbb{R}^d\to[0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, let $|||\cdot|||\colon\mathbb{R}^{d+1}\to[0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^{d+1}$, let $\|\cdot\|_F\colon\mathbb{R}^{d\times d}\to[0,\infty)$ be the Frobenius norm on $\mathbb{R}^{d\times d}$, let $(\Omega,\mathcal{F},\mathbb{P},(\mathbb{F}_s)_{s\in[0,T]})$ be a filtered probability space satisfying the usual conditions, let $W\colon[0,T]\times\Omega\to\mathbb{R}^d$ be a standard $(\mathbb{F}_s)_{s\in[0,T]}$-Brownian motion, let $\mu\in C^{0,1}([0,T]\times\mathbb{R}^d,\mathbb{R}^d)$, $\sigma\in C^{0,1}([0,T]\times\mathbb{R}^d,\mathbb{R}^{d\times d})$ satisfy for all $s\in[0,T]$, $x,y\in\mathbb{R}^d$, $v\in\mathbb{R}^d$ that*

$$\max\left\{\langle x-y,\mu(s,x)-\mu(s,y)\rangle,\tfrac{1}{2}\|\sigma(s,x)-\sigma(s,y)\|_F^2\right\}\leqslant\tfrac{c}{2}\|x-y\|^2,\tag{2.138}$$

$\max\{\langle x,\mu(t,x)\rangle,\|\sigma(t,x)\|_F^2\}\leqslant c(1+\|x\|^2)$, and $v^*\sigma(s,x)(\sigma(s,x))^*v\geqslant\alpha\|v\|^2$, assume for all $r\in(0,\infty)$, $j\in\{1,2,\dots,d\}$ that

$$\begin{aligned}\sup\Bigg(\Bigg\{&\tfrac{\|\frac{\partial\mu}{\partial x}(t,x)-\frac{\partial\mu}{\partial x}(t,y)\|_F+\|\frac{\partial\sigma}{\partial x_j}(t,x)-\frac{\partial\sigma}{\partial x_j}(t,y)\|_F}{\|x-y\|}:\\&t\in[0,T],x,y\in\{z\in\mathbb{R}^d,\|z\|\leqslant r\},x\neq y\Bigg\}\cup\{0\}\Bigg)<\infty,\end{aligned}\tag{2.139}$$

for every $t\in[0,T]$, $x\in\mathbb{R}^d$ let $X_t^x=(X_{t,s}^x)_{s\in[t,T]}\colon[t,T]\times\Omega\to\mathbb{R}^d$ be an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s\in[t,T]$ it holds a.s. that

$$X_{t,s}^x=x+\int_t^s\mu(r,X_{t,r}^x)\,\mathrm{d}r+\int_t^s\sigma(r,X_{t,r}^x)\,\mathrm{d}W_r,\tag{2.140}$$

assume for all $t\in[0,T]$, $\omega\in\Omega$ that $\big([t,T]\times\mathbb{R}^d\ni(s,x)\mapsto X_{t,s}^x(\omega)\in\mathbb{R}^d\big)\in C^{0,1}([t,T]\times\mathbb{R}^d,\mathbb{R}^d)$, for every $t\in[0,T]$, $x\in\mathbb{R}^d$ let $Z_t^x=(Z_{t,s}^x)_{s\in(t,T]}\colon(t,T]\times\Omega\to\mathbb{R}^{d+1}$ be an

$(\mathbb{F}_s)_{s\in(t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in (t,T]$ it holds a.s. that

$$Z^x_{t,s} = \left(\begin{matrix} 1 \\ \frac{1}{s-t} \int_t^s (\sigma(r, X^x_{t,r}))^{-1} \left(\frac{\partial}{\partial x} X^x_{t,r}\right) \mathrm{d}W_r \end{matrix} \right), \tag{2.141}$$

let $f \in C([0,T] \times \mathbb{R}^d \times \mathbb{R}^{d+1}, \mathbb{R})$, $g \in C(\mathbb{R}^d, \mathbb{R})$ be at most polynomially growing, and assume that for all $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$, $w_1, w_2 \in \mathbb{R}^{d+1}$ it holds that $\mathbb{E}[|g(X^x_{t,T})|^2] + \mathbb{E}[|f(s, X^x_{t,s}, w_1)|^2] < \infty$ and $|f(t,x,w_1) - f(t,x,w_2)| \leqslant L|||w_1 - w_2|||$. Then

(i) there exists a unique $v \in C([0,T] \times \mathbb{R}^d, \mathbb{R}) \cap C^{0,1}([0,T) \times \mathbb{R}^d, \mathbb{R})$ which satisfies that $((v, \nabla_x v)(t,x)\sqrt{T-t})_{t\in[0,T),x\in\mathbb{R}^d}$ grows at most polynomially and for all $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that $\mathbb{E}[|g(X^x_{t,T})| |||Z^x_{t,T}||| + \int_t^T |f(r, X^x_{t,r}, (v, \nabla_x v)(r, X^x_{t,r}))| |||Z^x_{t,r}||| \, \mathrm{d}r] < \infty$ and

$$(v, \nabla_x v)(t,x) = \mathbb{E}\left[g(X^x_{t,T}) Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, (v, \nabla_x v)(r, X^x_{t,r})) Z^x_{t,r} \, \mathrm{d}r \right], \tag{2.142}$$

(ii) there exists a unique at most polynomially growing viscosity solution $u \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$ of

$$\begin{aligned} &(\tfrac{\partial u}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x u)(t,x) \rangle \\ &+ \tfrac{1}{2} \operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^* (\operatorname{Hess}_x u)(t,x)) + f(t, x(u, \nabla_x u)(t,x)) = 0 \end{aligned} \tag{2.143}$$

with $u(T,x) = g(x)$ for $(t,x) \in (0,T) \times \mathbb{R}^d$, and

(iii) for all $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that $u(t,x) = v(t,x)$.

Proof of Corollary 2.3.3. Throughout this proof let $V_q \colon \mathbb{R}^d \to (0,\infty)$, $q \in (0,\infty)$, satisfy for all $q \in (0,\infty)$, $x \in \mathbb{R}^d$ that $V_q(x) = (1 + \|x\|^2)^{\frac{q}{2}}$. First note that [13, Lemma 3.3] (applied for every $q \in (0,\infty)$ with $p \curvearrowleft q$, $\mathcal{O} \curvearrowleft \mathbb{R}^d$ in the notation of [13, Lemma 3.3]) and the assumption that for all $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that $\max\{\langle x, \mu(t,x) \rangle, \|\sigma(t,x)\|_F^2\} \leqslant c(1 + \|x\|^2)$ demonstrate that

(I) for all $q \in (0,\infty)$ it holds that $V_q \in C^\infty(\mathbb{R}^d, (0,\infty))$ and

(II) for all $q \in (0,\infty)$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\langle \mu(t,x), (\nabla V_q)(x) \rangle + \tfrac{1}{2} \operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^* (\operatorname{Hess} V_q)(x)) \\ &\leqslant \tfrac{cq}{2} \max\{q+1, 3\} V_q(x). \end{aligned} \tag{2.144}$$

Observe that item (I) and the product rule imply that for all $q \in (0,\infty)$, $x \in \mathbb{R}^d$ it holds that

$$(\nabla V_q)(x) = qx(1 + \|x\|^2)^{\frac{q}{2}-1} = qV_q(x) \tfrac{x}{1+\|x\|^2}. \tag{2.145}$$

The fact that for all $a \in [0,\infty)$ it holds that $a \leqslant 1 + a^2$ therefore shows that for all $q \in (0,\infty)$, $x \in \mathbb{R}^d$ it holds that

$$\|(\nabla V_q)(x)\| = qV_q(x) \tfrac{\|x\|}{1+\|x\|^2} \leqslant qV_q(x). \tag{2.146}$$

Combining this with item (II) proves that for all $q \in (0,\infty)$, $t \in [0,T]$, $x \in O$ it holds that

$$\begin{aligned} &\langle \mu(t,x), (\nabla V_q)(x) \rangle + \tfrac{1}{2} \operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^* (\operatorname{Hess} V_q)(x)) \\ &+ L\|(\nabla V_q)(x)\| \leqslant (\tfrac{cq}{2} \max\{q+1, 3\} + Lq) V_q(x). \end{aligned} \tag{2.147}$$

Next note that for all $q \in (0, \infty)$ it holds that

$$\liminf_{r\to\infty}[\inf_{x\in\mathbb{R}^d,\|x\|>r} V_q(x)] = \infty. \tag{2.148}$$

Furthermore, observe that the assumption that f and g are at most polynomially growing ensures that there exists $p \in (0, \infty)$ which satisfies

$$\sup_{t\in[0,T]} \sup_{x\in\mathbb{R}^d} \left(\frac{|g(x)| + |f(t,x,0)|\sqrt{T-t}}{\sqrt{V_p(x)}} \right) < \infty. \tag{2.149}$$

This shows that for all $q \in [p, \infty)$ it holds that

$$\limsup_{r\to\infty} \left[\sup_{t\in[\max\{T-\frac{1}{r},0\},T)} \sup_{x\in\mathbb{R}^d,\|x\|>r} \left(\frac{|g(x) + f(t,x,0)\sqrt{T-t}|}{\sqrt{V_q(x)}} \right) \right] = 0. \tag{2.150}$$

Combining this with (2.148) and item (ii) of Corollary 2.3.2 (applied with $\rho \curvearrowleft (cp\max\{2p+1,3\} + 2pL)$, $O \curvearrowleft \mathbb{R}^d$, $V \curvearrowleft V_{2p}$ in the notation of Corollary 2.3.2) demonstrates that there exists a unique viscosity solution $u \in \{\mathbf{u} \in C([0,T] \times \mathbb{R}^d, \mathbb{R}) \colon \limsup_{r\to\infty} [\sup_{t\in[0,T]} \sup_{x\in\mathbb{R}^d,\|x\|>r} (\frac{|\mathbf{u}(t,x)|}{V_{2p}(x)})] = 0\}$ of

$$\begin{aligned}(\tfrac{\partial u}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x u)(t,x)\rangle + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x u)(t,x))\\ + f(t,x,(u,\nabla_x u)(t,x)) = 0\end{aligned} \tag{2.151}$$

with $u(T,x) = g(x)$ for $(t,x) \in (0,T) \times \mathbb{R}^d$. Let $v \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$ be an at most polynomially growing viscosity solution of

$$\begin{aligned}(\tfrac{\partial v}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x v)(t,x)\rangle + \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x v)(t,x))\\ + f(t,x,(v,\nabla_x v)(t,x)) = 0\end{aligned} \tag{2.152}$$

with $v(T,x) = g(x)$ for $(t,x) \in (0,T) \times \mathbb{R}^d$. Observe that the assumption that v is at most polynomially growing ensures that there exists $\beta \in [2p, \infty)$ which satisfies that

$$\limsup_{r\to\infty} \left[\sup_{t\in[0,T]} \sup_{x\in\mathbb{R}^d,\|x\|>r} \left(\frac{|v(t,x)|}{V_\beta(x)} \right) \right] = 0. \tag{2.153}$$

Item (ii) of Corollary 2.3.2 (applied with $\rho \curvearrowleft \rho_\beta$, $O \curvearrowleft \mathbb{R}^d$, $V \curvearrowleft V_\beta$ in the notation of Corollary 2.3.2), (2.151), and (2.152) therefore demonstrate that $u = v$. This establishes item (ii). Next observe that item (i) of Corollary 2.3.2 (applied with $\rho \curvearrowleft \rho_{2p}$, $O \curvearrowleft \mathbb{R}^d$, $V \curvearrowleft V_{2p}$ in the notation of Corollary 2.3.2) shows that for all $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that $\mathbb{E}[|g(X^x_{t,T})| \|| Z^x_{t,T} \|| + \int_t^T |f(r, X^x_{t,r}, (u, \nabla_x u)(r, X^x_{t,r}))| \|| Z^x_{t,r} \|| \, dr] < \infty$ and

$$(u, \nabla_x u)(t,x) = \mathbb{E}\left[g(X^x_{t,T}) Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, (u, \nabla_x u)(r, X^x_{t,r})) Z^x_{t,r} \, dr \right]. \tag{2.154}$$

Let $w \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$ satisfy that $((w, \nabla_x w)(t,x)\sqrt{T-t})_{t\in[0,T),x\in\mathbb{R}^d}$ grows at most polynomially and that for all $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\mathbb{E}[|g(X^x_{t,T})| \|| Z^x_{t,T} \|| + \int_t^T |f(r, X^x_{t,r}, (w, \nabla_x w)(r, X^x_{t,r}))| \|| Z^x_{t,r} \|| \, dr] < \infty \tag{2.155}$$

and

$$(w, \nabla_x w)(t,x) = \mathbb{E}\left[g(X^x_{t,T}) Z^x_{t,T} + \int_t^T f(r, X^x_{t,r}, (w, \nabla_x w)(r, X^x_{t,r})) Z^x_{t,r} \, dr \right]. \tag{2.156}$$

Note that the fact that $((w, \nabla_x w)(t,x)\sqrt{T-t})_{t\in[0,T),x\in\mathbb{R}^d}$ grows at most polynomially implies that there exists $\gamma \in [\beta, \infty)$ which satisfies that

$$\limsup_{r\to\infty}\left[\sup_{t\in[0,T)\setminus K_r}\ \sup_{x\in\mathbb{R}^d,\|x\|>r}\left(\frac{\|\!|(w,\nabla_x w)(t,x)|\!\|}{\sqrt{V_\gamma(t,x)}}\sqrt{T-t}\right)\right] = 0. \tag{2.157}$$

Items (i) and (iii) in Corollary 2.3.2 (applied with $\rho \curvearrowleft \rho_\gamma$, $O \curvearrowleft \mathbb{R}^d$, $V \curvearrowleft V_\gamma$ in the notation of Corollary 2.3.2), (2.154), and (2.156) therefore demonstrate that $u = v = w$. This establishes items (i) and (iii). The proof of Corollary 2.3.3 is thus complete. □

Chapter 3

Multi-level Picard (MLP) approximation of high-dimensional semilinear partial differential equations with gradient-dependent nonlinearities

This chapter is going to prove that the theoretical results from Chapters 1 and 2 lead to an approximation scheme that can solve semilinear partial differential equations with gradient-dependent nonlinearities numerically without suffering from the curse of dimensionality. The central goal of this chapter is to find a numerical approximation of the viscosity solutions $u\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ of partial differential equations of the form

$$\begin{aligned}&(\tfrac{\partial u}{\partial t})(t,x) + \langle \mu(t,x), (\nabla_x u)(t,x)\rangle \\ &+ \tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x u)(t,x)) + f(t,x,u(t,x),(\nabla_x u)(t,x)) = 0\end{aligned} \tag{3.1}$$

with $u(T,x) = g(x)$ for $(t,x) \in (0,T) \times \mathbb{R}^d$. To set up our approximation scheme, we combine the Feynman-Kac type representation result in Theorem 2.3.1 and the following approximation approach from the theory of ordinary differential equations. If we assume $f \in C(\mathbb{R},\mathbb{R})$ and $y\colon [0,T] \to \mathbb{R}$ which satisfy $y(0) = y_0$ and $y' = f(y)$ we can use the fundamental theorem of calculus to reformulate the problem into

$$y(t) = y_0 + \int_0^t f(y(s))\,\mathrm{d}s. \tag{3.2}$$

Moreover, if we assume f to be Lipschitz continuous, then Picard iterations for sufficiently small $t \in (0,T]$ lead to

$$y^0(t) = y_0, \qquad y^{n+1}(t) = y_0 + \int_0^t f(y^n(r))\,\mathrm{d}r, \quad n \in \mathbb{N}_0. \tag{3.3}$$

If we transfer this idea to the setting of stochastic fixed point equations like in Chapter 1 we receive

$$\begin{aligned}\Phi^0(t,x) &= \mathbb{E}[g(X^x_{t,T})Z^x_{t,T}], \\ \Phi^{n+1}(t,x) &= \mathbb{E}[g(X^x_{t,T})Z^x_{t,T}] + \int_t^T \mathbb{E}[f(\Phi^n(r,X^x_{t,r}))Z^x_{t,r}]\,\mathrm{d}r, \quad n \in \mathbb{N}_0.\end{aligned} \tag{3.4}$$

The proof of Theorem 1.2.5 in Chapter 1 shows - under suitable assumptions - that there exists a unique fixed point $\mathfrak{v} \in \mathcal{V}$ which satisfies $\Phi(\mathfrak{v}) = \mathfrak{v}$. This together with a telescoping sum argument leads to the following construction of an approximation scheme. Set $\Phi^{-1} = 0$ and define for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\Phi^{n+1}(t,x)\\
&= \mathbb{E}[g(X^x_{t,T})Z^x_{t,T}]\\
&\qquad + \int_t^T \Big[f(\Phi^0(r,X^x_{t,r})) + \sum_{k=1}^n \mathbb{E}[f(\Phi^k(r,X^x_{t,r})) - f(\Phi^{k-1}(r,X^x_{t,r}))]\Big] Z^x_{t,r}\,\mathrm{d}r \qquad (3.5)\\
&= \mathbb{E}[g(X^x_{t,T})Z^x_{t,T}] + \sum_{k=0}^n \int_t^T \mathbb{E}\Big[f(\Phi^k(r,X^x_{t,r})) - \mathbb{1}_{\mathbb{N}}(k) f(\Phi^{k-1}(r,X^x_{t,r}))\Big] Z^x_{t,r}\,\mathrm{d}r.
\end{aligned}$$

The next step is to approximate the continuous quantities in the above equation by discrete ones. For this, we replace the expected values by Monte Carlo averages and the time integral by GauSS-Legendre quadrature formulas, respectively. In the implementation of the Monte Carlo averages summands of large k are costly to compute but have small variance while summands of small k are cheap to compute but have large variance. The reason to use GauSS-Legendre quadrature for the discrete approximation of the time integral is based on the fact that for smooth PDE solutions the quadrature error can be exactly calculated (cf. Lemma 3.2.5). Putting everything together, our approximation scheme for the numerical calculation of $(u, \nabla_x u)$ is given by

$$\begin{aligned}
V^\theta_{n,m,M,Q}(t,x) &= g(x)\mathbf{e}_1 + \tfrac{\mathbb{1}_{\mathbb{N}}(m)}{M^m} \sum_{i=1}^{M^m} \big(g(Y^{(\theta,0,-i),x,n}_{t,T}) - g(x)\big) Z^{(\theta,0,-i),x,n}_{t,T}\\
&\quad + \sum_{l=0}^{m-1} \sum_{s\in(t,T)} \tfrac{q^{Q,[t,T]}(s)}{M^{m-l}} \qquad (3.6)\\
&\qquad \cdot \Bigg[\sum_{i=1}^{M^{m-l}} (F(V^{(\theta,l,i)}_{n,l,M,Q}) - \mathbb{1}_{\mathbb{N}}(l) F(V^{(\theta,-l,i)}_{n,l-1,M,Q}))(s, Y^{(\theta,l,i),x,n}_{t,s}) Z^{(\theta,l,i),x,n}_{t,s}\Bigg].
\end{aligned}$$

For the computations of $U^{(\cdots)}_{n,m,M,Q}$ the full history $U^{(\cdot)}_{n,0,M,Q}, U^{(\cdot)}_{n,1,M,Q}, \ldots, U^{(\cdot)}_{n,m-1,M,Q}$ needs to be computed that is why we have full recursive multi-level Picard approximation.

The chapter is structured as follows. Section 3.1 introduces and studies the Multilevel Picard approximation scheme. First, we properly define the MLP approximation scheme in Subsection 3.1.1. Subsections 3.1.2 - 3.1.5, then establish well-definedness and several other properties required for the error analysis in Section 3.2. This chapter's main result is Corollary 3.2.8 in Subsection 3.2.2 below. Essential for the proof of Corollary 3.2.8 are the upper error bound established in Theorem 3.2.4 and the calculation of the GauSS-Legendre quadrature error in Lemma 3.2.5. The proof of Theorem 3.2.4, in turn, uses the non-linear Feynman-Kac type formula developed in Subsection 3.2.1 and certain findings from Section 3.1.

3.1 Full-history recursive MLP approximations

In this section we study MLP approximations for viscosity solutions of a certain class of semilinear PDEs with gradient-dependent nonlinearities. In Subsection 3.1.1 we thoroughly introduce our MLP approximation scheme and the considered setting. Subsection 3.1.2 establishes upper bounds for certain GauSS-Legendre quadrature terms which will be needed for the error analysis in Section 3.2. Subsections 3.1.3 -3.1.5 prove several measurability, integrability, independence, differentiability, and boundedness properties of the considered MLP approximation scheme or parts thereof.

3.1.1 Mathematical description of MLP approximations

Throughout this chapter we are going to use the following setting. In contrast to Chapters 1 and 2, where we assumed the coefficients μ and σ to fulfil the global monotonicity assumption, we now need them to satisfy (3.7). In Lemma 3.1.2 we provide a result that the assumption in (3.7) implies that μ and σ are globally Lipschitz continuous. The reason why we need stronger assumptions on the coefficients than in the previous chapters lies in the convergence of the required Euler-Maruyama approximations.

Setting 3.1.1. *Let $d \in \mathbb{N}$, $\alpha, \beta, b, T \in (0,\infty)$, $K \in [1,\infty)$, $L \in \mathbb{R}^{d+1}$, let $\mathcal{L} \in [0,\infty)$ satisfy that $\mathcal{L} = \sum_{\nu=1}^{d+1}|L_\nu|$, let $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^d$, let $|||\cdot|||\colon \mathbb{R}^{d+1} \to [0,\infty)$ be the standard Euclidean norm on $\mathbb{R}^{d+1}$, let $\mu = (\mu_1,\ldots,\mu_d) \in C^1(\mathbb{R}^d,\mathbb{R}^d)$, $\sigma = (\sigma_{ij})_{i,j\in\{1,2,\ldots,d\}} \in C^1(\mathbb{R}^d,\mathbb{R}^{d\times d})$ satisfy for all $x,y \in \mathbb{R}^d$ that*

$$\max\left\{\sum_{i,j=1}^{d}\left|\frac{\partial\mu_i}{\partial x_j}(x)\right|, \left(\sum_{i,j,k=1}^{d}\left|\frac{\partial\sigma_{ij}}{\partial x_k}(x)\right|\right)^2\right\} \leqslant K \tag{3.7}$$

and $y^*\sigma(x)(\sigma(x))^*y \geqslant \alpha\|y\|^2$, for every $p \in \mathbb{N}$, $x \in \mathbb{R}^d$ let $\gamma_x^{(p)} \in [0,\infty)$ satisfy that

$$\gamma_x^{(p)} = 2^{p-1}(\|\mu(x)\|^p + p\|\sigma(x)\|_F^p), \tag{3.8}$$

let $f \in C([0,T]\times\mathbb{R}^d\times\mathbb{R}^{d+1},\mathbb{R})$, $F\colon \mathcal{M}(\mathcal{B}([0,T]\times\mathbb{R}^d),\mathcal{B}(\mathbb{R}^{d+1})) \to \mathcal{M}(\mathcal{B}([0,T]\times\mathbb{R}^d),\mathcal{B}(\mathbb{R}))$ satisfy for all $t\in[0,T]$, $x_1,x_2\in\mathbb{R}^d$, $y=(y_1,y_2,\ldots,y_{d+1})$, $z=(z_1,z_2,\ldots,z_{d+1})\in\mathbb{R}^{d+1}$, $w \in \mathcal{M}(\mathcal{B}([0,T]\times\mathbb{R}^d),\mathcal{B}(\mathbb{R}^{d+1}))$ that

$$(F(w))(t,x_1) = f(t,x_1,w(t,x_1)) \text{ and } |f(t,x_1,y)-f(t,x_2,z)| \leqslant \sum_{\nu=1}^{d+1} L_\nu|y_\nu - z_\nu|, \tag{3.9}$$

let $g \in C(\mathbb{R}^d,\mathbb{R})$ satisfy for all $x_1,x_2\in\mathbb{R}^d$ that

$$|g(x_1)-g(x_2)| \leqslant \mathcal{L}\|x_1-x_2\|, \tag{3.10}$$

let $u^\infty \in C^{1,2}([0,T]\times\mathbb{R}^d,\mathbb{R})$ satisfy for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that $u(T,x)=g(x)$ and

$$\begin{aligned}&(\tfrac{\partial u^\infty}{\partial t})(t,x) + \langle\mu(t,x),(\nabla_x u^\infty)(t,x)\rangle\\&+\tfrac{1}{2}\operatorname{Tr}(\sigma(t,x)[\sigma(t,x)]^*(\operatorname{Hess}_x u^\infty)(t,x)) + f(t,x,u^\infty(t,x),(\nabla_x u^\infty)(t,x)) = 0,\end{aligned} \tag{3.11}$$

let $v^\infty \in C([0,T]\times\mathbb{R}^d,\mathbb{R}^{d+1})$ satisfy for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that $v^\infty(t,x) = (u^\infty,\nabla_x u^\infty)(t,x)$, for every $N\in\mathbb{N}$, $t\in[0,T]$ let $(\mathfrak{c}_i^N)_{i\in\{1,2,\ldots,N\}} \subseteq [-1,1]$ be the N distinct roots of the Legendre polynomial $[-1,1]\ni x\mapsto \frac{1}{2^N N!}\frac{\partial^N}{\partial x^N}[(x^2-1)^N]\in\mathbb{R}$ and $q^{N,[t,T]}\colon [t,T]\to\mathbb{R}$ be the function which satisfies for all $s\in[t,T]$ that

$$q^{N,[t,T]}(s) = \begin{cases}\int_t^T\left[\prod_{\substack{i\in\{1,2,\ldots,N\},\\ \mathfrak{c}_i^N\neq\frac{2s-(t+T)}{T-t}}}\frac{2x-(T-t)\mathfrak{c}_i^N-(t+T)}{2s-(T-t)\mathfrak{c}_i^N-(t+T)}\right]dx & : (t<T), (\frac{2s-(t+T)}{T-t}\in\{\mathfrak{c}_1^N,\ldots,\mathfrak{c}_N^N\})\\ 0 & : \text{else},\end{cases} \tag{3.12}$$

let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_s)_{s\in[0,T]})$ *be a filtered probability space satisfying the usual conditions, let* $\Theta = \bigcup_{n\in\mathbb{N}} \mathbb{Z}^n$, *let* $W^\theta = (W^{\theta,i}_t)_{t\in[0,T],i\in\{1,2,\dots,d\}} \colon [0,T]\times\Omega \to \mathbb{R}^d$, $\theta\in\Theta$, *be i.i.d. standard* $(\mathbb{F}_s)_{s\in[0,T]}$*-Brownian motions, for every* $t\in[0,T]$ *let* $\lfloor\cdot\rfloor_t \colon \mathbb{R}\to\mathbb{R}$, *satisfy for all* $s\in[0,T]$ *that* $\lfloor s\rfloor_t = \sup([t,s]\cap(\frac{(T-t)}{n}\mathbb{N}))$, *for every* $n\in\mathbb{N}$, $\theta\in\Theta$, $t\in[0,T]$, $x\in\mathbb{R}^d$ *let* $Y^{\theta,x,n}_t = ((Y^{\theta,x,n}_{t,s})_i)_{s\in[t,T],i\in\{1,2,\dots,d\}} \colon [t,T]\times\Omega\to\mathbb{R}^d$ *satisfy for all* $s\in[t,T]$ *that* $Y^{\theta,x,n}_{t,t} = x$ *and*

$$\begin{aligned} &Y^{\theta,x,n}_{t,s} - Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}} \\ &= \mu(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})(s-\max\{t,\lfloor s\rfloor_t\}) + \sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})(W^\theta_s - W^\theta_{\max\{t,\lfloor s\rfloor_t\}}), \end{aligned} \tag{3.13}$$

for every $n\in\mathbb{N}$, $\theta\in\Theta$, $t\in[0,T]$, $x\in\mathbb{R}^d$ *let* $D^{\theta,x,n}_t = ((D^{\theta,x,n}_{t,s})_{ij})_{s\in[t,T],i,j\in\{1,2,\dots,d\}} \colon [t,T]\times\Omega\to\mathbb{R}^{d\times d}$ *satisfy for all* $i,j\in\{1,2,\dots,d\}$, $s\in[t,T]$ *that* $(D^{\theta,x,n}_{t,t})_{ij} = \delta_{ij}$ *and*

$$\begin{aligned} &(D^{\theta,x,n}_{t,s})_{ij} - (D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{ij} \\ &= \sum_{k=1}^d \bigg[\frac{\partial\mu_i}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{kj}(s-\max\{t,\lfloor s\rfloor_t\}) \\ &\quad + \sum_{m=1}^d \bigg[\frac{\partial\sigma_{im}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{kj}(W^{\theta,m}_s - W^{\theta,m}_{\max\{t,\lfloor s\rfloor_t\}})\bigg]\bigg], \end{aligned} \tag{3.14}$$

let $\mathbf{e}_1\in\mathbb{R}^{d+1}$ *satisfy* $\mathbf{e}_1 = (1,0,0,\dots,0)$, *for every* $n\in\mathbb{N}$, $\theta\in\Theta$, $t\in[0,T]$, $x\in\mathbb{R}^d$ *let* $Z^{\theta,x,n}_t = ((Z^{\theta,x,n}_{t,s})_i)_{s\in(t,T],i\in\{1,2,\dots,d+1\}} \colon [t,T]\times\Omega\to\mathbb{R}^{d+1}$ *satisfy for all* $s\in(t,T]$ *that* $Z^{\theta,x,n}_{t,t} = \mathbf{e}_1$ *and*

$$Z^{\theta,x,n}_{t,s} - Z^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}} = \begin{pmatrix} 0 \\ \frac{1}{s-t}(\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1} D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}(W^\theta_s - W^\theta_{\max\{t,\lfloor s\rfloor_t\}}) \end{pmatrix}, \tag{3.15}$$

let $\varphi\in C^{1,2}([0,T]\times\mathbb{R}^d,[0,\infty))$ *satisfy for all* $t\in[0,T]$, $x\in\mathbb{R}^d$ *that*

$$(\tfrac{\partial\varphi}{\partial t})(t,x) + \langle(\nabla_x\varphi)(t,x),\mu(x)\rangle + \tfrac{1}{2}\operatorname{Tr}(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x\varphi)(t,x)) \leqslant \beta\varphi(t,x) + b \tag{3.16}$$

and $\inf_{r\in[0,T]}\inf_{y\in\mathbb{R}^d}\varphi(r,y) > 0$, *and for every* n,M,Q *let* $V^\theta_{n,m,M,Q}\colon[0,T)\times\mathbb{R}^d\times\Omega\to\mathbb{R}^{d+1}$, $m\in\mathbb{Z}$, $\theta\in\Theta$, *satisfy for all* $\theta\in\Theta$, $m\in\mathbb{Z}$, $t\in[0,T]$, $x\in\mathbb{R}^d$ *that*

$$\begin{aligned} V^\theta_{n,m,M,Q}(t,x) &= g(x)\mathbf{e}_1 + \frac{\mathbb{1}_{\mathbb{N}}(m)}{M^m}\sum_{i=1}^{M^m}\big(g(Y^{(\theta,0,-i),x,n}_{t,T}) - g(x)\big)Z^{(\theta,0,-i),x,n}_{t,T} \\ &\quad + \sum_{l=0}^{m-1}\sum_{s\in(t,T)} \frac{q^{Q,[t,T]}(s)}{M^{m-l}} \\ &\quad \cdot \Bigg[\sum_{i=1}^{M^{m-l}} (F(V^{(\theta,l,i)}_{n,l,M,Q}) - \mathbb{1}_{\mathbb{N}}(l)F(V^{(\theta,-l,i)}_{n,l-1,M,Q}))(s,Y^{(\theta,l,i),x,n}_{t,s})Z^{(\theta,l,i),x,n}_{t,s}\Bigg]. \end{aligned} \tag{3.17}$$

The following elementary lemma proves that the assumption in (3.7) ensures that the coefficients μ and σ are Lipschitz continuous.

Lemma 3.1.2. *Let* $K\in[1,\infty)$ *and let* $\mu = (\mu_1,\dots,\mu_d)\in C^1(\mathbb{R}^d,\mathbb{R}^d)$, $\sigma = (\sigma_{ij})_{i,j\in\{1,2,\dots,d\}}\in C^1(\mathbb{R}^d,\mathbb{R}^{d\times d})$ *satisfy for all* $x\in\mathbb{R}^d$ *that*

$$\max\left\{\sum_{i,j=1}^d\left|\frac{\partial\mu_i}{\partial x_j}(x)\right|, \left(\sum_{i,j,k=1}^d\left|\frac{\partial\sigma_{ij}}{\partial x_k}(x)\right|\right)^2\right\} \leqslant K. \tag{3.18}$$

Then

(i) it holds for all $x \in \mathbb{R}^d$ that

$$\max\left\{\sqrt{\sum_{i,j=1}^{d}\left|\frac{\partial\mu_i}{\partial x_j}(x)\right|^2}, \sum_{i,j=1}^{d}\left(\sum_{k=1}^{d}\left|\frac{\partial\sigma_{ij}}{\partial x_k}(x)\right|\right)^2\right\} \leqslant K \tag{3.19}$$

and

(ii) it holds for all $x, y \in \mathbb{R}^d$ that

$$\max\left\{\|\mu(x)-\mu(y)\|, \|\sigma(x)-\sigma(y)\|_F\right\} \leqslant K\|x-y\|. \tag{3.20}$$

Proof of Lemma 3.1.2. First observe that (3.18) and the fact that for all $a, b \in [0,\infty)$ it holds that $\sqrt{a+b} \leqslant \sqrt{a}+\sqrt{b}$ imply that for all $x \in \mathbb{R}^d$ it holds that

$$\sqrt{\sum_{i,j=1}^{d}\left|\frac{\partial\mu_i}{\partial x_j}(x)\right|^2} \leqslant \sum_{i,j=1}^{d}\sqrt{\left|\frac{\partial\mu_i}{\partial x_j}(x)\right|^2} = \sum_{i,j=1}^{d}\left|\frac{\partial\mu_i}{\partial x_j}(x)\right| \leqslant K. \tag{3.21}$$

In addition, note that (3.18) and the fact that for all $a, b \in [0,\infty)$ it holds that $\sqrt{a+b} \leqslant \sqrt{a}+\sqrt{b}$ ensure that for all $x \in \mathbb{R}^d$ it holds that

$$\sqrt{\sum_{i,j=1}^{d}\left(\sum_{k=1}^{d}\left|\frac{\partial\sigma_{ij}}{\partial x_k}(x)\right|\right)^2} \leqslant \sum_{i,j=1}^{d}\sqrt{\left(\sum_{k=1}^{d}\left|\frac{\partial\sigma_{ij}}{\partial x_k}(x)\right|\right)^2} = \sum_{i,j,k=1}^{d}\left|\frac{\partial\sigma_{ij}}{\partial x_k}(x)\right| \leqslant \sqrt{K}. \tag{3.22}$$

Combining this with (3.21) establishes item (i). In the next step note that the mean value theorem and the assumption that $\mu \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ ensure that for all $i \in \{1, 2, \ldots, d\}$, $x, y \in \mathbb{R}^d$ there exists $\xi_i^{x,y} \in \mathbb{R}^d$ which satisfies that

$$|\mu_i(x)-\mu_i(y)| = \|(\nabla\mu_i)(\xi_i^{x,y})\|\|x-y\|. \tag{3.23}$$

Combining this with item (i) shows that for all $x, y \in \mathbb{R}^d$ it holds that

$$\begin{aligned}\|\mu(x)-\mu(y)\|^2 &= \sum_{i=1}^{d}|\mu_i(x)-\mu_i(y)|^2 = \sum_{i=1}^{d}\|(\nabla\mu_i)(\xi_i^{x,y})\|^2\|x-y\|^2 \\ &= \|x-y\|^2\sum_{i,j=1}^{d}\left|\frac{\partial\mu_i}{\partial x_j}(\xi_i^{x,y})\right|^2 \leqslant K^2\|x-y\|^2.\end{aligned} \tag{3.24}$$

Furthermore, observe that the mean value theorem and the assumption that $\sigma \in C^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$ show that for all $i, j \in \{1, 2, \ldots, d\}$, $x, y \in \mathbb{R}^d$ there exists $\rho_{ij}^{x,y} \in \mathbb{R}^d$ which satisfies that

$$|\sigma_{ij}(x)-\sigma_{ij}(y)| = \|(\nabla\sigma_{ij})(\rho_{ij}^{x,y})\|\|x-y\|. \tag{3.25}$$

This, (3.18), the assumption that $K \in [1,\infty)$, and the fact that for all $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ it holds that $\sum_{i=1}^{d}|x_i|^2 \leqslant (\sum_{i=1}^{d}|x_i|)^2$ demonstrate that for all $x, y \in \mathbb{R}^d$ it holds that

$$\begin{aligned}&\|\sigma(x)-\sigma(y)\|_F^2 = \sum_{i,j=1}^{d}|\sigma_{ij}(x)-\sigma_{ij}(y)|^2 = \sum_{i,j=1}^{d}\|(\nabla\sigma_{ij})(\rho_{ij}^{x,y})\|^2\|x-y\|^2 \\ &= \|x-y\|^2\sum_{i,j,k=1}^{d}\left|\frac{\partial\sigma_{ij}}{\partial x_k}(\rho_{ij}^{x,y})\right|^2 \leqslant \|x-y\|^2\left(\sum_{i,j,k=1}^{d}\left|\frac{\partial\sigma_{ij}}{\partial x_k}(\rho_{ij}^{x,y})\right|\right)^2 \\ &\leqslant K^2\|x-y\|^2.\end{aligned} \tag{3.26}$$

This establishes item (ii). The proof of Lemma 3.1.2 is thus complete. □

3.1.2 Preliminary results for GauSS-Legendre quadrature rules

In the following Lemma we prove upper bounds for the GauSS-Legendre quadrature of the functions $([t,T] \ni s \mapsto \frac{1}{\sqrt{s-t}} \in \mathbb{R})$ and $([t,T] \ni s \mapsto \frac{1}{\sqrt{T-s}\sqrt{s-t}} \in \mathbb{R})$ for $t \in [0,T]$. The proof follows the idea of [55, Lemma 3.3]. We will use Lemma 3.1.3 to ensure the integrability of the MLP approximation scheme in Subsection 3.1.5 and to calculate the global error in Section 3.2.

Lemma 3.1.3. *Assume Setting 3.1.1 and let $Q \in \mathbb{N}$. Then*

(i) it holds for all $t \in [0,T]$ that

$$\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\frac{1}{\sqrt{s-t}} \leqslant 2\sqrt{T-t} \tag{3.27}$$

and

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

$$\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\frac{1}{\sqrt{T-s}\sqrt{s-t}} \leqslant \pi. \tag{3.28}$$

Proof of Lemma 3.1.3. First note that for all $t \in [0,T)$, $r \in [0,1]$ with $2r-1 \in \{\mathfrak{c}_1^Q, \mathfrak{c}_2^Q, \ldots, \mathfrak{c}_Q^Q\}$ the substitution $[t,T] \ni x \mapsto \frac{x-t}{T-t} \in [0,1]$ implies that

$$\begin{aligned} &q^{Q,[t,T]}(r(T-t)+t) \\ &= \int_t^T \left[\prod_{\substack{i\in\{1,2,\ldots,Q\},\\ \mathfrak{c}_i^Q \neq \frac{2r(T-t)+2t-(t+T)}{T-t}}} \frac{2x-(T-t)\mathfrak{c}_i^Q-(t+T)}{2r(T-t)+2t-(T-t)\mathfrak{c}_i^Q-(t+T)} \right] dx \\ &= \int_t^T \left[\prod_{\substack{i\in\{1,2,\ldots,Q\},\\ \mathfrak{c}_i^Q\neq 2r-1}} \frac{2(x-t)-(T-t)\mathfrak{c}_i^Q-(T-t)}{(T-t)(2r-\mathfrak{c}_i^Q-1)} \right] dx \\ &= (T-t)\int_0^1 \left[\prod_{\substack{i\in\{1,2,\ldots,Q\},\\ \mathfrak{c}_i^Q\neq 2r-1}} \frac{2y-\mathfrak{c}_i^Q-1}{2r-\mathfrak{c}_i^Q-1} \right] dy. \end{aligned} \tag{3.29}$$

Combining this with (3.12) shows that for all $t \in [0,T)$, $r \in [0,1]$ it holds that

$$q^{Q,[t,T]}(r(T-t)+t) = (T-t)q^{Q,[0,1]}(r). \tag{3.30}$$

This implies that for all $t \in [0,T)$ it holds that

$$\sum_{s\in(t,T)} \frac{q^{Q,[t,T]}(s)}{\sqrt{s-t}} = \sum_{r\in(0,1)} \frac{q^{Q,[t,T]}(r(T-t)+t)}{\sqrt{r(T-t)}} = \sqrt{T-t}\sum_{r\in(0,1)} \frac{q^{Q,[0,1]}(r)}{\sqrt{r}}. \tag{3.31}$$

Furthermore, observe that for all $\varepsilon \in (0,\infty)$, $r \in (0,1)$ it holds that

$$\frac{\partial^{2Q}}{\partial r^{2Q}}\frac{1}{\sqrt{r+\varepsilon}} = (r+\varepsilon)^{-(2Q+\frac{1}{2})}(-1)^{2Q}\prod_{i=0}^{2Q-1}\left(\tfrac{1}{2}+i\right) \geqslant 0. \tag{3.32}$$

In addition, observe that the error presentation for the GauSS-Legendre quadrature rule (cf., e.g., [28, Display (2.7.12)] (applied with $a \curvearrowleft 0$, $b \curvearrowleft 1$, $n \curvearrowleft Q$) in the notation of [28, Display (2.7.12)]) proves that for every $\varepsilon \in (0,\infty)$ there exists $\xi \in (0,1)$ which satisfies

$$\sum_{r\in(0,1)} \frac{q^{Q,[0,1]}(r)}{\sqrt{r+\varepsilon}} = \int_0^1 \frac{1}{\sqrt{r+\varepsilon}}\,\mathrm{d}r - \frac{(Q!)^4}{(2Q+1)[(2Q)!]^3}\left(\frac{\partial^{2Q}}{\partial r^{2Q}}\frac{1}{\sqrt{r+\varepsilon}}\right)\bigg|_{r=\xi}. \tag{3.33}$$

This and (3.32) show that for all $\varepsilon \in (0,1)$ it holds that

$$\sum_{r\in(0,1)} \frac{q^{Q,[0,1]}(r)}{\sqrt{r+\varepsilon}} \leqslant \int_0^1 \frac{1}{\sqrt{r+\varepsilon}}\,\mathrm{d}r \leqslant \int_0^1 \frac{1}{\sqrt{r}}\,\mathrm{d}r = 2. \tag{3.34}$$

Combining this with (3.31) and letting $\varepsilon \to 0$ implies that for all $t \in [0,T]$ it holds that

$$\sum_{s\in(t,T)} \frac{q^{Q,[t,T]}(s)}{\sqrt{s-t}} = \sqrt{T-t}\sum_{r\in(0,1)} \frac{q^{Q,[0,1]}(r)}{\sqrt{r}} \leqslant 2\sqrt{T-t}. \tag{3.35}$$

This establishes item (i). Next observe that (3.30) ensures that for all $t \in [0,T)$ it holds that

$$\sum_{s\in(t,T)} \frac{q^{Q,[t,T]}(s)}{\sqrt{T-s}\sqrt{s-t}} = \sum_{r\in(0,1)} \frac{q^{Q,[t,T]}(r(T-t)+t)}{\sqrt{(T-t)(1-r)}\sqrt{r(T-t)}} = \sum_{r\in(0,1)} \frac{q^{Q,[0,1]}(r)}{\sqrt{r(1-r)}}. \tag{3.36}$$

Moreover, note that the general Leibniz rule ensures that for all $\varepsilon \in (0,\infty)$, $r \in (0,1)$ it holds that

$$\begin{aligned}
&\frac{\partial^{2Q}}{\partial r^{2Q}}\left(\frac{1}{\sqrt{r+\varepsilon}\sqrt{1-r+\varepsilon}}\right)\\
&= \sum_{k=1}^{2Q}\binom{2Q}{k}\left[(r+\varepsilon)^{-(2Q-k+\frac{1}{2})}\prod_{i=0}^{2Q-k-1}\left(-\frac{1}{2}-i\right)\right]\\
&\quad\cdot\left[(1-r+\varepsilon)^{-(k+\frac{1}{2})}\prod_{i=0}^{k-1}\left(-\frac{1}{2}-i\right)\right]\\
&= \sum_{k=1}^{2Q}\binom{2Q}{k}\left[(r+\varepsilon)^{-(2Q-k+\frac{1}{2})}(-1)^{2Q-k}\prod_{i=0}^{2Q-k-1}\left(\frac{1}{2}+i\right)\right]\\
&\quad\cdot\left[(1-r+\varepsilon)^{-(k+\frac{1}{2})}(-1)^{k}\prod_{i=0}^{k-1}\left(\frac{1}{2}+i\right)\right] \geqslant 0.
\end{aligned} \tag{3.37}$$

In addition, note that the error presentation for the GauSS-Legendre quadrature rule (cf., e.g., [28, Display (2.7.12)] (applied with $a \curvearrowleft 0$, $b \curvearrowleft 1$, $n \curvearrowleft Q$ in the notation of [28, Display (2.7.12)])) demonstrates that for every $\varepsilon \in (0,\infty)$ there exists $\eta \in (0,1)$ which satisfies

$$\begin{aligned}
&\sum_{r\in(0,1)} \frac{q^{Q,[0,1]}(r)}{\sqrt{r+\varepsilon}\sqrt{1-r+\varepsilon}}\\
&= \int_0^1 \frac{1}{\sqrt{r+\varepsilon}\sqrt{1-r+\varepsilon}}\,\mathrm{d}r - \frac{(Q!)^4}{(2Q+1)[(2Q)!]^3}\left(\frac{\partial^{2Q}}{\partial r^{2Q}}\frac{1}{\sqrt{r+\varepsilon}\sqrt{1-r+\varepsilon}}\right)\bigg|_{r=\eta}.
\end{aligned} \tag{3.38}$$

Combining this with item (i) of Lemma 1.1.2 (applied with $a \curvearrowleft 0$, $b \curvearrowleft 1$, $\lambda \curvearrowleft 0$ in the notation of Lemma 1.1.2) and (3.37) proves that for all $\varepsilon \in (0,1)$ it holds that

$$\sum_{r\in(0,1)} \frac{q^{Q,[0,1]}(r)}{\sqrt{r+\varepsilon}\sqrt{1-r+\varepsilon}} \leqslant \int_0^1 \frac{1}{\sqrt{r+\varepsilon}\sqrt{1-r+\varepsilon}}\,\mathrm{d}r \leqslant \int_0^1 \frac{1}{\sqrt{r(1-r)}}\,\mathrm{d}r = \pi. \tag{3.39}$$

This, (3.36), and letting $\varepsilon \to 0$ show that for all $t \in [0,T]$ it holds that

$$\sum_{s\in(t,T)} \frac{q^{Q,[t,T]}(s)}{\sqrt{T-s}\sqrt{s-t}} = \sum_{r\in(0,1)} \frac{q^{Q,[0,1]}(r)}{\sqrt{r(1-r)}} \leqslant \pi. \tag{3.40}$$

This establishes item (ii). The proof of Lemma 3.1.3 is thus complete. □

3.1.3 Measurability and independence properties for MLP approximations

The following Lemma establishes measurability, distributional, and independence properties of the MLP approximation scheme and its components.

Lemma 3.1.4. *Assume Setting 3.1.1, $n, M, Q \in \mathbb{N}$, and let $\Lambda \subseteq [0,T]^2$ satisfy $\Lambda = \{(t,s) \in [0,T]^2 : t \leqslant s\}$. Then*

(i) it holds for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ that $Y^{\theta,x,n}_{t,s}$ is measurable and $D^{\theta,x,n}_{t,s}$ is measurable,

(ii) it holds for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ that $Z^{\theta,x,n}_{t,s}$ is measurable,

(iii) it holds for all $m \in \mathbb{N}_0$, $\theta \in \Theta$ that $V^{\theta}_{n,m,M,Q}$ is a continuous random field,

(iv) it holds for all $m_1, m_2 \in \mathbb{N}_0$, $i,j,k,l \in \mathbb{Z}$, $\theta \in \Theta$ with $(i,j) \neq (k,l)$ that $(V^{(\theta,i,j)}_{n,m_1,M,Q}(t,x))_{(t,x)\in[0,T)\times\mathbb{R}^d}$ and $(V^{(\theta,k,l)}_{n,m_2,M,Q}(t,x))_{(t,x)\in[0,T)\times\mathbb{R}^d}$ are independent,

(v) it holds for all $m \in \mathbb{N}_0$, $\theta \in \Theta$ that $(V^{\theta}_{n,m,M,Q}(t,x))_{(t,x)\in[0,T)\times\mathbb{R}^d}$ and $(Y^{\theta,x,n}_{t,s})_{(t,s,x)\in\Lambda\times\mathbb{R}^d}$ are independent, and

(vi) it holds for all $m \in \mathbb{N}_0$, $t \in [0,T)$, $x \in \mathbb{R}^d$ that $V^{\theta}_{n,m,M,Q}(t,x)$, $\theta \in \Theta$, are identically distributed.

Proof of Lemma 3.1.4. First let $A_1 \subseteq \mathbb{N}_0 \cap [0,n)$ satisfy that

$$\begin{aligned} A_1 = \{&k \in \mathbb{N}_0 \cap [0,n) \colon \forall \theta \in \Theta, t \in [0,T], s \in [t,T] \cap (\tfrac{kT}{n}, \tfrac{(k+1)T}{n}], x \in \mathbb{R}^d \\ &\text{it holds that } Y^{\theta,x,n}_{t,s} \colon \Omega \to \mathbb{R}^d \text{ is measurable and it holds that} \\ &D^{\theta,x,n}_{t,s} \colon \Omega \to \mathbb{R}^{d\times d} \text{ is measurable}\}. \end{aligned} \tag{3.41}$$

Observe that the fact that for all $i,j \in \{1,2,\dots,d\}$, $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T] \cap (0, \frac{T}{n}]$, $x \in \mathbb{R}^d$ it holds that

$$Y^{\theta,x,n}_{t,\max\{t,0\}} = x + \mu(x)(s - \max\{t,0\}) + \sigma(x)(W^{\theta}_s - W^{\theta}_{\max\{t,0\}}) \tag{3.42}$$

and

$$\begin{aligned} \left(D^{\theta,x,n}_{t,s}\right)_{ij} = \delta_{ij} + \sum_{k=1}^{d} \Bigg[&\frac{\partial \mu_i}{\partial x_k}(x)(s - \max\{t,0\}) \\ &+ \sum_{m=1}^{d} \left[\frac{\partial \sigma_{im}}{\partial x_k}(x)\big(W^{\theta,m}_s - W^{\theta,m}_{\max\{t,0\}}\big)\right]\Bigg], \end{aligned} \tag{3.43}$$

the assumption that $\mu \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, $\sigma \in C^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$, and the assumption that W^θ, $\theta \in \Theta$, are Brownian motions ensure that $0 \in A_1$. For the induction step assume $k \in A_1 \cap [0, n-1)$. Combining this with (3.13), (3.14), the fact that $\mu \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ and $\sigma \in C^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$, and the assumption that W^θ, $\theta \in \Theta$, are Brownian motions implies that $k+1 \in A_1$. Induction therefore proves item (i). In the next step let $A_2 \subseteq \mathbb{N}_0 \cap [0, n)$ satisfy that

$$\begin{aligned} A_2 = \{&k \in \mathbb{N}_0 \cap [0, n) \colon \forall \theta \in \Theta, t \in [0, T), s \in [t, T] \cap \left(\tfrac{kT}{n}, \tfrac{(k+1)T}{n}\right], x \in \mathbb{R}^d \\ &\text{it holds that } Z^{\theta,x,n}_{t,s} \colon \Omega \to \mathbb{R}^{d+1} \text{ is measurable}\}. \end{aligned} \tag{3.44}$$

Note that item (i), the fact that for all $\theta \in \Theta$, $t \in [0, T)$, $s \in [t, T] \cap (0, \frac{T}{n}]$, $x \in \mathbb{R}^d$ it holds that

$$Z^{\theta,x,n}_{t,s} = \begin{pmatrix} 1 \\ \frac{1}{s-t}(\sigma(Y^{\theta,x,n}_{t,\max\{t,0\}}))^{-1} D^{\theta,x,n}_{t,\max\{t,0\}}(W^\theta_s - W^\theta_{\max\{t,0\}}) \end{pmatrix}, \tag{3.45}$$

the fact that $\sigma^{-1} \in C(\mathbb{R}^d, \mathbb{R}^{d\times d})$, and the assumption that W^θ, $\theta \in \Theta$, are Brownian motions show that $0 \in A_2$. For the induction step assume that $k \in A_2 \cap [0, n-1)$. This, item (i), (3.15), the fact that $\sigma^{-1} \in C(\mathbb{R}^d, \mathbb{R}^{d\times d})$, and the assumption that W^θ, $\theta \in \Theta$, are Brownian motions demonstrate that $k+1 \in A_2$. Induction therefore establishes item (ii). Next let $A_3 \subseteq \mathbb{N}$ satisfy that

$$\begin{aligned} A_3 = \{&m \in \mathbb{N} \colon \forall \theta \in \Theta, \mathfrak{m} \in \mathbb{N}_0 \cap [0, m) \text{ it holds that} \\ &V^\theta_{n,\mathfrak{m},M,Q} \colon [0, T) \times \mathbb{R}^d \times \Omega \to \mathbb{R}^{d+1} \text{ is a continuous random field}\}. \end{aligned} \tag{3.46}$$

Observe that the fact that for all $\theta \in \Theta$, $t \in [0, T)$, $x \in \mathbb{R}^d$ it holds that $V^\theta_{n,0,M,Q}(t, x) = 0$ ensures that $1 \in A_3$. For the induction step assume $m \in A_3$. This, item (i), the assumption that W^θ, $\theta \in \Theta$, are Brownian motions, the assumption that $\mu \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, $\sigma \in C^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$, f and g are continuous, and (3.17) show that for all $\theta \in \Theta$ it holds that $V^\theta_{n,m,M,Q}$ is a continuous random field. This implies that $m+1 \in A_3$. Induction hence establishes item (iii). Next note that (3.17) guarantees that for all $m \in \mathbb{N}_0$, $\theta \in \Theta$ it holds that

$$\sigma(V^\theta_{n,m,M,Q}) \subseteq \sigma\big((W^{(\theta,\vartheta)})_{\vartheta\in\Theta}\big). \tag{3.47}$$

Combining this with the fact that for all $i, j, k, l \in \mathbb{Z}$, $\theta \in \Theta$ with $(i, j) \neq (k, l)$ it holds that $(W^{(\theta,i,j,\vartheta)})_{\vartheta\in\Theta}$ and $(W^{(\theta,k,l,\vartheta)})_{\vartheta\in\Theta}$ are independent establishes item (iv). In the next step observe that (3.13) ensures that for all $\theta \in \Theta$, $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that

$$\sigma(Y^{\theta,x,n}_t) \subseteq \sigma\big(W^\theta\big). \tag{3.48}$$

This, (3.47), and the fact that for all $\theta \in \Theta$ it holds that W^θ and $(W^{(\theta,\vartheta)})_{\vartheta\in\Theta}$ are independent establish item (v). For the next step let $A_4 \subseteq \mathbb{N}$ satisfy that

$$\begin{aligned} A_4 = \{&m \in \mathbb{N} \colon \forall \mathfrak{m} \in \mathbb{N}_0 \cap [0, m) \text{ it holds that} \\ &V^\theta_{n,\mathfrak{m},M,Q} \colon [0, T) \times \mathbb{R}^d \times \Omega \to \mathbb{R}^{d+1}, \theta \in \Theta, \text{ are identically distributed}\}. \end{aligned} \tag{3.49}$$

Note that the fact that for all $\theta \in \Theta$, $t \in [0, T)$, $x \in \mathbb{R}^d$ it holds that $V^\theta_{n,0,M,Q}(t, x) = 0$ ensures that $1 \in A_4$. For the induction step assume $m \in A_4$. Observe that the assumption that W^θ, $\theta \in \Theta$, are Brownian motions implies that for all $t \in [0, T]$, $s \in [t, T]$, $x \in \mathbb{R}^d$ it holds that $Y^{\theta,x,n}_{t,s}$, $\theta \in \Theta$, are identically distributed and $D^{\theta,x,n}_{t,s}$, $\theta \in \Theta$, are identically distributed. Therefore, we obtain that for all $t \in [0, T]$, $s \in [t, T]$, $x \in \mathbb{R}^d$ it holds that $Z^{\theta,x,n}_{t,s}$, $\theta \in \Theta$, are identically distributed. Combining this with items (iii)-(iv) and [59, Corollary 2.5] demonstrates that for all $\theta \in \Theta$ it holds that $V^\theta_{n,m,M,Q}$, $\theta \in \Theta$, are identically distributed. This implies that $m+1 \in A_4$. Induction hence establishes item (vi). The proof of Lemma 3.1.4 is thus complete. □

3.1.4 Preliminary results for Euler-Maruyama approximations

The following lemma shows that the Euler-Maruyama approximations in (3.13), (3.14), and (3.15) can be written as Itô processes.

Lemma 3.1.5. *Assume Setting 3.1.1 and let $n \in \mathbb{N}$. Then*

(i) it holds for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ that

$$Y^{\theta,x,n}_{t,s} = x + \int_t^s \mu(Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})\,\mathrm{d}r + \int_t^s \sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})\,\mathrm{d}W^\theta_r, \tag{3.50}$$

(ii) it holds for all $\theta \in \Theta$, $i,j \in \{1,2,\ldots,d\}$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}(D^{\theta,x,n}_{t,s})_{ij} = \delta_{ij} + \sum_{k=1}^d \Bigg[\int_t^s \frac{\partial \mu_i}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})_{kj}\,\mathrm{d}r \\ + \sum_{m=1}^d \bigg[\int_t^s \frac{\partial \sigma_{im}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})_{kj}\,\mathrm{d}W^{\theta,m}_r \bigg]\Bigg],\end{aligned} \tag{3.51}$$

and

(iii) it holds for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ that

$$Z^{\theta,x,n}_{t,s} = \begin{pmatrix} 1 \\ \frac{1}{s-t}\int_t^s (\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}}))^{-1} D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}}\,\mathrm{d}W^\theta_r \end{pmatrix}. \tag{3.52}$$

Proof of Lemma 3.1.5. First observe that (3.13) implies that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}&Y^{\theta,x,n}_{t,s} \\ &= x + \sum_{\substack{k \in \mathbb{N}_0 :\\ t \leqslant \frac{(k+1)(T-t)}{n},\, \frac{k(T-t)}{n} < s}} \Big(\mu(Y^{\theta,x,n}_{t,\max\{t,\frac{k(T-t)}{n}\}})(\min\{s,\tfrac{(k+1)(T-t)}{n}\} - \max\{t,\tfrac{k(T-t)}{n}\}) \\ &\quad + \sigma(Y^{\theta,x,n}_{t,\max\{t,\frac{k(T-t)}{n}\}})(W^\theta_{\min\{s,\frac{(k+1)(T-t)}{n}\}} - W^\theta_{\max\{t,\frac{k(T-t)}{n}\}})\Big) \\ &= x + \sum_{k \in \mathbb{N}_0 :\, t \leqslant \frac{(k+1)(T-t)}{n},\, \frac{k(T-t)}{n} < s} \int_{\max\{t,\frac{k(T-t)}{n}\}}^{\min\{s,\frac{(k+1)(T-t)}{n}\}} \mu(Y^{\theta,x,n}_{t,\max\{t,\frac{k(T-t)}{n}\}})\,\mathrm{d}r \\ &\quad + \int_{\max\{t,\frac{k(T-t)}{n}\}}^{\min\{s,\frac{(k+1)(T-t)}{n}\}} \sigma(Y^{\theta,x,n}_{t,\max\{t,\frac{k(T-t)}{n}\}})\,\mathrm{d}W^\theta_r \\ &= x + \int_t^s \mu(Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})\,\mathrm{d}r + \int_t^s \sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})\,\mathrm{d}W^\theta_r.\end{aligned} \tag{3.53}$$

This establishes item (i). Furthermore, note that (3.14) ensures that for all $\theta \in \Theta$,

$i, j \in \{1, 2, \ldots, d\}$, $t \in [0, T]$, $s \in [t, T]$, $x \in \mathbb{R}^d$ it holds that

$$
\begin{aligned}
(D^{\theta,x,n}_{t,s})_{ij} &= \delta_{ij} + \sum_{\substack{k \in \mathbb{N}_0 : \\ t \leqslant \frac{(k+1)(T-t)}{n}, \frac{k(T-t)}{n} < s}} \Bigg(\sum_{l=1}^{d} \bigg[\frac{\partial \mu_i}{\partial x_l} (Y^{\theta,x,n}_{t,\max\{t, \frac{k(T-t)}{n}\}}) (D^{\theta,x,n}_{t,\max\{t, \frac{k(T-t)}{n}\}})_{lj} \\
&\quad \cdot (\min\{s, \tfrac{(k+1)(T-t)}{n}\} - \max\{t, \tfrac{k(T-t)}{n}\}) \\
&\quad + \sum_{m=1}^{d} \bigg[\frac{\partial \sigma_{im}}{\partial x_l} (Y^{\theta,x,n}_{t,\max\{t, \frac{k(T-t)}{n}\}}) (D^{\theta,x,n}_{t,\max\{t, \frac{k(T-t)}{n}\}})_{lj} \\
&\quad \cdot (W^{\theta,m}_{\min\{s, \frac{(k+1)(T-t)}{n}\}} - W^{\theta,m}_{\max\{t, \frac{k(T-t)}{n}\}}) \bigg] \bigg] \Bigg) \\
&= \delta_{ij} + \sum_{\substack{k \in \mathbb{N}_0 : \\ t \leqslant \frac{(k+1)(T-t)}{n}, \frac{k(T-t)}{n} < s}} \Bigg(\sum_{l=1}^{d} \bigg[\int_{\max\{t, \frac{k(T-t)}{n}\}}^{\min\{s, \frac{(k+1)(T-t)}{n}\}} \frac{\partial \mu_i}{\partial x_l} (Y^{\theta,x,n}_{t,\max\{t, \frac{k(T-t)}{n}\}}) \\
&\quad \cdot (D^{\theta,x,n}_{t,\max\{t, \frac{k(T-t)}{n}\}})_{lj} \, \mathrm{d}r \\
&\quad + \sum_{m=1}^{d} \bigg[\int_{\max\{t, \frac{k(T-t)}{n}\}}^{\min\{s, \frac{(k+1)(T-t)}{n}\}} \frac{\partial \sigma_{im}}{\partial x_l} (Y^{\theta,x,n}_{t,\max\{t, \frac{k(T-t)}{n}\}}) (D^{\theta,x,n}_{t,\max\{t, \frac{k(T-t)}{n}\}})_{lj} \, \mathrm{d}W^{\theta,m}_r \bigg] \bigg] \Bigg) \\
&= \delta_{ij} + \sum_{l=1}^{d} \bigg[\int_t^s \frac{\partial \mu_i}{\partial x_l} (Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}}) (D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})_{lj} \, \mathrm{d}r \\
&\quad + \sum_{m=1}^{d} \bigg[\int_t^s \frac{\partial \sigma_{im}}{\partial x_l} (Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}}) (D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})_{lj} \, \mathrm{d}W^{\theta,m}_r \bigg] \bigg].
\end{aligned}
\tag{3.54}
$$

This establishes item (ii). Next observe that for all $\theta \in \Theta$, $t \in [0, T]$, $s \in (t, T]$, $x \in \mathbb{R}^d$ it holds that

$$
\begin{aligned}
&\sum_{k \in \mathbb{N}_0 : \, t \leqslant \frac{(k+1)(T-t)}{n}, \frac{k(T-t)}{n} < s} (\sigma(Y^{\theta,x,n}_{t,\max\{t, \frac{k(T-t)}{n}\}}))^{-1} D^{\theta,x,n}_{t,\max\{t, \frac{k(T-t)}{n}\}} \\
&\quad \cdot (W^{\theta}_{\min\{s, \frac{(k+1)(T-t)}{n}\}} - W^{\theta}_{\max\{t, \frac{k(T-t)}{n}\}})) \\
&= \sum_{k \in \mathbb{N}_0 : \, t \leqslant \frac{(k+1)(T-t)}{n}, \frac{k(T-t)}{n} < s} \int_{\max\{t, \frac{k(T-t)}{n}\}}^{\min\{s, \frac{(k+1)(T-t)}{n}\}} (\sigma(Y^{\theta,x,n}_{t,\max\{t, \frac{k(T-t)}{n}\}}))^{-1} D^{\theta,x,n}_{t,\max\{t, \frac{k(T-t)}{n}\}} \, \mathrm{d}W^{\theta}_r \\
&= \int_t^s (\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}}))^{-1} D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}} \, \mathrm{d}W^{\theta}_r.
\end{aligned}
\tag{3.55}
$$

Combining this with (3.15) and the assumption that for all $\theta \in \Theta$, $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that $Z^{\theta,x,n}_{t,t} = \mathbf{e}_1$ establish item (iii). The proof of Lemma 3.1.5 is thus complete. □

Next we prove in the following lemma that the approximation scheme defined in (3.17) is differentiable in the starting point $x \in \mathbb{R}^d$ and that the derivative process agrees with the process in (3.14).

Lemma 3.1.6. *Assume Setting 3.1.1, let $n \in \mathbb{N}$, and let $t \in [0, T]$. Then*

(i) it holds for all $\theta \in \Theta$, $s \in [t, T]$, $\omega \in \Omega$ that $(\mathbb{R}^d \ni x \mapsto Y^{\theta,x,n}_{t,s}(\omega) \in \mathbb{R}^d) \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ and

(ii) it holds for all $\theta \in \Theta$, $s \in [t, T]$, $\omega \in \Omega$ that $\frac{\partial}{\partial x} Y^{\theta,x,n}_{t,s}(\omega) = D^{\theta,x,n}_{t,s}(\omega)$.

Proof of Lemma 3.1.6. Throughout this proof let $k_0 \in \mathbb{N}$ satisfy $k_0 = \inf\{k \in \mathbb{N} \colon ([t,T] \cap \frac{k(T-t)}{n}) \neq \varnothing\}$. We prove that for all $k \in \mathbb{N} \cap [k_0, \infty)$, $\theta \in \Theta$, $s \in (\max\{t, \frac{(k-1)(T-t)}{n}\}, \min\{T, \frac{k(T-t)}{n}\}]$, $\omega \in \Omega$ it holds that

$$(\mathbb{R}^d \ni x \mapsto Y^{\theta,x,n}_{t,s}(\omega) \in \mathbb{R}^d) \in C^1(\mathbb{R}^d, \mathbb{R}^d) \tag{3.56}$$

and

$$\frac{\partial}{\partial x} Y^{\theta,x,n}_{t,s}(\omega) = D^{\theta,x,n}_{t,s}(\omega) \tag{3.57}$$

by induction on $k \in \mathbb{N} \cap [k_0, \infty)$. First observe that $\max\{t, \frac{(k_0-1)(T-t)}{n}\} = t$ and $\min\{T, \frac{k_0(T-t)}{n}\} = \frac{k_0(T-t)}{n}$. Furthermore, note that the fact that for all $\theta \in \Theta$, $x \in \mathbb{R}^d$, $\omega \in \Omega$ it holds that $Y^{\theta,x,n}_{t,t}(\omega) = x$ ensures that for all $\omega \in \Omega$ it holds that $(\mathbb{R}^d \ni x \mapsto Y^{\theta,x,n}_{t,t}(\omega) \in \mathbb{R}^d) \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ and $\frac{\partial}{\partial x} Y^{\theta,x,n}_{t,t}(\omega) = I_d = D^{\theta,x,n}_{t,t}(\omega)$. Moreover, observe that (3.13) and the fact that for all $\theta \in \Theta$, $x \in \mathbb{R}^d$, $\omega \in \Omega$ it holds that $Y^{\theta,x,n}_{t,t}(\omega) = x$ imply that for all $\theta \in \Theta$, $s \in (t, \frac{k_0(T-t)}{n}]$, $x \in \mathbb{R}^d$, $\omega \in \Omega$ it holds that

$$Y^{\theta,x,n}_{t,s}(\omega) = x + \mu(x)(s-t) + \sigma(x)(W^\theta_s(\omega) - W^\theta_t(\omega)). \tag{3.58}$$

The assumption that $\mu \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ and $\sigma \in C^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$ hence demonstrate that for all $i, j \in \{1, 2, \ldots, d\}$, $\theta \in \Theta$, $s \in (t, \frac{k_0(T-t)}{n}]$, $\omega \in \Omega$ it holds that $(\mathbb{R}^d \ni x \mapsto Y^{\theta,x,n}_{t,s}(\omega) \in \mathbb{R}^d) \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ and

$$\begin{aligned}
&\frac{\partial}{\partial x_j}(Y^{\theta,x,n}_{t,s}(\omega))_i \\
&= \delta_{ij} + \sum_{k=1}^d \left[\frac{\partial \mu_i}{\partial x_k}(x)\delta_{kj}(s-t) + \sum_{m=1}^d \left[\frac{\partial \sigma_{im}}{\partial x_j}(x)\delta_{kj}(W^{\theta,m}_s(\omega) - W^{\theta,m}_t(\omega)) \right] \right] \\
&= (D^{\theta,x,n}_{t,s}(\omega))_{ij}.
\end{aligned} \tag{3.59}$$

This proves (3.56) in the base case $k = k_0$. For the induction step assume that (3.56) and (3.57) hold for $k \in \mathbb{N}_0 \cap [k_0, \infty)$. Note that (3.13) ensures that for all $t \in [0,T]$, $s \in (\frac{k(T-t)}{n}, \min\{T, \frac{(k+1)(T-t)}{n}\}]$, $x \in \mathbb{R}^d$, $\omega \in \Omega$ it holds that

$$\begin{aligned}
&Y^{\theta,x,n}_{t,s}(\omega) \\
&= Y^{\theta,x,n}_{t,\frac{k(T-t)}{n}}(\omega) + \mu(Y^{\theta,x,n}_{t,\frac{k(T-t)}{n}}(\omega))(s - \tfrac{k(T-t)}{n}) \\
&\quad + \sigma(Y^{\theta,x,n}_{t,\frac{k(T-t)}{n}}(\omega))(W^\theta_s(\omega) - W^\theta_{\frac{k(T-t)}{n}}(\omega)).
\end{aligned} \tag{3.60}$$

Combining this with the assumption that $\mu \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ and $\sigma \in C^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$ and the fact that for all $\theta \in \Theta$, $s \in (\frac{(k-1)(T-t)}{n}, \min\{T, \frac{k(T-t)}{n}\}]$, $\omega \in \Omega$ it holds that $(\mathbb{R}^d \ni x \mapsto Y^{\theta,x,n}_{t,s}(\omega) \in \mathbb{R}^d) \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ with $\frac{\partial}{\partial x} Y^{\theta,x,n}_{t,s}(\omega) = D^{\theta,x,n}_{t,s}(\omega)$ shows that for all $i, j \in \{1, 2, \ldots, d\}$, $\theta \in \Theta$, $s \in (\frac{k(T-t)}{n}, \min\{T, \frac{(k+1)(T-t)}{n}\}]$, $\omega \in \Omega$ it holds that $(\mathbb{R}^d \ni x \mapsto$

$Y_{t,s}^{\theta,x,n}(\omega) \in \mathbb{R}^d) \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ and

$$\begin{aligned}
\frac{\partial}{\partial x_j}(Y_{t,s}^{\theta,x,n}(\omega))_i &= \frac{\partial}{\partial x_j}(Y_{t,\frac{k(T-t)}{n}}^{\theta,x,n}(\omega))_i \\
&\quad + \sum_{k=1}^d \left[\frac{\partial \mu_i}{\partial x_k}\left(Y_{t,\frac{k(T-t)}{n}}^{\theta,x,n}(\omega)\right)\left(\frac{\partial}{\partial x_j}(Y_{t,\frac{k(T-t)}{n}}^{\theta,x,n})_k(\omega)\right)\left(s - \tfrac{k(T-t)}{n}\right) \right. \\
&\quad + \sum_{m=1}^d \left[\frac{\partial \sigma_{im}}{\partial x_k}\left(Y_{t,\frac{k(T-t)}{n}}^{\theta,x,n}(\omega)\right)\left(\frac{\partial}{\partial x_j}(Y_{t,\frac{k(T-t)}{n}}^{\theta,x,n}(\omega))_k\right) \right. \\
&\quad \left.\left. \cdot \left(W_s^{\theta,m}(\omega) - W_{\frac{k(T-t)}{n}}^{\theta,m}(\omega)\right)\right]\right] \\
&= \left(D_{t,\frac{k(T-t)}{n}}^{\theta,x,n}(\omega)\right)_{ij} \\
&\quad + \sum_{k=1}^d \left[\frac{\partial \mu_i}{\partial x_k}\left(Y_{t,\frac{k(T-t)}{n}}^{\theta,x,n}(\omega)\right)\left(D_{t,\frac{k(T-t)}{n}}^{\theta,x,n}(\omega)\right)_{kj}\left(s - \tfrac{k(T-t)}{n}\right) \right. \\
&\quad + \sum_{m=1}^d \left[\frac{\partial \sigma_{im}}{\partial x_k}\left(Y_{t,\frac{k(T-t)}{n}}^{\theta,x,n}(\omega)\right)\left(D_{t,\frac{k(T-t)}{n}}^{\theta,x,n}(\omega)\right)_{kj} \right. \\
&\quad \left.\left. \cdot \left(W_s^{\theta,m}(\omega) - W_{\frac{k(T-t)}{n}}^{\theta,m}(\omega)\right)\right]\right] = \left(D_{t,s}^{\theta,x,n}(\omega)\right)_{ij}.
\end{aligned} \tag{3.61}$$

Induction therefore proves (3.56) and (3.57). The proof of Lemma 3.1.6 is thus complete. □

The following Lemma establishes moment estimates for the difference of the Euler-Maruyama approximations in (3.13) and (3.14) and their starting points. It allows us to control these differences uniformly in the discretization step size.

Lemma 3.1.7. *Assume Setting 3.1.1, let $n \in \mathbb{N}$, and let $p \in [2, \infty)$. Then*

(i) for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\sup_{r\in[t,s]} \mathbb{E}\big[\|Y_{t,r}^{\theta,x,n} - x\|^p\big] \leqslant \gamma_x^{(p)}(s-t)\exp\Big(\big(\tfrac{p(p-1)}{2} + 2^{p-1}pK^p\big)(s-t)\Big), \tag{3.62}$$

(ii) for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\|Y_{t,T}^{\theta,x,n} - Y_{t,s}^{\theta,n,x}\|^4\right]\right)^{\frac{1}{4}} \\
&\leqslant \sqrt{T-s}(\sqrt{T}+\sqrt{6})\left(2K\left(\gamma_x^{(4)}(T-s)\exp\left(6+32K^4T\right)\right)^{\frac{1}{4}} + \gamma_x^{(1)}\right),
\end{aligned} \tag{3.63}$$

(iii) for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\sup_{r\in[t,s]} \mathbb{E}\Big[\big\|D_{t,r}^{\theta,x,n} - I_d\big\|^2\Big] \leqslant 3K(s-t)\exp\Big(5K(s-t)\Big), \tag{3.64}$$

and

(iv) for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\sup_{r\in[t,s]} \mathbb{E}\Big[\big\|D_{t,r}^{\theta,x,n} - I_d\big\|^4\Big] \leqslant 36K^2(s-t)(1+(s-t))\exp\Big(160K^2(s-t)(1+(s-t))\Big). \tag{3.65}$$

Proof of Lemma 3.1.7. First note that item (ii) of Lemma 3.1.2, the triangle inequality, and the fact that for all $a, b \in [0, \infty)$ it holds that $(a+b)^p \leqslant 2^{p-1}(a^p + b^p)$ ensure that for all $\theta \in \Theta$, $t \in [0, T]$, $s \in [t, T]$, $x \in \mathbb{R}^d$ it holds a.s. that

$$\begin{aligned}
&\|\mu(Y^{\theta,x,n}_{t,s})\|^p + (p-1)\|\sigma(Y^{\theta,x,n}_{t,s})\|_F^p \\
&\leqslant 2^{p-1}\left(\|\mu(Y^{\theta,x,n}_{t,s}) - \mu(x)\|^p + \|\mu(x)\|^p\right) \\
&\quad + 2^{p-1}(p-1)\left(\|\sigma(Y^{\theta,x,n}_{t,s}) - \sigma(x)\|_F^p + \|\sigma(x)\|_F^p\right) \\
&\leqslant 2^{p-1}\left(K^p\|Y^{\theta,x,n}_{t,s} - x\|^p + \|\mu(x)\|^p\right) \\
&\quad + 2^{p-1}(p-1)\left(K^p\|Y^{\theta,x,n}_{t,s} - x\|^p + \|\sigma(x)_F\|^p\right) \\
&= 2^{p-1}pK^p\|Y^{\theta,x,n}_{t,s} - x\| + 2^{p-1}\left(\|\mu(x)\|^p + (p-1)\|\sigma(x)\|_F^p\right).
\end{aligned} \tag{3.66}$$

Next observe that for all $i, j \in \{1, 2, \ldots, d\}$, $x \in \mathbb{R}^d$ it holds that $\frac{\partial}{\partial x_i}\|x\|^p = p\|x\|^{p-2}x_i$ and

$$\frac{\partial^2}{\partial x_i \partial x_j}\|x\|^p = \mathbb{1}_{\mathbb{R}\setminus\{0\}}(x)p(p-2)\|x\|^{p-4}x_i x_j + \mathbb{1}_j(i)p\|x\|^{p-2}. \tag{3.67}$$

Combining this with Itô's formula and item (i) of Lemma 3.1.5 demonstrates that for all $\theta \in \Theta$, $t \in [0, T]$, $s \in [t, T]$, $x \in \mathbb{R}^d$ it holds a.s. that

$$\begin{aligned}
&\|Y^{\theta,x,n}_{t,s} - x\|^p \\
&= \int_t^s p\|Y^{\theta,x,n}_{t,r} - x\|^{p-2}\langle Y^{\theta,x,n}_{t,r} - x, \mu(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\rangle \,\mathrm{d}r \\
&\quad + \int_t^s p\|Y^{\theta,x,n}_{t,r} - x\|^{p-2}\langle Y^{\theta,x,n}_{t,r} - x, \sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\,\mathrm{d}W^\theta_r\rangle \\
&\quad + \frac{1}{2}\sum_{i,j=1}^d \int_t^s \Big[\mathbb{1}_{\mathbb{R}\setminus\{0\}}(Y^{\theta,x,n}_{t,r} - x)p(p-2)\|Y^{\theta,x,n}_{t,r} - x\|^{p-4} \\
&\quad \cdot (Y^{\theta,x,n}_{t,r} - x)_i(Y^{\theta,x,n}_{t,r} - x)_j \\
&\quad + \mathbb{1}_j(i)p\|Y^{\theta,x,n}_{t,r} - x\|^{p-2}\Big]\sum_{k=1}^d \sigma_{ik}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\sigma_{jk}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\,\mathrm{d}r.
\end{aligned} \tag{3.68}$$

The Cauchy-Schwarz inequality, Young's inequality, item (i) of Lemma 3.1.5, (3.8), and

(3.66) therefore imply that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds a.s. that

$$\begin{aligned}
&\|Y^{\theta,x,n}_{t,s} - x\|^p \\
&\leqslant \int_t^s p\|Y^{\theta,x,n}_{t,r} - x\|^{p-1}\|\mu(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\|\,\mathrm{d}r \\
&\quad + \int_t^s p\|Y^{\theta,x,n}_{t,r} - x\|^{p-1}\|\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\|_F\,\mathrm{d}W^\theta_r \\
&\quad + \int_t^s \frac{p(p-1)}{2}\|Y^{\theta,x,n}_{t,r} - x\|^{p-2}\|\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\|_F^2\,\mathrm{d}r \\
&\leqslant \int_t^s (p-1)\|Y^{\theta,x,n}_{t,r} - x\|^p + \|\mu(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\|^p\,\mathrm{d}r \\
&\quad + \int_t^s \frac{(p-1)(p-2)}{2}\|Y^{\theta,x,n}_{t,r} - x\|^p + (p-1)\|\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\|_F^p\,\mathrm{d}r \\
&\quad + \int_t^s p\|Y^{\theta,x,n}_{t,r} - x\|^{p-1}\|\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\|_F\,\mathrm{d}W^\theta_r \\
&\leqslant \gamma^{(p)}_x(s-t) + \int_t^s \frac{p(p-1)}{2}\|Y^{\theta,x,n}_{t,r} - x\|^p + 2^{p-1}pK^p\|Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}} - x\|^p\,\mathrm{d}r \\
&\quad + \int_t^s p\|Y^{\theta,x,n}_{t,r} - x\|^{p-1}\|\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\|_F\,\mathrm{d}W^\theta_r.
\end{aligned} \tag{3.69}$$

In the next step for every $k \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ let $\tau^{t,x}_k \colon \Omega \to [t,T]$ satisfy $\tau^{t,x}_k = \inf(\{s \in [t,T] \colon \|Y^{\theta,x,n}_{t,s} - x\| \geqslant k\} \cup \{T\})$. This and item (ii) of Lemma 3.1.2 ensure that for all $k \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\mathbb{E}\bigg[\int_t^{\min\{s,\tau^{t,x}_k\}} p\|Y^{\theta,x,n}_{t,\min\{r,\tau^{t,x}_k\}} - x\|^{p-1}\|\sigma(Y^{\theta,x,n}_{t,\max\{t,\min\{\lfloor r\rfloor_t,\tau^{t,x}_k\}\}})\|\,\mathrm{d}W^\theta_r\bigg] = 0. \tag{3.70}$$

Combining this with Fubini's theorem and (3.68) shows that for all $k \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\sup_{\eta\in[t,s]} \mathbb{E}\Big[\|Y^{\theta,x,n}_{t,\min\{\eta,\tau^{t,x}_k\}} - x\|^p\Big] \\
&\leqslant \mathbb{E}\bigg[\gamma^{(p)}_x(s-t) + \int_t^{\min\{s,\tau^{t,x}_k\}} \frac{p(p-1)}{2}\|Y^{\theta,x,n}_{t,\min\{r,\tau^{t,x}_k\}} - x\|^p \\
&\qquad + 2^{p-1}pK^p\|Y^{\theta,x,n}_{t,\max\{t,\min\{\lfloor r\rfloor_t,\tau^{t,x}_k\}\}} - x\|^p\,\mathrm{d}r\bigg] \\
&\leqslant \gamma^{(p)}_x(s-t) + \Big(\frac{p(p-1)}{2} + 2^{p-1}pK^p\Big)\int_t^s \sup_{\eta\in[t,r]} \mathbb{E}\big[\|Y^{\theta,x,n}_{t,\min\{\eta,\tau^{t,x}_k\}} - x\|^p\big]\,\mathrm{d}r.
\end{aligned} \tag{3.71}$$

Fatou's Lemma, the Gronwall inequality, and the fact that for all $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that $\lim_{k\to\infty} \tau^{t,x}_k = T$ hence prove that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\sup_{r\in[t,s]} \mathbb{E}\big[\|Y^{\theta,x,n}_{t,r} - x\|^p\big] = \sup_{r\in[t,s]} \mathbb{E}\big[\lim_{k\to\infty}\|Y^{\theta,x,n}_{t,\min\{r,\tau^{t,x}_k\}} - x\|^p\big] \\
&\leqslant \sup_{k\in\mathbb{N}}\sup_{r\in[t,s]} \mathbb{E}\big[\|Y^{\theta,x,n}_{t,\min\{r,\tau^{t,x}_k\}} - x\|^p\big] \\
&\leqslant \gamma^{(p)}_x(s-t)\exp\Big(\big(\frac{p(p-1)}{2} + 2^{p-1}pK^p\big)(s-t)\Big).
\end{aligned} \tag{3.72}$$

This establishes item (i). Next observe that item (i) of Lemma 3.1.5 ensures that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$Y_{t,T}^{\theta,x,n} - Y_{t,s}^{\theta,x,n} = \int_s^T \mu(Y_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n}) \,\mathrm{d}r + \int_s^T \sigma(Y_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n}) \,\mathrm{d}W_r^\theta. \tag{3.73}$$

The triangle inequality and the Burkholder-Davis-Gundy inequality hence demonstrate that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\|Y_{t,T}^{\theta,x,n} - Y_{t,s}^{\theta,n,x}\|^4\right]\right)^{\frac{1}{4}} \\
&\leqslant \left(\mathbb{E}\left[\left\|\int_s^T \mu(Y_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n}) \,\mathrm{d}r\right\|^4\right]\right)^{\frac{1}{4}} + \left(\mathbb{E}\left[\left\|\int_s^T \sigma(Y_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n}) \,\mathrm{d}W_r^\theta\right\|^4\right]\right)^{\frac{1}{4}} \\
&\leqslant \int_s^T \left(\mathbb{E}\left[\|\mu(Y_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n})\|^4\right]\right)^{\frac{1}{4}} \mathrm{d}r \\
&\quad + \sqrt{6}\left(\int_s^T \left(\mathbb{E}\left[\left\|\sigma(Y_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n})\right\|_F^4\right]\right)^{\frac{1}{2}} \mathrm{d}r\right)^{\frac{1}{2}} \\
&\leqslant (T-s) \sup_{u\in[s,T]} \left(\mathbb{E}\left[\|\mu(Y_{t,u}^{\theta,x,n}) \,\mathrm{d}r\|^4\right]\right)^{\frac{1}{4}} \\
&\quad + \sqrt{6(T-s)} \sup_{u\in[s,T]} \left(\mathbb{E}\left[\|\sigma(Y_{t,u}^{\theta,x,n})\|_F^4\right]\right)^{\frac{1}{4}}.
\end{aligned} \tag{3.74}$$

Furthermore, note that the triangle inequality, item (ii) of Lemma 3.1.2, item (i) (applied with $p \curvearrowleft 4$), and (3.8) prove that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\sup_{u\in[s,T]} \left(\mathbb{E}\left[\|\mu(Y_{t,u}^{\theta,x,n})\|^4\right]\right)^{\frac{1}{4}} + \sup_{u\in[s,T]} \left(\mathbb{E}\left[\|\sigma(Y_{t,u}^{\theta,x,n})\|_F^4\right]\right)^{\frac{1}{4}} \\
&\leqslant 2K \sup_{u\in[s,T]} \left(\mathbb{E}\left[\|Y_{t,u}^{\theta,x,n} - x\|^4\right]\right)^{\frac{1}{4}} + \|\mu(x)\| + \|\sigma(x)\| \\
&\leqslant 2K \left(\gamma_x^{(4)}(T-s)\exp\left(6+32K^4T\right)\right)^{\frac{1}{4}} + \gamma_x^{(1)}.
\end{aligned} \tag{3.75}$$

Combining this with (3.74) shows that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\|Y_{t,T}^{\theta,x,n} - Y_{t,s}^{\theta,n,x}\|^4\right]\right)^{\frac{1}{4}} \\
&\leqslant \sqrt{T-s}(\sqrt{T}+\sqrt{6})\left(2K\left(\gamma_x^{(4)}(T-s)\exp\left(6+32K^4T\right)\right)^{\frac{1}{4}} + \gamma_x^{(1)}\right).
\end{aligned} \tag{3.76}$$

This establishes item (ii). Next note that Itô's formula, item (ii) of Lemma 3.1.5, and (3.67) imply that for all $j \in \{1,2,\ldots,d\}$, $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds a.s.

that

$$
\begin{aligned}
&\left\|\frac{\partial}{\partial x_j} Y^{\theta,x,n}_{t,s} - \mathbf{e}_j\right\|^2\\
&= \int_t^s 2\Big\langle \frac{\partial}{\partial x_j} Y^{\theta,x,n}_{t,r} - \mathbf{e}_j, \sum_{k=1}^d \frac{\partial \mu}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\Big(\frac{\partial}{\partial x_j} Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}\Big)_k\Big\rangle \,\mathrm{d}r\\
&\quad + \int_t^s 2\Big\langle \frac{\partial}{\partial x_j} Y^{\theta,x,n}_{t,r} - \mathbf{e}_j, \sum_{k=1}^d \frac{\partial \sigma}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\Big(\frac{\partial}{\partial x_j} Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}\Big)_k \,\mathrm{d}W^\theta_r\Big\rangle\\
&\quad + \frac{1}{2}\sum_{a,i=1}^d \int_t^s 2\mathbb{1}_i(a) \sum_{k=1}^d \Bigg(\sum_{l=1}^d \frac{\partial \sigma_{ik}}{\partial x_l}(Y^{\theta,x,n}_{\max\{t,\lfloor r\rfloor_t\}})\Big(\frac{\partial}{\partial x_j} Y^{\theta,x,n}_{\max\{t,\lfloor r\rfloor_t\}}\Big)_l\Bigg)\\
&\quad \cdot \Bigg(\sum_{l=1}^d \frac{\partial \sigma_{ak}}{\partial x_l}(Y^{\theta,x,n}_{\max\{t,\lfloor r\rfloor_t\}})\Big(\frac{\partial}{\partial x_j} Y^{\theta,x,n}_{\max\{t,\lfloor r\rfloor_t\}}\Big)_l\Bigg)\,\mathrm{d}r\\
&= \int_t^s 2\sum_{i=1}^d \Big[\frac{\partial}{\partial x_j}(Y^{\theta,x,n}_{t,r})_i - \delta_{ij}\Big]\sum_{k=1}^d \frac{\partial \mu_i}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\Big(\frac{\partial}{\partial x_j} Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}\Big)_k \,\mathrm{d}r\\
&\quad + \int_t^s 2\sum_{i=1}^d \Big[\frac{\partial}{\partial x_j}(Y^{\theta,x,n}_{t,r})_i - \delta_{ij}\Big]\sum_{l,k=1}^d \frac{\partial \sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\Big(\frac{\partial}{\partial x_j} Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}\Big)_k \,\mathrm{d}W^{\theta,l}_r\\
&\quad + \int_t^s \sum_{i,k=1}^d \Bigg(\sum_{l=1}^d \frac{\partial \sigma_{ik}}{\partial x_l}(Y^{\theta,x,n}_{\max\{t,\lfloor r\rfloor_t\}})\Big(\frac{\partial}{\partial x_j} Y^{\theta,x,n}_{\max\{t,\lfloor r\rfloor_t\}}\Big)_l\Bigg)^2 \,\mathrm{d}r.
\end{aligned}
\tag{3.77}
$$

Hence, we obtain that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds a.s. that

$$
\begin{aligned}
&\left\|D^{\theta,x,n}_{t,s} - I_d\right\|_F^2 = \sum_{j=1}^d \left\|\frac{\partial}{\partial x_j} Y^{\theta,x,n}_{t,s} - \mathbf{e}_j\right\|^2\\
&= 2\int_t^s \sum_{i,j,k=1}^d \Big[(D^{\theta,x,n}_{t,r})_{ij} - \delta_{ij}\Big]\frac{\partial \mu_i}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})_{kj}\,\mathrm{d}r\\
&\quad + 2\int_t^s \sum_{i,j,k,l=1}^d \Big[(D^{\theta,x,n}_{t,r})_{ij} - \delta_{ij}\Big]\frac{\partial \sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})_{kj}\,\mathrm{d}W^{\theta,l}_r\\
&\quad + \int_t^s \sum_{i,j,k=1}^d \Bigg(\sum_{l=1}^d \frac{\partial \sigma_{ik}}{\partial x_l}(Y^{\theta,x,n}_{\max\{t,\lfloor r\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})_{lj}\Bigg)^2\,\mathrm{d}r.
\end{aligned}
\tag{3.78}
$$

Next note that Young's inequality, (3.7), and the fact that for all $a, b \in \mathbb{R}$ it holds that

$2ab \leqslant a^2 + b^2$ ensure that for all $\theta \in \Theta$, $t \in [0, T]$, $s \in [t, T]$, $x \in \mathbb{R}^d$ it holds a.s. that

$$
\begin{aligned}
&2 \int_t^s \sum_{i,j,k=1}^d \Big[(D_{t,r}^{\theta,x,n})_{ij} - \delta_{ij} \Big] \frac{\partial \mu_i}{\partial x_k} (Y_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n}) (D_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n})_{kj} \, \mathrm{d}r \\
&= 2 \int_t^s \sum_{i,j,k=1}^d \Big[(D_{t,r}^{\theta,x,n})_{ij} - \delta_{ij} \Big] \frac{\partial \mu_i}{\partial x_k} (Y_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n}) \big((D_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n})_{kj} - \delta_{kj} \big) \, \mathrm{d}r \\
&\quad + 2 \int_t^s \sum_{i,j=1}^d \Big[(D_{t,r}^{\theta,x,n})_{ij} - \delta_{ij} \Big] \frac{\partial \mu_i}{\partial x_j} (Y_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n}) \, \mathrm{d}r \\
&\leqslant 2 \int_t^s \sum_{j=1}^d \Big[\max_{l \in \{1,2,\dots,d\}} \Big((D_{t,r}^{\theta,x,n})_{lj} - \delta_{lj} \Big) \Big] \Big[\max_{a \in \{1,2,\dots,d\}} \big((D_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n})_{aj} - \delta_{aj} \big) \Big] \\
&\quad \cdot \sum_{i,k=1}^d \frac{\partial \mu_i}{\partial x_k} (Y_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n}) \, \mathrm{d}r + 2 \int_t^s \big\| D_{t,r}^{\theta,x,n} - I_d \big\|_F \sum_{i,j=1}^d \frac{\partial \mu_i}{\partial x_j} (Y_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n}) \, \mathrm{d}r \\
&\leqslant 2K \int_t^s \sum_{j=1}^d \Big[\max_{l \in \{1,2,\dots,d\}} \Big((D_{t,r}^{\theta,x,n})_{lj} - \delta_{lj} \Big) \Big] \Big[\max_{a \in \{1,2,\dots,d\}} \big((D_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n})_{aj} - \delta_{aj} \big) \Big] \, \mathrm{d}r \qquad (3.79) \\
&\quad + 2K \int_t^s \big\| D_{t,r}^{\theta,x,n} - I_d \big\|_F \, \mathrm{d}r \\
&\leqslant K \int_t^s \sum_{j=1}^d \bigg(\Big[\max_{l \in \{1,2,\dots,d\}} \Big((D_{t,r}^{\theta,x,n})_{lj} - \delta_{lj} \Big)^2 \Big] \\
&\quad + \Big[\max_{a \in \{1,2,\dots,d\}} \big((D_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n})_{aj} - \delta_{aj} \big)^2 \Big] \bigg) \, \mathrm{d}r + K \int_t^s \Big(1 + \big\| D_{t,r}^{\theta,x,n} - I_d \big\|_F^2 \Big) \, \mathrm{d}r \\
&\leqslant K \int_t^s \Big(\big\| D_{t,r}^{\theta,x,n} - I_d \big\|_F^2 + \big\| D_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n} - I_d \big\|_F^2 \Big) \, \mathrm{d}r + K(s-t) \\
&\quad + K \int_t^s \big\| D_{t,r}^{\theta,x,n} - I_d \big\|_F^2 \, \mathrm{d}r \\
&= K(s-t) + K \int_t^s \Big(2 \big\| D_{t,r}^{\theta,x,n} - I_d \big\|_F^2 + \big\| D_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n} - I_d \big\|_F^2 \Big) \, \mathrm{d}r.
\end{aligned}
$$

In addition, observe that item (i) of Lemma 3.1.2 and the fact that for all $a, b \in \mathbb{R}$ it holds that $(a+b)^2 \leqslant 2(a^2 + b^2)$ imply that for all $\theta \in \Theta$, $t \in [0, T]$, $s \in [t, T]$, $x \in \mathbb{R}^d$ it holds a.s. that

$$
\begin{aligned}
&\int_t^s \sum_{i,j,k=1}^d \left(\sum_{l=1}^d \frac{\partial \sigma_{ik}}{\partial x_l}(Y^{\theta,x,n}_{\max\{t,\lfloor r \rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})_{lj} \right)^2 dr \\
&\leqslant 2\int_t^s \sum_{i,j,k=1}^d \left(\sum_{l=1}^d \frac{\partial \sigma_{ik}}{\partial x_l}(Y^{\theta,x,n}_{\max\{t,\lfloor r \rfloor_t\}})\Big((D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})_{lj} - \delta_{lj}\Big) \right)^2 dr \\
&\quad + 2\int_t^s \sum_{i,j,k=1}^d \left(\sum_{l=1}^d \frac{\partial \sigma_{ik}}{\partial x_l}(Y^{\theta,x,n}_{\max\{t,\lfloor r \rfloor_t\}})\delta_{lj} \right)^2 dr \\
&\leqslant 2\int_t^s \sum_{j=1}^d \Big[\max_{a\in\{1,2,\ldots,d\}} (D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})_{aj} - \delta_{aj} \Big)^2 \Big] \sum_{i,k=1}^d \left(\sum_{l=1}^d \frac{\partial \sigma_{ik}}{\partial x_l}(Y^{\theta,x,n}_{\max\{t,\lfloor r \rfloor_t\}}) \right)^2 dr \\
&\quad + 2\int_t^s \sum_{i,k=1}^d \left(\sum_{l=1}^d \frac{\partial \sigma_{ik}}{\partial x_l}(Y^{\theta,x,n}_{\max\{t,\lfloor r \rfloor_t\}}) \right)^2 dr \\
&\leqslant 2K \int_t^s \sum_{j=1}^d \Big[\max_{a\in\{1,2,\ldots,d\}} \Big((D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})_{aj} - \delta_{aj}\Big)^2 \Big] dr + 2K\int_t^s dr \\
&\leqslant 2K(s-t) + 2K\int_t^s \big\| D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}} - I_d \big\|_F^2 \, dr.
\end{aligned} \tag{3.80}
$$

Combining this with (3.78) and (3.79) shows that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds a.s. that

$$
\begin{aligned}
&\Big\| D^{\theta,x,n}_{t,s} - I_d \Big\|_F^2 \\
&\qquad - 2\int_t^s \sum_{i,j,k,l=1}^d \Big[(D^{\theta,x,n}_{t,r})_{ij} - \delta_{ij} \Big] \frac{\partial \sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})_{kj} \, dW^{\theta,l}_r \\
&= 2\int_t^s \sum_{i,j,k=1}^d \Big[(D^{\theta,x,n}_{t,r})_{ij} - \delta_{ij} \Big] \frac{\partial \mu_i}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})_{kj} \, dr \\
&\qquad + \int_t^s \sum_{i,j,k=1}^d \left(\sum_{l=1}^d \frac{\partial \sigma_{ik}}{\partial x_l}(Y^{\theta,x,n}_{\max\{t,\lfloor r \rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}})_{lj} \right)^2 dr \\
&\leqslant 3K(s-t) + K\int_t^s \Big(2\big\| D^{\theta,x,n}_{t,r} - I_d \big\|_F^2 + 3\big\| D^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}} - I_d \big\|_F^2 \Big) \, dr.
\end{aligned} \tag{3.81}
$$

In the next step for every $k \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ let $\mathfrak{t}^{t,x}_k \colon \Omega \to [t,T]$ satisfy $\mathfrak{t}^{t,x}_k = \inf(\{s \in [t,T] \colon \|D^{\theta,x,n}_{t,s}\| \geqslant k\} \cup \{T\})$. Observe that item (ii) of Lemma 3.1.2 implies that for all $k \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$
\begin{aligned}
&\mathbb{E}\Bigg[\int_t^{\min\{s,\mathfrak{t}^{t,x}_k\}} \sum_{i,j,l,m=1}^d \Big[(D^{\theta,x,n}_{t,\min\{r,\mathfrak{t}^{t,x}_k\}})_{ij} - \delta_{ij} \Big] \\
&\qquad \cdot \frac{\partial \sigma_{im}}{\partial x_l}(Y^{\theta,x,n}_{t,\max\{t,\min\{\lfloor r \rfloor_t,\mathfrak{t}^{t,x}_k\}\}})(D^{\theta,x,n}_{t,\max\{t,\min\{\lfloor r \rfloor_t,\mathfrak{t}^{t,x}_k\}\}})_{lj} \, dW^{\theta,m}_r \Bigg] = 0.
\end{aligned} \tag{3.82}
$$

This, Fubini's theorem, and (3.81) show that for all $k \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\sup_{s\in[t,T]} \mathbb{E}\Big[\Big\|D^{\theta,x,n}_{t,\max\{t,\min\{s,\mathfrak{t}^{t,x}_k\}\}} - I_d\Big\|^2_F\Big] \\
&\leqslant \mathbb{E}\bigg[3K(s-t) \\
&\quad + K\int_t^{\max\{s,\mathfrak{t}^t_k\}} \Big(2\big\|D^{\theta,x,n}_{t,\min\{r,\mathfrak{t}^{t,x}_k\}} - I_d\big\|^2_F + 3\big\|D^{\theta,x,n}_{t,\max\{t,\min\{\lfloor r\rfloor_t,\mathfrak{t}^{t,x}_k\}\}} - I_d\big\|^2_F\Big)\,\mathrm{d}r\bigg] \\
&\leqslant 3K(s-t) + 5K\int_t^{\max\{s,\mathfrak{t}^t_k\}} \mathbb{E}\Big[\sup_{\nu\in[t,r]} \big\|D^{\theta,x,n}_{t,\nu} - I_d\big\|^2_F\Big]\,\mathrm{d}r.
\end{aligned} \tag{3.83}$$

Fatou's lemma, Gronwall's inequality, and the fact that for all $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that $\lim_{k\to\infty} \mathfrak{t}^{t,x}_k = T$ therefore prove that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\sup_{r\in[t,s]} \mathbb{E}\Big[\big\|D^{\theta,x,n}_{t,r} - I_d\big\|^2_F\Big] = \sup_{r\in[t,s]} \mathbb{E}\Big[\lim_{k\to\infty} \big\|D^{\theta,x,n}_{t,\min\{r,\mathfrak{t}^{t,x}_k\}} - I_d\big\|^2_F\Big] \\
&\leqslant \sup_{k\in\mathbb{N}} \sup_{r\in[t,s]} \mathbb{E}\Big[\big\|D^{\theta,x,n}_{t,\min\{r,\mathfrak{t}^{t,x}_k\}} - I_d\big\|^2_F\Big] \leqslant 3K(s-t)\exp\Big(5K(s-t)\Big).
\end{aligned} \tag{3.84}$$

This establishes item (iii). Next note that (3.81), the Cauchy-Schwarz inequality for square-integrable functions, and the fact that for all $x_1, x_2, x_3, x_4 \in \mathbb{R}$ it holds that $(\sum_{i=1}^4 x_i)^2 \leqslant 4(\sum_{i=1}^4 x_i^2)$ show that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds a.s. that

$$\begin{aligned}
&\big\|D^{\theta,x,n}_{t,s} - I_d\big\|^4_F = \Big(\Big\|D^{\theta,x,n}_{t,s} - I_d\Big\|^2_F\Big)^2 \\
&\leqslant \bigg(3K(s-t) + K\int_t^s \Big(2\big\|D^{\theta,x,n}_{t,r} - I_d\big\|^2_F + 3\big\|D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}} - I_d\big\|^2_F\Big)\,\mathrm{d}r \\
&\quad + 2\int_t^s \sum_{i,j,k,l=1}^d \Big[(D^{\theta,x,n}_{t,r})_{ij} - \delta_{ij}\Big]\frac{\partial\sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})_{kj}\,\mathrm{d}W^{\theta,l}_r\bigg)^2 \\
&\leqslant 4\big(3K(s-t)\big)^2 + 4\Big(2K\int_t^s \big\|D^{\theta,x,n}_{t,r} - I_d\big\|^2_F\,\mathrm{d}r\Big)^2 \\
&\quad + 4\Big(3K\int_t^s \big\|D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}} - I_d\big\|^2_F\,\mathrm{d}r\Big)^2 \\
&\quad + 4\bigg(2\int_t^s \sum_{i,j,k,l=1}^d \Big[(D^{\theta,x,n}_{t,r})_{ij} - \delta_{ij}\Big]\frac{\partial\sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})_{kj}\,\mathrm{d}W^{\theta,l}_r\bigg)^2 \\
&\leqslant 36K^2(s-t)^2 + K^2(s-t)\int_t^s 16\big\|D^{\theta,x,n}_{t,r} - I_d\big\|^4_F + 36\big\|D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}} - I_d\big\|^4_F\,\mathrm{d}r \\
&\quad + 4\bigg(2\int_t^s \sum_{i,j,k,l=1}^d \Big[(D^{\theta,x,n}_{t,r})_{ij} - \delta_{ij}\Big]\frac{\partial\sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})_{kj}\,\mathrm{d}W^{\theta,l}_r\bigg)^2.
\end{aligned} \tag{3.85}$$

Next note that (3.7) and the fact that for all $a, b \in \mathbb{R}$ it holds that $(a+b)^2 \leqslant 2(a^2+b^2)$ and $2ab \leqslant a^2 + b^2$ demonstrate that for all $\theta \in \Theta$, $t \in [0,T]$, $r \in [t,T]$, $x \in \mathbb{R}^d$ it holds a.s.

that

$$
\begin{aligned}
&\bigg(\sum_{i,j,k,l=1}^{d}\Big[(D^{\theta,x,n}_{t,r})_{ij}-\delta_{ij}\Big]\frac{\partial\sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})_{kj}\bigg)^2\\
&\leqslant 2\bigg(\sum_{i,j,k,l=1}^{d}\Big[(D^{\theta,x,n}_{t,r})_{ij}-\delta_{ij}\Big]\frac{\partial\sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\Big((D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})_{kj}-\delta_{kj}\Big)\bigg)^2\\
&\quad+2\bigg(\sum_{i,j,l=1}^{d}\Big[(D^{\theta,x,n}_{t,r})_{ij}-\delta_{ij}\Big]\frac{\partial\sigma_{il}}{\partial x_j}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\bigg)^2\\
&\leqslant 2\bigg(\sum_{j=1}^{d}\Big[\max_{a\in\{1,2,\dots,d\}}(D^{\theta,x,n}_{t,r})_{aj}-\delta_{aj}\Big]\Big[\max_{a\in\{1,2,\dots,d\}}(D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})_{aj}-\delta_{aj}\Big]\\
&\quad\cdot\sum_{i,k,l=1}^{d}\frac{\partial\sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\bigg)^2+2\|D^{\theta,x,n}_{t,r}-I_d\|^2\bigg(\sum_{i,j,l=1}^{d}\frac{\partial\sigma_{il}}{\partial x_j}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})\bigg)^2\\
&\leqslant 2K^2\bigg(\sum_{j=1}^{d}\Big[\max_{a\in\{1,2,\dots,d\}}(D^{\theta,x,n}_{t,r})_{aj}-\delta_{aj}\Big]\Big[\max_{a\in\{1,2,\dots,d\}}(D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})_{aj}-\delta_{aj}\Big]\bigg)^2\\
&\quad+2K^2\|D^{\theta,x,n}_{t,r}-I_d\|^2\\
&\leqslant 2K^2\bigg(\sum_{j=1}^{d}\bigg(\Big[\max_{a\in\{1,2,\dots,d\}}\big((D^{\theta,x,n}_{t,r})_{aj}-\delta_{aj}\big)^2\Big]\\
&\quad+\Big[\max_{a\in\{1,2,\dots,d\}}\big((D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})_{aj}-\delta_{aj}\big)^2\Big]\bigg)\bigg)^2+2K^2\big(1+\|D^{\theta,x,n}_{t,r}-I_d\|^4\big)\\
&\leqslant 2K^2\Big(\|D^{\theta,x,n}_{t,r}-I_d\|^2+\|D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}-I_d\|^2\Big)^2+2K^2\big(1+\|D^{\theta,x,n}_{t,r}-I_d\|^4\big)\\
&\leqslant 4K^2\Big(\|D^{\theta,x,n}_{t,r}-I_d\|^4+\|D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}-I_d\|^4\Big)+2K^2\big(1+\|D^{\theta,x,n}_{t,r}-I_d\|^4\big)
\end{aligned}
\tag{3.86}
$$

The Burkholder-Davis-Gundy inequality therefore implies that for all $\theta\in\Theta$, $t\in[0,T]$, $s\in[t,T]$, $x\in\mathbb{R}^d$ it holds that

$$
\begin{aligned}
&\mathbb{E}\bigg[\bigg(2\int_t^s\sum_{i,j,k,l=1}^{d}\Big[(D^{\theta,x,n}_{t,r})_{ij}-\delta_{ij}\Big]\frac{\partial\sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})_{kj}\,\mathrm{d}W^{\theta,l}_r\bigg)^2\bigg]\\
&\leqslant 4\int_t^s\mathbb{E}\bigg[\Big|\sum_{i,j,k,l=1}^{d}\Big[(D^{\theta,x,n}_{t,r})_{ij}-\delta_{ij}\Big]\frac{\partial\sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}})_{kj}\Big|^2\bigg]\,\mathrm{d}r\\
&\leqslant 4K^2\int_t^s\Big(6\mathbb{E}\Big[\|D^{\theta,x,n}_{t,r}-I_d\|^4\Big]+4\mathbb{E}\Big[\|D^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}-I_d\|^4\Big]\Big)\,\mathrm{d}r+8K^2(s-t).
\end{aligned}
\tag{3.87}
$$

Combining this with Fubini's theorem and (3.85) shows that for all $\theta\in\Theta$, $t\in[0,T]$,

$s \in [t,T]$ $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\sup_{q\in[t,s]} \mathbb{E}\Big[\Big\|D^{\theta,x,n}_{t,\min\{q,\mathfrak{t}^{t,x}_k\}} - I_d\Big\|^4\Big] \\
&\leqslant 36K^2(s-t)^2 + \frac{K^2}{s-t}\int_t^{\max\{s,\mathfrak{t}^t_k\}} \Big(16\mathbb{E}\big[\big\|D^{\theta,x,n}_{t,\min\{r,\mathfrak{t}^{t,x}_k\}} - I_d\big\|_F^4\big] \\
&\quad + 36\mathbb{E}\big[\big\|D^{\theta,x,n}_{t,\max\{t,\min\{\lfloor r\rfloor_t,\mathfrak{t}^{t,x}_k\}\}} - I_d\big\|_F^4\big]\Big)\,\mathrm{d}r \\
&\quad + 32K^2(s-t) + 16K^2\int_t^{\max\{s,\mathfrak{t}^t_k\}} \Big(6\mathbb{E}\left[\|D^{\theta,x,n}_{t,\min\{r,\mathfrak{t}^{t,x}_k\}} - I_d\|^4\right] \\
&\quad + 4\mathbb{E}\left[\|D^{\theta,x,n}_{t,\max\{t,\min\{\lfloor r\rfloor_t,\mathfrak{t}^{t,x}_k\}\}} - I_d\|^4\right]\Big)\,\mathrm{d}r \\
&\leqslant 36K^2(s-t)^2 + 32K^2(s-t) \\
&\quad + (52(s-t)+160)K^2\int_t^s \mathbb{E}\Big[\sup_{\nu\in[t,r]} \big\|D^{\theta,x,n}_{t,\nu} - I_d\big\|_F^4\Big]\,\mathrm{d}r.
\end{aligned} \tag{3.88}$$

Fatou's lemma, Gronwall's inequality, and the fact that for all $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that $\lim_{k\to\infty} \mathfrak{t}^{t,x}_k = T$ therefore prove that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\sup_{r\in[t,s]} \mathbb{E}\Big[\big\|D^{\theta,x,n}_{t,r} - I_d\big\|^4\Big] = \sup_{r\in[t,s]} \mathbb{E}\Big[\lim_{k\to\infty} \big\|D^{\theta,x,n}_{t,\max\{t,\min\{r,\mathfrak{t}^{t,x}_k\}\}} - I_d\big\|^4\Big] \\
&\leqslant \sup_{k\in\mathbb{N}}\sup_{r\in[t,s]} \mathbb{E}\Big[\big\|D^{\theta,x,n}_{t,\max\{t,\min\{r,\mathfrak{t}^{t,x}_k\}\}} - I_d\big\|^4\Big] \\
&\leqslant 36K^2(s-t)(1+(s-t))\exp\Big(160K^2(s-t)(1+(s-t))\Big).
\end{aligned} \tag{3.89}$$

This establishes item (iv). The proof of Lemma 3.1.7 is thus complete. □

The findings in Lemma 3.1.7 lead to the following moment estimates for the stochastic processes displayed in (3.15).

Lemma 3.1.8. *Assume Setting 3.1.1 and let $n \in \mathbb{N}$. Then*

(i) for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\left(\mathbb{E}\left[\Big\|Z^{\theta,x,n}_{t,s}\Big\|^2\right]\right)^{\frac{1}{2}} \leqslant 1 + \frac{1}{\sqrt{\alpha(s-t)}}\Big(1 + \sqrt{3K(T-t)}\exp\left(\tfrac{5}{2}K(s-t)\right)\Big), \tag{3.90}$$

(ii) for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\Big\|Z^{\theta,x,n}_{t,s}\Big\|^4\right]\right)^{\frac{1}{4}} \\
&\leqslant 1 + \sqrt{\frac{6}{\alpha(s-t)}}\Big(1 + \sqrt{6K(1+T-t)}\exp\left(40K^2(1+T-t)^2\right)\Big),
\end{aligned} \tag{3.91}$$

and

(iii) for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
\mathbb{E}\left[\varphi(s, Y^{\theta,x,n}_{t,s})\Big\|Z^{\theta,x,n}_{t,s}\Big\|\right] &\leqslant \frac{\varphi(t,x)}{\sqrt{s-t}}\exp(\beta T)\Big(1 + \tfrac{b}{\beta[\inf_{r\in[0,T]}\inf_{y\in\mathbb{R}^d}\varphi(r,y)]}\Big) \\
&\quad \cdot \left[\sqrt{T-t} + \frac{1}{\sqrt{\alpha}}\Big(1 + \sqrt{3K(T-t)}\exp\left(\tfrac{5}{2}K(T-t)\right)\Big)\right].
\end{aligned} \tag{3.92}$$

Proof of Lemma 3.1.8. First observe that the assumption that for all $x, y \in \mathbb{R}^d$ it holds that $y^*\sigma(x)(\sigma(x))^*y \geqslant \alpha\|y\|^2$ ensures that for all $x, y \in \mathbb{R}^d$ it holds that

$$\begin{aligned} \|y\|^2 = y^*y = y^*(\sigma(x))^{-1}\sigma(x)(\sigma(x))^*((\sigma(x))^{-1})^*y \\ \leqslant \alpha\|((\sigma(x))^{-1})^*y\|^2 \leqslant \alpha\|(\sigma(x))^{-1}\|_F^2\|y\|^2. \end{aligned} \tag{3.93}$$

This implies that for all $x \in \mathbb{R}^d$ it holds that

$$\left\|(\sigma(x))^{-1}\right\|_F^2 \leqslant \frac{1}{\alpha}. \tag{3.94}$$

Item (iii) of Lemma 3.1.5, the triangle inequality, and the Burkholder-Davis-Gundy inequality therefore show that for all $\theta \in \Theta$, $t \in [0, T]$, $s \in [t, T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\left(\mathbb{E}\left[\left\|\left|Z_{t,s}^{\theta,x,n}\right\|\right|^2\right]\right)^{\frac{1}{2}} \\ &\leqslant 1 + \frac{1}{s-t}\left(\mathbb{E}\left[\left\|\int_t^s (\sigma(Y_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n}))^{-1}(D_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n} - I_d)\,\mathrm{d}W_r^\theta\right\|^2\right]\right)^{\frac{1}{2}} \\ &\quad + \frac{1}{s-t}\left(\mathbb{E}\left[\left\|\int_t^s (\sigma(Y_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n}))^{-1}\,\mathrm{d}W_r^\theta\right\|^2\right]\right)^{\frac{1}{2}} \\ &\leqslant 1 + \frac{1}{s-t}\left(\int_t^s \mathbb{E}\left[\|(\sigma(Y_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n}))^{-1}\|_F^2\|D_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n} - I_d\|_F^2\right]\mathrm{d}r\right)^{\frac{1}{2}} \\ &\quad + \frac{1}{s-t}\left(\int_t^s \mathbb{E}\left[\|(\sigma(Y_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n}))^{-1}\|_F^2\right]\mathrm{d}r\right)^{\frac{1}{2}} \\ &\leqslant 1 + \frac{1}{(s-t)\sqrt{\alpha}}\left(\int_t^s \mathbb{E}\left[\|D_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n} - I_d\|_F^2\right]\mathrm{d}r\right)^{\frac{1}{2}} + \frac{1}{(s-t)\sqrt{\alpha}}\left(\int_t^s \mathrm{d}r\right)^{\frac{1}{2}}. \end{aligned} \tag{3.95}$$

Combining this with item (iii) of Lemma 3.1.7 proves that for all $\theta \in \Theta$, $t \in [0, T]$, $s \in [t, T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\left(\mathbb{E}\left[\left\|\left|Z_{t,s}^{\theta,x,n}\right\|\right|^2\right]\right)^{\frac{1}{2}} \\ &\leqslant 1 + \frac{1}{(s-t)\sqrt{\alpha}}\left(\int_t^s 3K(s-t)\exp\left(5K(s-t)\right)\mathrm{d}r\right)^{\frac{1}{2}} + \frac{1}{\sqrt{\alpha(s-t)}} \\ &\leqslant 1 + \frac{1}{(s-t)\sqrt{\alpha}}\sqrt{3K(T-t)}\exp\left(\tfrac{5}{2}K(s-t)\right)\left(\int_t^s \mathrm{d}r\right)^{\frac{1}{2}} + \frac{1}{\sqrt{\alpha(s-t)}} \\ &\leqslant 1 + \frac{1}{\sqrt{\alpha(s-t)}}\left(1 + \sqrt{3K(T-t)}\exp\left(\tfrac{5}{2}K(s-t)\right)\right). \end{aligned} \tag{3.96}$$

This establishes item (i). Next note that the triangle inequality, the Burkholder-Davis-Gundy inequality, item (iii) of Lemma 3.1.5, and (3.94) ensure that for all $\theta \in \Theta$, $t \in [0, T]$,

$s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$
\begin{aligned}
&\left(\mathbb{E}\left[\left\|Z_{t,s}^{\theta,x,n}\right\|^4\right]\right)^{\frac{1}{4}} \\
&\leqslant 1 + \frac{1}{s-t}\left(\mathbb{E}\left[\left\|\int_t^s (\sigma(Y_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n}))^{-1}(D_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n} - I_d)\,\mathrm{d}W_r^\theta\right\|^4\right]\right)^{\frac{1}{4}} \\
&\quad + \frac{1}{s-t}\left(\mathbb{E}\left[\left\|\int_t^s (\sigma(Y_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n}))^{-1}\,\mathrm{d}W_r^\theta\right\|^4\right]\right)^{\frac{1}{4}} \\
&\leqslant 1 + \frac{\sqrt{6}}{s-t}\left(\int_t^s \left(\mathbb{E}\big[\|(\sigma(Y_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n}))^{-1}\|_F^4\|D_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n} - I_d\|_F^4\big]\right)^{\frac{1}{2}}\mathrm{d}r\right)^{\frac{1}{2}} \\
&\quad + \frac{\sqrt{6}}{s-t}\left(\int_t^s \left(\mathbb{E}\big[\|(\sigma(Y_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n}))^{-1}\|_F^4\big]\right)^{\frac{1}{2}}\mathrm{d}r\right)^{\frac{1}{2}} \\
&\leqslant 1 + \sqrt{\frac{6}{\alpha(s-t)^2}}\left(\int_t^s \left(\mathbb{E}\big[\|D_{t,\max\{t,\lfloor r\rfloor_t\}}^{\theta,x,n} - I_d\|_F^4\big]\right)^{\frac{1}{2}}\mathrm{d}r\right)^{\frac{1}{2}} \\
&\quad + \sqrt{\frac{6}{\alpha(s-t)^2}}\left(\int_t^s \mathrm{d}r\right)^{\frac{1}{2}}.
\end{aligned}
\tag{3.97}
$$

This and item (iv) of Lemma 3.1.7 prove that for all $\theta \in \Theta$, $t \in [0,T)$, $s \in (t,T]$, $x \in \mathbb{R}^d$ it holds that

$$
\begin{aligned}
&\left(\mathbb{E}\left[\left\|Z_{t,s}^{\theta,x,n}\right\|^4\right]\right)^{\frac{1}{4}} \\
&\leqslant 1 + \sqrt{\frac{6}{\alpha(s-t)}} \\
&\quad + \sqrt{\frac{6}{\alpha(s-t)^2}}\left(\int_t^s 6K(1+(s-t))\exp\left(80K^2(1+(s-t))^2\right)\mathrm{d}r\right)^{\frac{1}{2}} \\
&\leqslant 1 + \sqrt{\frac{6}{\alpha(s-t)}} + 6\sqrt{\frac{K(1+T-t)}{\alpha(s-t)^2}}\exp\left(40K^2(1+s-t)^2\right)\left(\int_t^s \mathrm{d}r\right)^{\frac{1}{2}} \\
&\leqslant 1 + \sqrt{\frac{6}{\alpha(s-t)}}\left(1 + \sqrt{6K(1+T-t)}\exp\left(40K^2(1+T-t)^2\right)\right).
\end{aligned}
\tag{3.98}
$$

This establishes item (ii). Next observe that (3.16) ensures that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds a.s. that

$$
\begin{aligned}
&(\tfrac{\partial\varphi}{\partial t})(s, Y_{t,s}^{\theta,x,n}) + \langle(\nabla_x\varphi)(s, Y_{t,s}^{\theta,x,n}), \mu(Y_{t,s}^{\theta,x,n})\rangle \\
&\quad + \tfrac{1}{2}\operatorname{Tr}(\sigma(Y_{t,s}^{\theta,x,n})[\sigma(Y_{t,s}^{\theta,x,n})]^*(\operatorname{Hess}_x\varphi)(s, Y_{t,s}^{\theta,x,n})) \\
&\quad + \tfrac{1}{2}\tfrac{\|[(\nabla_x\varphi)(s,Y_{t,s}^{\theta,x,n})]^*\sigma(Y_{t,s}^{\theta,x,n})\|^2}{\varphi(s,Y_{t,s}^{\theta,x,n})} \leqslant \beta\varphi(s, Y_{t,s}^{\theta,x,n}) + b.
\end{aligned}
\tag{3.99}
$$

Furthermore, observe that the assumption that W has continuous sample paths implies that Y has continuous sample paths. Combining this, (3.99), and [50, Theorem 2.4 (i)] (applied for every $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ with $H \curvearrowleft \mathbb{R}^d$, $U \curvearrowleft \mathbb{R}^d$, $T \curvearrowleft T-t$, $O \curvearrowleft \mathbb{R}^d$, $a \curvearrowleft ([0,T-t] \times \Omega \ni (s,\omega) \mapsto \mu(Y_{t,\max\{t,\lfloor t+s\rfloor\}}^{\theta,x,n}(\omega)) \in \mathbb{R}^d)$, $b \curvearrowleft ([0,T-t] \times \Omega \ni (s,\omega) \mapsto \sigma(Y_{t,\max\{t,\lfloor t+s\rfloor\}}^{\theta,x,n}(\omega)) \in \mathbb{R}^{d\times d})$, $X \curvearrowleft ([0,T-t] \times \Omega \ni (s,\omega) \mapsto Y_{t,t+s}^{\theta,x,n}(\omega) \in \mathbb{R}^d)$, $p \curvearrowleft 2$,

$\alpha \curvearrowleft ([0,T-t]\times\Omega \ni (s,\omega) \mapsto \beta \in \mathbb{R})$, $\beta \curvearrowleft ([0,T-t]\times\Omega \ni (s,\omega) \mapsto b \in \mathbb{R})$, $q_1 \curvearrowleft 2$, $q_2 \curvearrowleft \infty$ in the notation of [50, Theorem 2.4]) demonstrates that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}\Big(\mathbb{E}\big[|\varphi(s,Y^{\theta,x,n}_{t,s})|^2\big]\Big)^{\frac{1}{2}} &\leqslant \exp(\beta T)\Big(\varphi(t,x) + b\int_0^T \exp(-\beta s)\,\mathrm{d}s\Big)\\ &\leqslant \exp(\beta T)\Big(\varphi(t,x)+\tfrac{b}{\beta}\big(1-\exp(-\beta T)\big)\Big) \leqslant \exp(\beta T)\Big(\varphi(t,x)+\tfrac{b}{\beta}\Big).\end{aligned} \tag{3.100}$$

Combining this with Hölder's inequality, item (i), and the assumption that $\inf_{r\in[0,T]}\inf_{y\in\mathbb{R}^d}\varphi(r,y) > 0$ proves that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}&\mathbb{E}\left[\varphi(s,Y^{\theta,x,n}_{t,s})\Big\|\Big|Z^{\theta,x,n}_{t,s}\Big\|\Big|\right] \leqslant \Big(\mathbb{E}\big[\|\varphi(s,Y^{\theta,x,n}_{t,s})\|^2\big]\Big)^{\frac{1}{2}}\left(\mathbb{E}\left[\Big\|\Big|Z^{\theta,x,n}_{t,s}\Big\|\Big|^2\right]\right)^{\frac{1}{2}}\\ &\leqslant \exp(\beta T)\Big(\varphi(t,x)+\tfrac{b}{\beta}\Big)\\ &\quad\cdot\left[\frac{\sqrt{T-t}}{\sqrt{s-t}} + \frac{1}{\sqrt{\alpha(s-t)}}\Big(1+\sqrt{3K(T-t)}\exp\big(\tfrac{5}{2}K(s-t)\big)\Big)\right]\\ &\leqslant \frac{\varphi(t,x)}{\sqrt{s-t}}\exp(\beta T)\Big(1+\tfrac{b}{\beta[\inf_{r\in[0,T]}\inf_{y\in\mathbb{R}^d}\varphi(r,y)]}\Big)\\ &\quad\cdot\left[\sqrt{T-t}+\frac{1}{\sqrt{\alpha}}\Big(1+\sqrt{3K(T-t)}\exp\big(\tfrac{5}{2}K(T-t)\big)\Big)\right].\end{aligned} \tag{3.101}$$

This establishes item (iii). The proof of Lemma 3.1.8 is thus complete. □

3.1.5 Integrability properties for MLP approximations

The following lemma establishes integrability of the MLP approximation scheme defined in (3.17). The proof of Lemma 3.1.9 uses the Lipschitz assumptions of the coefficients and the results in Lemma 3.1.3 and Lemma 3.1.8.

Lemma 3.1.9. *Assume Setting 3.1.1, let $n, M, Q \in \mathbb{N}$, and assume*

$$\sup_{t\in[0,T)}\sup_{x\in\mathbb{R}^d}\left[\frac{|g(x)|}{\varphi(T,x)} + \frac{|(F(0))(t,x)|}{\varphi(t,x)}\sqrt{T-t}\right] < \infty. \tag{3.102}$$

Then

(i) it holds for all $m \in \mathbb{N}_0$, $\theta \in \Theta$ that

$$\sup_{t\in[0,T)}\sup_{x\in\mathbb{R}^d}\mathbb{E}\left[\frac{\big\|\big|V^\theta_{n,m,M,Q}(t,x)\big\|\big|}{\varphi(t,x)}\right] < \infty, \tag{3.103}$$

(ii) it holds for all $m \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T)$, $x \in \mathbb{R}^d$ that

$$\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\mathbb{E}\left[\Big\|\Big|(F(V^\theta_{n,m,M,Q}))(s,Y^{\theta,x,n}_{t,s})Z^{\theta,x,n}_{t,s}\Big\|\Big|\right] < \infty, \tag{3.104}$$

and

(iii) for all $m \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}&\mathbb{E}[V^\theta_{n,m,M,Q}(t,x)] = \mathbb{E}[g(Y^{0,x,n}_{t,T})Z^{0,x,n}_{t,T}]\\ &+\mathbb{E}\left[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)(F(V^\theta_{n,m-1,M,Q}))(s,Y^{0,x,n}_{t,s})Z^{0,x,n}_{t,s}\right].\end{aligned} \tag{3.105}$$

Proof of Lemma 3.1.9. First observe that (3.9) that for all $t \in [0,T]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \dots, y_{d+1}), z = (z_1, z_2, \dots, z_{d+1}) \in \mathbb{R}^{d+1}$ it holds that

$$|f(t,x,y) - f(t,x,z)| \leqslant \sum_{\nu=1}^{d+1} L_\nu |y_\nu - z_\nu| \leqslant \|y-z\| \sum_{\nu=1}^{d+1} L_\nu \leqslant \mathcal{L}\|y-z\|. \tag{3.106}$$

We are going to prove (3.103) by induction on $m \in \mathbb{N}_0$. Note that the fact that for all $\theta \in \Theta$, $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that $V^\theta_{n,0,M,Q}(t,x) = 0$ establishes (3.103) in the base case $m = 0$. For the induction step assume that (3.103) holds for all $\mathfrak{m} \in \{0,1,\dots,m\}$. Observe that the Cauchy-Schwarz inequality, item (i) of Lemma 3.1.7 (applied with $p \curvearrowleft 2$ in the notation of Lemma 3.1.7), item (i) of Lemma 3.1.8, and (3.10) show that for all $\theta \in \Theta$, $i \in \mathbb{R}$, $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\mathbb{E}\left[\left\|\left|(g(Y^{(\theta,0,-i),x,n}_{t,T}) - g(x))Z^{(\theta,0,-i),x,n}_{t,T}\right\|\right|\right] \\
&\leqslant \mathcal{L}\mathbb{E}\left[\|Y^{(\theta,0,-i),x,n}_{t,T} - x\|\left\|\left|Z^{(\theta,0,-i),x,n}_{t,T}\right\|\right|\right] \\
&\leqslant \mathcal{L}\Big(\mathbb{E}\big[\|Y^{(\theta,0,-i),x,n}_{t,T} - x\|^2\big]\Big)^{\frac{1}{2}}\left(\mathbb{E}\left[\left\|\left|Z^{(\theta,0,-i),x,n}_{t,T}\right\|\right|^2\right]\right)^{\frac{1}{2}} \\
&\leqslant \mathcal{L}\sqrt{\gamma^{(2)}_x(T-t)}\exp\left(\left(\tfrac{1}{2}+2K^2\right)(T-t)\right) \\
&\qquad\cdot\left[1 + \frac{1}{\sqrt{\alpha(T-t)}}\left(1 + \sqrt{3K(T-t)}\exp\left(3K^2(T-t)\right)\right)\right] \\
&\leqslant \mathcal{L}\sqrt{\gamma^{(2)}_x}\exp\left(\left(\tfrac{1}{2}+2K^2\right)(T-t)\right) \\
&\qquad\cdot\left[\sqrt{T-t} + \frac{1}{\sqrt{\alpha}}\left(1 + \sqrt{3K(T-t)}\exp\left(3K^2(T-t)\right)\right)\right].
\end{aligned} \tag{3.107}$$

Next note that item (iii) of Lemma 3.1.8 ensures that there exists $c_Z \in [0,\infty)$ which satisfies for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\mathbb{E}\left[\varphi(s, Y^{\theta,x,n}_{t,s})\left\|\left|Z^{\theta,x,n}_{t,s}\right\|\right|\right] \leqslant \frac{c_Z}{\sqrt{s-t}}\varphi(t,x). \tag{3.108}$$

In addition, observe that (3.102) implies that there exists $a \in [0,\infty)$ which satisfies for all $t \in [0,T)$, $x \in \mathbb{R}^d$ that

$$|g(x)| \leqslant a\varphi(T,x) \qquad \text{and} \qquad |(F(0))(t,x)|\sqrt{T-t} \leqslant a\varphi(t,x). \tag{3.109}$$

Furthermore, observe that triangle inequality and (3.106) demonstrate that for all $\theta \in \Theta$, $i \in \mathbb{R}$, $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\sum_{l=0}^{m}\sum_{s\in(t,T)} q^{Q,[t,T]}(s) \\
&\quad\cdot\mathbb{E}\left[\left|((F(V^{(\theta,l,1)}_{n,l,M,Q})) - \mathbb{1}_{\mathbb{N}}(l)(F(V^{(\theta,-l,1)}_{n,l-1,M,Q})))(s, Y^{(\theta,l,1),x,n}_{t,s})\right|\left\|\left|Z^{(\theta,l,1),x,n}_{t,s}\right\|\right|\right] \\
&\leqslant \sum_{s\in(t,T)} q^{Q,[t,T]}(s)\mathbb{E}\left[\left|(F(0))(s, Y^{(\theta,l,1),x,n}_{t,s})\right|\left\|\left|Z^{(\theta,l,1),x,n}_{t,s}\right\|\right|\right] \\
&\quad+\sum_{l=1}^{m}\sum_{s\in(t,T)} \mathcal{L}q^{Q,[t,T]}(s) \\
&\quad\mathbb{E}\left[\left\|\left|(V^{(\theta,l,1)}_{n,l,M,Q} - V^{(\theta,-l,1)}_{n,l-1,M,Q})(s, Y^{(\theta,l,1),x,n}_{t,s})\right\|\right|\left\|\left|Z^{(\theta,l,1),x,n}_{t,s}\right\|\right|\right].
\end{aligned} \tag{3.110}$$

Next observe that (3.108), (3.109), and item (ii) of Lemma 3.1.3 show that for all $\theta \in \Theta$, $i \in \mathbb{R}$, $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\mathbb{E}\Big[|(F(0))(s,Y^{(\theta,0,1),x,n}_{t,s})|\Big\|Z^{(\theta,0,1),x,n}_{t,s}\Big\|\Big] \\ &\leqslant \sum_{s\in(t,T)} aq^{Q,[t,T]}(s)\mathbb{E}\bigg[\frac{\varphi(s,Y^{(\theta,0,1),x,n}_{t,s})}{\sqrt{T-s}}\Big\|Z^{(\theta,0,1),x,n}_{t,s}\Big\|\bigg] \\ &\leqslant ac_z\varphi(t,x)\sum_{s\in(t,T)}\frac{q^{Q,[t,T]}(s)}{\sqrt{T-s}\sqrt{s-t}} \leqslant ac_z\pi\varphi(t,x). \end{aligned} \tag{3.111}$$

Moreover, observe that (3.108) and, e.g., [59, Lemma 2.2] prove that for all $l \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\mathbb{E}\Big[\Big\|(V^{(\theta,l,1)}_{n,l,M,Q}-V^{(\theta,-l,1)}_{n,l-1,M,Q})(s,Y^{(\theta,l,1),x,n}_{t,s})\Big\|\Big\|Z^{(\theta,l,1),x,n}_{t,s}\Big\|\Big] \\ &=\mathbb{E}\Bigg[\mathbb{E}\Bigg[\frac{\Big\|(V^{(\theta,l,1)}_{n,l,M,Q}-V^{(\theta,-l,1)}_{n,l-1,M,Q})(s,y)\Big\|}{\varphi(s,y)}\Bigg]\Bigg|_{y=Y^{(\theta,l,1),x,n}_{t,s}}\Big\|Z^{(\theta,l,1),x,n}_{t,s}\Big\|\varphi(s,Y^{(\theta,l,1),x,n}_{t,s})\Bigg] \\ &\leqslant\Bigg[\sup_{r\in[0,T)}\sup_{y\in\mathbb{R}^d}\frac{\mathbb{E}\Big[\Big\|(V^{(\theta,l,1)}_{n,l,M,Q}-V^{(\theta,-l,1)}_{n,l-1,M,Q})(r,y)\Big\|\Big]}{\varphi(r,y)}\Bigg] \\ &\quad\cdot\mathbb{E}\Big[\Big\|Z^{(\theta,l,1),x,n}_{t,s}\Big\|\varphi(s,Y^{(\theta,l,1),x,n}_{t,s})\Big] \\ &\leqslant\Bigg[\sup_{r\in[0,T)}\sup_{y\in\mathbb{R}^d}\frac{\mathbb{E}\Big[\Big\|(V^{(\theta,l,1)}_{n,l,M,Q}-V^{(\theta,-l,1)}_{n,l-1,M,Q})(r,y)\Big\|\Big]}{\varphi(r,y)}\Bigg]\frac{c_Z\varphi(t,x)}{\sqrt{s-t}}. \end{aligned} \tag{3.112}$$

Item (i) of Lemma 3.1.3 hence demonstrates that for all $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\sum_{l=1}^{m}\sum_{s\in(t,T)}\mathcal{L}q^{Q,[t,T]}(s)\mathbb{E}\Big[\Big\|(V^{(\theta,l,1)}_{n,l,M,Q}-V^{(\theta,-l,1)}_{n,l-1,M,Q})(s,Y^{(\theta,l,1),x,n}_{t,s})\Big\|\Big\|Z^{(\theta,l,1),x,n}_{t,s}\Big\|\Big] \\ &\leqslant c_Z\mathcal{L}\varphi(t,x)\sum_{l=1}^{m}\Bigg[\sup_{r\in[0,T)}\sup_{y\in\mathbb{R}^d}\frac{\mathbb{E}\Big[\Big\|(V^{(\theta,l,1)}_{n,l,M,Q}-V^{(\theta,-l,1)}_{n,l-1,M,Q})(r,y)\Big\|\Big]}{\varphi(r,y)}\Bigg] \\ &\qquad\cdot\sum_{s\in(t,T)}\frac{q^{Q,[t,T]}(s)}{\sqrt{s-t}} \\ &\leqslant 2c_Z\mathcal{L}\sqrt{T-t}\varphi(t,x) \\ &\qquad\cdot\sum_{l=1}^{m}\Bigg[\sup_{r\in[0,T)}\sup_{y\in\mathbb{R}^d}\frac{\mathbb{E}\Big[\Big\|(V^{(\theta,l,1)}_{n,l,M,Q}-V^{(\theta,-l,1)}_{n,l-1,M,Q})(r,y)\Big\|\Big]}{\varphi(r,y)}\Bigg]. \end{aligned} \tag{3.113}$$

This, (3.107), (3.109), (3.110), and (3.111) show that for all $\theta \in \Theta$ it holds that

$$
\begin{aligned}
&\sup_{t\in[0,T)}\sup_{x\in\mathbb{R}^d}\mathbb{E}\Bigg[\frac{\big\|\big|V^{\theta}_{n,m+1,M,Q}(t,x)\big\|\big|}{\varphi(t,x)}\Bigg]\\
&\leqslant \sup_{x\in\mathbb{R}^d}\frac{|g(x)|}{\varphi(T,x)}\\
&\quad+\sup_{t\in[0,T)}\sup_{x\in\mathbb{R}^d}\frac{1}{M^{m+1}}\sum_{i=1}^{M^{m+1}}\mathbb{E}\Bigg[\frac{\big\|\big|(g(Y^{(\theta,0,-i),x,n}_{t,T})-g(x))Z^{(\theta,0,-i),x,n}_{t,T}\big\|\big|}{\varphi(t,x)}\Bigg]\\
&\quad+\sup_{t\in[0,T)}\sup_{x\in\mathbb{R}^d}\sum_{l=0}^{m}\sum_{s\in(t,T)}\frac{q^{Q,[t,T]}}{M^{m+1-l}}\\
&\quad\cdot\sum_{i=1}^{M^{m+1-l}}\mathbb{E}\Bigg[\frac{\big\|\big|((F(V^{(\theta,l,i)}_{n,l,M,Q}))-\mathbb{1}_{\mathbb{N}}(l)(F(V^{(\theta,-l,i)}_{n,l-1,M,Q})))(s,Y^{(\theta,l,i),x,n}_{t,s})Z^{(\theta,l,i),x,n}_{t,s}\big\|\big|}{\varphi(t,x)}\Bigg]\\
&=\sup_{x\in\mathbb{R}^d}\frac{|g(x)|}{\varphi(T,x)}+\sup_{t\in[0,T)}\sup_{x\in\mathbb{R}^d}\mathbb{E}\Bigg[\frac{\big\|\big|(g(Y^{(\theta,0,1),x,n}_{t,T})-g(x))Z^{(\theta,0,1),x,n}_{t,T}\big\|\big|}{\varphi(t,x)}\Bigg]\\
&\quad+\sup_{t\in[0,T)}\sup_{x\in\mathbb{R}^d}\sum_{l=0}^{m}\sum_{s\in(t,T)}q^{Q,[t,T]}\\
&\quad\cdot\mathbb{E}\Bigg[\frac{\big\|\big|(F(V^{(\theta,l,1)}_{n,l,M,Q})-\mathbb{1}_{\mathbb{N}}(l)F(V^{(\theta,-l,1)}_{n,l-1,M,Q}))(s,Y^{(\theta,l,1),x,n}_{t,s})Z^{(\theta,l,1),x,n}_{t,s}\big\|\big|}{\varphi(t,x)}\Bigg]\\
&\leqslant a+ac_Z\pi+\frac{\mathcal{L}\sqrt{\gamma^{(2)}_x}\exp((\frac{1}{2}+2K^2)T)}{\inf_{r\in[0,T]}\inf_{y\in\mathbb{R}^d}\varphi(r,y)}\\
&\quad\cdot\left[\sqrt{T}+\frac{1}{\sqrt{\alpha}}\Big(1+\sqrt{3KT}\exp\big(3K^2T\big)\Big)\right]\\
&\quad+2c_Z\mathcal{L}\sqrt{T}\sum_{l=1}^{m}\left[\sup_{r\in[0,T)}\sup_{y\in\mathbb{R}^d}\frac{\mathbb{E}\Big[\big\|\big|(V^{(\theta,l,1)}_{n,l,M,Q}-V^{(\theta,-l,1)}_{n,l-1,M,Q})(r,y)\big\|\big|\Big]}{\varphi(r,y)}\right].
\end{aligned}
\tag{3.114}
$$

Induction therefore proves (3.103). This establishes item (i). Next note that the triangle inequality, item (v) of Lemma 3.1.4, (3.106), and (3.102) imply that for all $m \in \mathbb{N}$, $\theta \in \Theta$,

$t \in [0, T)$, $s \in [t, T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\mathbb{E}\left[\left|\!\left|\!\left|(F(V^\theta_{n,m,M,Q}))(s, Y^{\theta,x,n}_{t,s})Z^{\theta,x,n}_{t,s}\right|\!\right|\!\right|\right] \\
&\leqslant \mathbb{E}\left[\left|\!\left|\!\left|(F(0))(s, Y^{\theta,x,n}_{t,s})Z^{\theta,x,n}_{t,s}\right|\!\right|\!\right|\right] + \mathcal{L}\mathbb{E}\left[\left|\!\left|\!\left|V^\theta_{n,m,M,Q}(s, Y^{\theta,x,n}_{t,s})Z^{\theta,x,n}_{t,s}\right|\!\right|\!\right|\right] \\
&= \mathbb{E}\left[\frac{|(F(0))(s, Y^{\theta,x,n}_{t,s})|}{\varphi(s, Y^{\theta,x,n}_{t,s})}\sqrt{T-s}\,\frac{\varphi(s, Y^{\theta,x,n}_{t,s})}{\sqrt{T-s}}\left|\!\left|\!\left|Z^{\theta,x,n}_{t,s}\right|\!\right|\!\right|\right] \\
&\quad + \mathcal{L}\mathbb{E}\left[\frac{\left|\!\left|\!\left|V^\theta_{n,m,M,Q}(s, Y^{\theta,x,n}_{t,s})\right|\!\right|\!\right|}{\varphi(s, Y^{\theta,x,n}_{t,s})}\varphi(s, Y^{\theta,x,n}_{t,s})\left|\!\left|\!\left|Z^{\theta,x,n}_{t,s}\right|\!\right|\!\right|\right] \\
&\leqslant \left[\sup_{r\in[t,s]}\sup_{y\in\mathbb{R}^d}\left(\frac{|(F(0))(r,y)|}{\varphi(r,y)}\sqrt{T-r} + \mathcal{L}\sqrt{T}\frac{\mathbb{E}\left[|\!|\!|V^\theta_{n,m,M,Q}(r,y)|\!|\!|\right]}{\varphi(r,y)}\right)\right] \\
&\quad \cdot \frac{1}{\sqrt{T-s}}\mathbb{E}\left[\varphi(s, Y^{\theta,x,n}_{t,s})\left|\!\left|\!\left|Z^{\theta,x,n}_{t,s}\right|\!\right|\!\right|\right] \\
&\leqslant \left[\sup_{r\in[t,s]}\sup_{y\in\mathbb{R}^d}\left(\frac{|(F(0))(r,y)|}{\varphi(r,y)}\sqrt{T-r} + \mathcal{L}\sqrt{T}\frac{\mathbb{E}\left[|\!|\!|V^\theta_{n,m,M,Q}(r,y)|\!|\!|\right]}{\varphi(r,y)}\right)\right] \\
&\quad \cdot \frac{c_Z\varphi(t,x)}{\sqrt{T-s}\sqrt{s-t}}.
\end{aligned} \tag{3.115}$$

Item (ii) of Lemma 3.1.3 hence ensures that for all $m \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0, T)$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\mathbb{E}\left[\left|\!\left|\!\left|(F(V^\theta_{n,m,M,Q}))(s, Y^{\theta,x,n}_{t,s})Z^{\theta,x,n}_{t,s}\right|\!\right|\!\right|\right] \\
&\leqslant \left[\sup_{r\in[t,s]}\sup_{y\in\mathbb{R}^d}\frac{|(F(0))(r,y)|}{\varphi(r,y)}\sqrt{T-r} + \mathcal{L}\sqrt{T}\frac{\mathbb{E}\left[|\!|\!|V^\theta_{n,m,M,Q}(r,y)|\!|\!|\right]}{\varphi(r,y)}\right] \\
&\quad \cdot \sum_{s\in(t,T)} q^{Q,[t,T]}(s)\frac{c_Z\varphi(t,x)}{\sqrt{T-s}\sqrt{s-t}} \\
&\leqslant \left[\sup_{r\in[t,s]}\sup_{y\in\mathbb{R}^d}\frac{|(F(0))(r,y)|}{\varphi(r,y)}\sqrt{T-r} + \mathcal{L}\sqrt{T}\frac{\mathbb{E}\left[|\!|\!|V^\theta_{n,m,M,Q}(r,y)|\!|\!|\right]}{\varphi(r,y)}\right]c_z\pi\varphi(t,x).
\end{aligned} \tag{3.116}$$

Combining this with item (i) and (3.102) establishes item (ii). In the next step observe that Fubini's theorem, item (ii), item (vi) of Lemma 3.1.4, (3.17), and a telescoping sum argument demonstrate that for all $m \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0, T)$, $x \in \mathbb{R}^d$ it holds that

$$
\begin{aligned}
&\mathbb{E}[V^{\theta}_{n,m,M,Q}(t,x)] \\
&= g(x)\mathbf{e}_1 + \frac{1}{M^m}\sum_{i=1}^{M^m}\mathbb{E}\big[(g(Y^{(\theta,0,-i),x,n}_{t,T}) - g(x))Z^{(\theta,0,-i),x,n}_{t,T}\big] \\
&\quad + \sum_{l=0}^{m-1}\sum_{s\in(t,T)}\frac{q^{Q,[t,T]}(s)}{M^{m-l}} \\
&\quad \cdot \mathbb{E}\Bigg[\sum_{i=1}^{M^{m-l}}(F(V^{(\theta,l,i)}_{n,l,M,Q}) - \mathbb{1}_{\mathbb{N}}(l)F(V^{(\theta,-l,i)}_{n,l-1,M,Q}))(s, Y^{(\theta,l,i),x,n}_{t,s})Z^{(\theta,l,i),x,n}_{t,s}\Bigg] \\
&= \mathbb{E}\big[g(Y^{0,x,n}_{t,T})Z^{0,x,n}_{t,T}\big] \\
&\quad + \sum_{l=0}^{m-1}\sum_{s\in(t,T)}\frac{q^{Q,[t,T]}(s)}{M^{m-l}}\sum_{i=1}^{M^{m-l}}\mathbb{E}\big[(F(V^{(\theta,l,i)}_{n,l,M,Q}))(s, Y^{(\theta,l,i),x,n}_{t,s})Z^{(\theta,l,i),x,n}_{t,s}\big] \\
&\quad - \sum_{l=1}^{m-1}\sum_{i=1}^{M^{m-l}}\sum_{s\in(t,T)}\frac{q^{Q,[t,T]}(s)}{M^{m-l}}\mathbb{E}\big[(F(V^{(\theta,-l,i)}_{n,l-1,M,Q}))(s, Y^{(\theta,l,i),x,n}_{t,s})Z^{(\theta,l,i),x,n}_{t,s}\big] \\
&= \mathbb{E}\big[g(Y^{0,x,n}_{t,T})Z^{0,x,n}_{t,T}\big] + \sum_{l=0}^{m-1}\sum_{s\in(t,T)}q^{Q,[t,T]}(s)\mathbb{E}\big[(F(V^{\theta}_{n,l,M,Q}))(s, Y^{0,x,n}_{t,s})Z^{0,x,n}_{t,s}\big] \\
&\quad - \sum_{l=1}^{m-1}\sum_{s\in(t,T)}q^{Q,[t,T]}(s)\mathbb{E}\big[(F(V^{\theta}_{n,l-1,M,Q}))(s, Y^{0,x,n}_{t,s})Z^{0,x,n}_{t,s}\big] \\
&= \mathbb{E}\big[g(Y^{0,x,n}_{t,T})Z^{0,x,n}_{t,T}\big] + \sum_{l=0}^{m-1}\sum_{s\in(t,T)}q^{Q,[t,T]}(s)\Big(\mathbb{E}\big[(F(V^{\theta}_{n,l,M,Q}))(s, Y^{0,x,n}_{t,s})Z^{0,x,n}_{t,s}\big] \\
&\quad - \mathbb{1}_{\mathbb{N}}(l)\mathbb{E}\big[(F(V^{\theta}_{n,l-1,M,Q}))(s, Y^{0,x,n}_{t,s})Z^{0,x,n}_{t,s}\big]\Big) \\
&= \mathbb{E}\big[g(Y^{0,x,n}_{t,T})Z^{0,x,n}_{t,T}\big] + \sum_{s\in(t,T)}q^{Q,[t,T]}(s)\mathbb{E}\big[(F(V^{\theta}_{n,m-1,M,Q}))(s, Y^{0,x,n}_{t,s})Z^{0,x,n}_{t,s}\big].
\end{aligned}
\tag{3.117}
$$

This establishes item (iii). The proof of Lemma 3.1.9 is thus complete. □

3.2 Computational complexity analysis for MLP approximations

This section's aim is to estimate the overall complexity of our MLP approximation scheme which is achieved in this chapter's main result, Theorem 3.2.4 in Subsection 3.2.2 below. Essential for the proof of Theorem 3.2.4 are the nonlinear Feynman-Kac type formula in Lemma 3.2.1 in Subsection 3.2.1 and a helpful upper bound on error terms arising from the nonlinear Feynman-Kac type formula which is given in Lemma 3.2.2. Combining Theorem 3.2.4 with an explicit calculation of the error arising from the GauSS-Legendre quadrature (cf. Lemma 3.2.5) and the assumption of a smooth PDE solution leads to the more summarized global error bound in Corollary 3.2.6. Corollary 3.2.8 holds our main result of this chapter. It demonstrates that - under the assumption of a smooth solution - the computational effort of the approximation scheme in (3.17) is of order $O(d\varepsilon^{-(6+\delta)})$ for all $\delta \in (0, \infty)$ if d denotes the dimension of the problem and ε is the prescribed accuracy. In this section we follow closely the proof strategies developed in [54, Section 4] but for a larger class of semilinear PDEs under additional assumptions.

3.2.1 Nonlinear Feynman-Kac formula for the approximation scheme

The following lemma introduces a nonlinear Feynman-Kac formula for the approximation scheme displayed in (3.17). An essential part of the proof is the application of the Bismut-Elworthy-Li formula developed in Theorem 2.2.3. Lemma 3.2.1 is a generalization of the Bismut-Elworthy-Li type formula in [55, Lemma 4.2].

Lemma 3.2.1. *Assume Setting 3.1.1, let $n \in \mathbb{N}$, and assume that*

$$\sup_{t\in[0,T]}\sup_{x\in\mathbb{R}^d}\frac{|||v^\infty(t,x)|||}{\varphi(t,x)}+\sup_{t\in[0,T]}\sup_{x\in\mathbb{R}^d}\frac{|(F(0))(t,x)|}{\varphi(t,x)}<\infty. \tag{3.118}$$

Then

(i) *for all $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that*

$$u^\infty(t,x)=\mathbb{E}[g(Y^{0,x,n}_{t,T})]+\mathbb{E}\bigg[\int_t^T(F(v^\infty))(r,Y^{0,x,n}_{t,r})\,\mathrm{d}r\bigg] \tag{3.119}$$

and

(ii) *for all $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that*

$$\begin{aligned}(\nabla_x u^\infty)(t,x)&=\mathbb{E}\bigg[g(Y^{\theta,x,n}_{t,T})\tfrac{1}{T-t}\int_t^T(\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1}\frac{\partial}{\partial x}Y^{\theta,x,n}_{t,s}\,\mathrm{d}W^\theta_s\bigg]\\&\quad+\mathbb{E}\bigg[\int_t^T(F(v^\infty))(r,Y^{0,x,n}_{t,r})\tfrac{1}{r-t}\int_t^r(\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1}\frac{\partial}{\partial x}Y^{\theta,x,n}_{t,s}\,\mathrm{d}W^\theta_s\,\mathrm{d}r\bigg].\end{aligned} \tag{3.120}$$

Proof of Lemma 3.2.1. First observe that the triangle inequality, (3.9), and (3.118) guarantee that

$$\begin{aligned}&\sup_{t\in[0,T)}\sup_{x\in\mathbb{R}^d}\frac{|(F(v^\infty))(t,x)|}{\varphi(t,x)}\\&\leqslant\sup_{t\in[0,T)}\sup_{x\in\mathbb{R}^d}\frac{\mathcal{L}|||v^\infty(t,x)|||}{\varphi(t,x)}+\sup_{t\in[0,T)}\sup_{x\in\mathbb{R}^d}\frac{|(F(0))(t,x)|}{\varphi(t,x)}<\infty.\end{aligned} \tag{3.121}$$

Furthermore, note that Itô's formula and (3.11) show that for all $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds a.s. that

$$\begin{aligned}&g(Y^{0,x,n}_{t,T})-u^\infty(t,x)=u^\infty(T,Y^{0,x,n}_{t,T})-u^\infty(t,Y^{0,x,n}_{t,t})\\&=\int_t^T\Big[(\tfrac{\partial u^\infty}{\partial t})(r,Y^{0,x,n}_{t,r})+\tfrac12\operatorname{Tr}(\sigma(r,Y^{0,x,n}_{t,r})[\sigma(r,Y^{0,x,n}_{t,r})]^*(\operatorname{Hess}_x u^\infty)(r,Y^{0,x,n}_{t,r}))\\&\qquad+\langle\mu(r,Y^{0,x,n}_{t,r}),(\nabla_x u^\infty)(r,Y^{0,x,n}_{t,r})\rangle\Big]\,\mathrm{d}r+\int_t^T((\nabla_x u^\infty)\sigma)(r,Y^{0,x,n}_{t,r})\,\mathrm{d}W_r\\&=-\int_t^T f(r,Y^{0,x,n}_{t,r},u^\infty(r,Y^{0,x,n}_{t,r}),(\nabla_x u^\infty)(r,Y^{0,x,n}_{t,r}))\,\mathrm{d}r\\&\qquad+\int_t^T((\nabla_x u^\infty)\sigma)(r,Y^{0,x,n}_{t,r})\,\mathrm{d}W_r.\end{aligned} \tag{3.122}$$

Combining this with (3.118) and (3.121) implies that for all $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that $\mathbb{E}[\sup_{s\in[t,T]} \|\int_t^s ((\nabla_x u^\infty)\sigma)(r, Y_{t,r}^{0,x,n})\, dW_r\|] < \infty$. This proves that for all $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$\int_t^T ((\nabla_x u^\infty)\sigma)(r, Y_{t,r}^{0,x,n})\, dW_r = 0. \tag{3.123}$$

Combining this with (3.122) demonstrates that for all $t \in [0,T)$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &u^\infty(t,x) \\ &= \mathbb{E}\big[g(Y_{t,T}^{0,x,n})\big] + \mathbb{E}\bigg[\int_t^T f(r, Y_{t,r}^{0,x,n}, u^\infty(r, Y_{t,r}^{0,x,n}), (\nabla_x u^\infty)(r, Y_{t,r}^{0,x,n}))\, dr\bigg] \\ &= \mathbb{E}\big[g(Y_{t,T}^{0,x,n})\big] + \mathbb{E}\bigg[\int_t^T F(v^\infty)(r, Y_{t,r}^{0,x,n})\, dr\bigg]. \end{aligned} \tag{3.124}$$

This establishes item (i). Next observe that item (i) of Lemma 3.1.5 shows that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$Y_{t,s}^{\theta,x,n} = x + \int_t^s \mu(Y_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n})\, dr + \int_t^s \sigma(Y_{t,\max\{t,\lfloor r \rfloor_t\}}^{\theta,x,n})\, dW_r^\theta. \tag{3.125}$$

Item (ii) of Lemma 3.1.2, item (i) of Lemma 3.1.6, and Theorem 2.2.3 (applied for every $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ with $O \curvearrowleft \mathbb{R}^d$, $\mu \curvearrowleft ([0,T] \times \mathbb{R}^d \ni (s,y) \mapsto \mu(Y_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n}) \in \mathbb{R}^d)$, $\sigma \curvearrowleft ([0,T] \times \mathbb{R}^d \ni (s,y) \mapsto \sigma(Y_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n}) \in \mathbb{R}^{d\times d})$, $f \curvearrowleft g$ in the notation of Theorem 2.2.3) therefore demonstrate that for all $\theta \in \Theta$, $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$\frac{\partial}{\partial x}\mathbb{E}\big[g(Y_{t,T}^{\theta,x,n})\big] = \mathbb{E}\bigg[g(Y_{t,T}^{\theta,x,n})\tfrac{1}{T-t}\int_t^T (\sigma(Y_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n}))^{-1}\frac{\partial}{\partial x}Y_{t,s}^{\theta,x,n}\, dW_s^\theta\bigg]. \tag{3.126}$$

Moreover, note that Item (ii) of Lemma 3.1.2, item (i) of Lemma 3.1.6, and Theorem 2.2.3 (applied for every $\theta \in \Theta$, $t \in [0,T)$, $x \in \mathbb{R}^d$ with $O \curvearrowleft \mathbb{R}^d$, $\mu \curvearrowleft ([0,T] \times \mathbb{R}^d \ni (s,y) \mapsto \mu(Y_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n}) \in \mathbb{R}^d)$, $\sigma \curvearrowleft ([0,T] \times \mathbb{R}^d \ni (s,y) \mapsto \sigma(Y_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n}) \in \mathbb{R}^{d\times d})$ in the notation of Theorem 2.2.3) ensure that for all $\theta \in \Theta$, $t \in [0,T)$, $r \in (t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\frac{\partial}{\partial x}\mathbb{E}[(F(v^\infty))(r, Y_{t,r}^{\theta,x,n})] \\ &= \mathbb{E}\bigg[(F(v^\infty))(r, Y_{t,r}^{\theta,x,n})\tfrac{1}{r-t}\int_t^r (\sigma(Y_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n}))^{-1}\frac{\partial}{\partial x}Y_{t,s}^{\theta,x,n}\, dW_s^\theta\bigg]. \end{aligned} \tag{3.127}$$

Leibniz integral rule hence implies that for all $\theta \in \Theta$, $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\frac{\partial}{\partial x}\int_t^T \mathbb{E}\big[(F(v^\infty))(r, Y_{t,r}^{\theta,x,n})\big]\, dr = \int_t^T \frac{\partial}{\partial x}\mathbb{E}\big[(F(v^\infty))(r, Y_{t,r}^{\theta,x,n})\big]\, dr \\ &= \int_t^T \mathbb{E}\bigg[(F(v^\infty))(r, Y_{t,r}^{\theta,x,n})\tfrac{1}{r-t}\int_t^r (\sigma(Y_{t,\max\{t,\lfloor s \rfloor_t\}}^{\theta,x,n}))^{-1}\frac{\partial}{\partial x}Y_{t,s}^{\theta,x,n}\, dW_s^\theta\bigg]\, dr. \end{aligned} \tag{3.128}$$

Combining this and (3.126) establishes item (ii). The proof of Lemma 3.2.1 is thus complete. □

3.2.2 Error analysis for multi-level Picard approximations with GauSS-Legendre quadrature rules

The following lemma, Lemma 3.2.2, establishes an upper bound for a specific error appearing in (3.151) in the proof of Theorem 3.2.4 which arises from the application of Lemma 3.2.1.

Lemma 3.2.2. *Assume Setting 3.1.1, let $n \in \mathbb{N}$, and assume that*

$$\sup_{x\in\mathbb{R}^d} \tfrac{|g(x)|}{\varphi(T,x)} + \sup_{t\in[0,T]} \sup_{x\in\mathbb{R}^d} \tfrac{\|v^\infty(t,x)\|}{\varphi(t,x)} + \sup_{t\in[0,T]} \sup_{x\in\mathbb{R}^d} \tfrac{|(F(0))(t,x)|}{\varphi(t,x)} < \infty. \tag{3.129}$$

Then

(i) it holds for all $\theta \in \Theta$, $t \in [0,T)$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} &\left\|\mathbb{E}\left[g(Y^{\theta,x,n}_{t,T})\left(Z^{\theta,x,n}_{t,T} - \left(\frac{1}{\frac{1}{T-t}\int_t^T (\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1} D^{\theta,x,n}_{t,s}\,\mathrm{d}W^\theta_s}\right)\right)\right]\right\| \\ &\leqslant \left[\sup_{y\in\mathbb{R}^d}|g(y)|\right] \\ &\quad\cdot 2K\sqrt{\frac{1}{\alpha}\left(\frac{(T-t)}{n^2} + \frac{1}{n}\right)\left(1 + 3K(T-t)\exp\left(5K(T-t)\right)\right)} \end{aligned} \tag{3.130}$$

and

(ii) it holds for all $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} &\left\|\mathbb{E}\left[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)(F(v^\infty))(s, Y^{\theta,x,n}_{t,s})\right.\right. \\ &\qquad \left.\left.\cdot\left(Z^{\theta,x,n}_{t,s} - \left(\frac{1}{\frac{1}{s-t}\int_t^s (\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}))^{-1} D^{\theta,x,n}_{t,r}\,\mathrm{d}W^\theta_r}\right)\right)\right]\right\| \\ &\leqslant \left[\sup_{s\in[t,T]}\sup_{y\in\mathbb{R}^d}\left(|(F(0))(s,y)| + \mathcal{L}\|v^\infty(s,y)\|\right)\right] \\ &\quad\cdot 4K\sqrt{\frac{1}{\alpha}\left(\frac{(T-t)^3}{n^2} + \frac{(T-t)^2}{n}\right)\left(1 + 3K(T-t)\exp\left(5K(T-t)\right)\right)}. \end{aligned} \tag{3.131}$$

Proof of Lemma 3.2.2. First note that (3.14) ensures that for all $i, j \in \{1, 2, \dots, d\}$, $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\mathbb{E}\left[\left|(D^{\theta,x,n}_{t,s})_{ij} - (D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{ij}\right|^2\right] \\ &= \mathbb{E}\left[\left|\sum_{k=1}^d \frac{\partial \mu_i}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{kj}(s - \max\{t,\lfloor s\rfloor_t\})\right.\right. \\ &\qquad \left.\left. + \sum_{k,l=1}^d \frac{\partial \sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{kj}(W^{\theta,l}_s - W^{\theta,l}_{\max\{t,\lfloor s\rfloor_t\}})\right|^2\right]. \end{aligned} \tag{3.132}$$

The fact that for all $a, b \in \mathbb{R}$ it holds that $(a+b)^2 \leqslant 2(a^2 + b^2)$ hence implies that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\mathbb{E}\left[\left\|D^{\theta,x,n}_{t,s} - D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}\right\|_F^2\right] \\ &\leqslant 2\mathbb{E}\left[\sum_{i,j=1}^d \left|\sum_{k=1}^d \frac{\partial \mu_i}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{kj}(s - \max\{t,\lfloor s\rfloor_t\})\right|^2\right] \\ &\quad + 2\mathbb{E}\left[\sum_{i,j=1}^d \left|\sum_{k,l=1}^d \frac{\partial \sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{kj}(W^{\theta,l}_s - W^{\theta,l}_{\max\{t,\lfloor s\rfloor_t\}})\right|^2\right]. \end{aligned} \tag{3.133}$$

Next observe that (3.7) and the fact that for all $a, b \in [0, \infty)$ it holds that $a^2 + b^2 \leqslant (a+b)^2 \leqslant 2(a^2+b^2)$ show that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&2\mathbb{E}\bigg[\sum_{i,j=1}^{d}\bigg|\sum_{k=1}^{d}\frac{\partial\mu_i}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{kj}(s-\max\{t,\lfloor s\rfloor_t\})\bigg|^2\bigg]\\
&\leqslant 4(s-\max\{t,\lfloor s\rfloor_t\})^2\Bigg(\mathbb{E}\bigg[\sum_{i,j=1}^{d}\bigg|\sum_{k=1}^{d}\frac{\partial\mu_i}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})\delta_{kj}\bigg|^2\bigg]\\
&\quad+\mathbb{E}\bigg[\sum_{i,j=1}^{d}\bigg|\sum_{k=1}^{d}\frac{\partial\mu_i}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})\big((D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{kj}-\delta_{kj}\big)\bigg|^2\bigg]\Bigg)\\
&\leqslant \frac{4(T-t)^2}{n^2}\Bigg(\mathbb{E}\bigg[\sum_{i=1}^{d}\bigg|\sum_{k=1}^{d}\frac{\partial\mu_i}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})\bigg|^2\bigg]\\
&\quad+\mathbb{E}\bigg[\sum_{j=1}^{d}\Big[\max_{a\in\{1,2,\ldots,d\}}\big((D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{aj}-\delta_{aj}\big)^2\Big]\sum_{i=1}^{d}\bigg|\sum_{k=1}^{d}\frac{\partial\mu_i}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})\bigg|^2\bigg]\Bigg)\\
&\leqslant \frac{4K^2(T-t)^2}{n^2}\bigg(1+\mathbb{E}\Big[\big\|D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}-I_d\big\|_F^2\Big]\bigg).
\end{aligned} \tag{3.134}$$

Moreover, note that item (i) of Lemma 3.1.2 the fact that for all $n \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $r \in [s,T]$, $x \in \mathbb{R}^d$ it holds that $Y^{\theta,n,x}_{t,s}$ and $W^\theta_r - W^\theta_s$ are independent, and the fact that for all $a, b \in \mathbb{R}$ it holds that $(a+b)^2 \leqslant 2(a^2+b^2)$ ensure that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&2\mathbb{E}\bigg[\sum_{i,j=1}^{d}\bigg|\sum_{k,l=1}^{d}\frac{\partial\sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})(D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{kj}(W^{\theta,l}_s-W^{\theta,l}_{\max\{t,\lfloor s\rfloor_t\}})\bigg|^2\bigg]\\
&\leqslant 4\mathbb{E}\Big[\max_{a\in\{1,2,\ldots,d\}}\big|(W^{\theta,a}_s-W^{\theta,a}_{\max\{t,\lfloor s\rfloor_t\}})\big|^2\Big]\Bigg(\mathbb{E}\bigg[\sum_{i,j=1}^{d}\bigg|\sum_{k,l=1}^{d}\frac{\partial\sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})\delta_{kj}\bigg|^2\bigg]\\
&\quad+\mathbb{E}\bigg[\sum_{i,j=1}^{d}\bigg|\sum_{k,l=1}^{d}\frac{\partial\sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})\big((D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{kj}-\delta_{kj}\big)\bigg|^2\bigg]\Bigg)\\
&\leqslant 4(s-\max\{t,\lfloor s\rfloor_t\})\Bigg(\mathbb{E}\bigg[\sum_{i=1}^{d}\bigg|\sum_{k,l=1}^{d}\frac{\partial\sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})\bigg|^2\bigg]\\
&\quad+\mathbb{E}\bigg[\sum_{j=1}^{d}\Big[\max_{a\in\{1,2,\ldots,d\}}\big((D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})_{aj}-\delta_{aj}\big)^2\Big]\sum_{i=1}^{d}\bigg|\sum_{k,l=1}^{d}\frac{\partial\sigma_{il}}{\partial x_k}(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}})\bigg|^2\bigg]\Bigg)\\
&\leqslant \frac{4K^2(T-t)}{n}\bigg(1+\mathbb{E}\Big[\big\|D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}-I_d\big\|_F^2\Big]\bigg).
\end{aligned} \tag{3.135}$$

Combining this with item (iii) of Lemma 3.1.7, (3.133), and (3.134) proves that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\mathbb{E}\Big[\big\|D^{\theta,x,n}_{t,s}-D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}\big\|_F^2\Big]\\
&\leqslant 4K^2\Big(\frac{(T-t)^2}{n^2}+\frac{(T-t)}{n}\Big)\bigg(1+\mathbb{E}\Big[\big\|D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}-I_d\big\|_F^2\Big]\bigg)\\
&\leqslant 4K^2\Big(\frac{(T-t)^2}{n^2}+\frac{(T-t)}{n}\Big)\Big(1+3K(s-t)\exp\big(5K(s-t)\big)\Big).
\end{aligned} \tag{3.136}$$

Item (iv) of Lemma 1.2.2, the Burkholder-Davis-Gundy inequality, and the fact that for all $t \in [0,T]$, $r \in [t,T]$, $x \in \mathbb{R}^d$ it holds that $(Z^{0,x,n}_{t,r})_1 = 1$ therefore imply that for all $\theta \in \Theta$, $t \in [0,T]$, $r \in (t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left\|Z^{\theta,x,n}_{t,r} - \begin{pmatrix} 1 \\ \frac{1}{r-t}\int_t^r (\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1} D^{\theta,x,n}_{t,s}\,\mathrm{d}W^\theta_s \end{pmatrix}\right\|^2\right]\right)^{\frac{1}{2}} \\
&= \left(\mathbb{E}\left[\left\|\frac{1}{r-t}\int_t^r (\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1}\left[D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}} - D^{\theta,x,n}_{t,s}\right]\mathrm{d}W^\theta_s\right\|^2\right]\right)^{\frac{1}{2}} \\
&\leqslant \frac{1}{r-t}\left(\int_t^r \mathbb{E}\left[\left\|(\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1}\right\|^2_{L(\mathbb{R}^d)}\left\|D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}} - D^{\theta,x,n}_{t,s}\right\|^2_F\right]\mathrm{d}s\right)^{\frac{1}{2}} \\
&\leqslant \frac{1}{(r-t)\sqrt{\alpha}}\left(\int_t^r \mathbb{E}\left[\left\|D^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}} - D^{\theta,x,n}_{t,s}\right\|^2_F\right]\mathrm{d}s\right)^{\frac{1}{2}} \\
&\leqslant 2K\sqrt{\frac{1}{\alpha(r-t)}\left(\frac{(T-t)^2}{n^2} + \frac{(T-t)}{n}\right)\left(1 + 3K(T-t)\exp\left(5K(T-t)\right)\right)}.
\end{aligned} \tag{3.137}$$

This implies that for all $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left\|Z^{\theta,x,n}_{t,T} - \begin{pmatrix} 1 \\ \frac{1}{T-t}\int_t^T (\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1} D^{\theta,x,n}_{t,s}\,\mathrm{d}W^\theta_s \end{pmatrix}\right\|^2\right]\right)^{\frac{1}{2}} \\
&\leqslant 2K\sqrt{\frac{1}{\alpha}\left(\frac{(T-t)}{n^2} + \frac{1}{n}\right)\left(1 + 3K(T-t)\exp\left(5K(T-t)\right)\right)}.
\end{aligned} \tag{3.138}$$

Hölder's inequality hence shows that for all $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\left\|\mathbb{E}\left[g(Y^{\theta,x,n}_{t,T})\left(Z^{\theta,x,n}_{t,T} - \begin{pmatrix} 1 \\ \frac{1}{T-t}\int_t^T (\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1} D^{\theta,x,n}_{t,s}\,\mathrm{d}W^\theta_s \end{pmatrix}\right)\right]\right\| \\
&\leqslant \left[\sup_{y\in\mathbb{R}^d}|g(y)|\right]\mathbb{E}\left[\left\|Z^{\theta,x,n}_{t,T} - \begin{pmatrix} 1 \\ \frac{1}{T-t}\int_t^T (\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1} D^{\theta,x,n}_{t,s}\,\mathrm{d}W^\theta_s \end{pmatrix}\right\|\right] \\
&\leqslant \left[\sup_{y\in\mathbb{R}^d}|g(y)|\right]2K\sqrt{\frac{1}{\alpha}\left(\frac{(T-t)}{n^2} + \frac{1}{n}\right)\left(1 + 3K(T-t)\exp\left(5K(T-t)\right)\right)}.
\end{aligned} \tag{3.139}$$

This establishes item (i). Next note that Hölder's inequality, item (i) of Lemma 3.1.3, and (3.137) prove that for all $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\left\| \mathbb{E}\Bigg[\sum_{s \in (t,T)} q^{Q,[t,T]}(s)(F(v^\infty))(s, Y^{\theta,x,n}_{t,s}) \\
&\qquad \cdot \left(Z^{\theta,x,n}_{t,s} - \left(\frac{1}{\frac{1}{s-t}\int_t^s (\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}}))^{-1} D^{\theta,x,n}_{t,r}\, \mathrm{d}W^\theta_r} \right)\right)\Bigg] \right\| \\
&\leqslant \sum_{s \in (t,T)} q^{Q,[t,T]}(s)\mathbb{E}\Bigg[|(F(v^\infty))(s, Y^{\theta,x,n}_{t,s})| \\
&\qquad \cdot \left\| Z^{\theta,x,n}_{t,s} - \left(\frac{1}{\frac{1}{s-t}\int_t^s (\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}}))^{-1} D^{\theta,x,n}_{t,r}\, \mathrm{d}W^\theta_r} \right)\right\| \Bigg] \\
&\leqslant \left[\sup_{s \in [t,T]} \sup_{y \in \mathbb{R}^d} |(F(v^\infty))(s,y)| \right] \sum_{s \in (t,T)} q^{Q,[t,T]}(s) \\
&\qquad \cdot \mathbb{E}\left[\left\| Z^{\theta,x,n}_{t,s} - \left(\frac{1}{\frac{1}{s-t}\int_t^s (\sigma(Y^{\theta,x,n}_{t,\max\{t,\lfloor r \rfloor_t\}}))^{-1} D^{\theta,x,n}_{t,r}\, \mathrm{d}W^\theta_r} \right)\right\| \right] \\
&\leqslant \left[\sup_{s \in [t,T]} \sup_{y \in \mathbb{R}^d} |(F(v^\infty))(s,y)| \right] \sum_{s \in (t,T)} \frac{q^{Q,[t,T]}(s)}{\sqrt{s-t}} \\
&\qquad \cdot 2K\sqrt{\frac{1}{\alpha}\left(\frac{(T-t)^2}{n^2} + \frac{(T-t)}{n}\right)}\Big(1 + 3K(T-t)\exp\big(5K(T-t)\big)\Big) \\
&\leqslant \left[\sup_{s \in [t,T]} \sup_{y \in \mathbb{R}^d} |(F(v^\infty))(s,y)| \right] \\
&\qquad \cdot 4K\sqrt{\frac{1}{\alpha}\left(\frac{(T-t)^3}{n^2} + \frac{(T-t)^2}{n}\right)}\Big(1 + 3K(T-t)\exp\big(5K(T-t)\big)\Big).
\end{aligned} \tag{3.140}$$

This and (3.9) demonstrate that for all $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\left\| \mathbb{E}\Bigg[\sum_{s \in (t,T)} q^{Q,[t,T]}(s)(F(v^\infty))(s, Y^{\theta,x,n}_{t,s}) \\
&\qquad \cdot \left(Z^{\theta,x,n}_{t,s} - \frac{1}{s-t}\int_t^s (\sigma(Y^{\theta,x,n}_{t,\max\{\lfloor r \rfloor_t\}}))^{-1} D^{\theta,x,n}_{t,r}\, \mathrm{d}W^\theta_r \right)\Bigg] \right\| \\
&\leqslant \left[\sup_{s \in [t,T]} \sup_{y \in \mathbb{R}^d} \left(|(F(0))(s,y)| + \mathcal{L}\|v^\infty(s,y)\| \right) \right] \\
&\qquad \cdot 4K\sqrt{\frac{1}{\alpha}\left(\frac{(T-t)^3}{n^2} + \frac{(T-t)^2}{n}\right)}\Big(1 + 3K(T-t)\exp\big(5K(T-t)\big)\Big).
\end{aligned} \tag{3.141}$$

This establishes item (ii). The proof of Lemma 3.2.2 is thus complete. □

The following theorem, Theorem 3.2.4, provides a recursive upper bound for the approximation error of the MLP approximation scheme. In the proof we follow the idea of [59, Lemma 4.3] and first study the Monte Carlo error and the time discretization error separately before combining them with a discrete Gronwall inequality to obtain a global error bound.

Lemma 3.2.3. *Assume Setting 3.1.1, let $n, M, Q \in \mathbb{N}$, assume that*

$$\sup_{x \in \mathbb{R}^d} \tfrac{|g(x)|}{\varphi(T,x)} + \sup_{t \in [0,T]} \sup_{x \in \mathbb{R}^d} \tfrac{|(F(0))(t,x)|}{\varphi(t,x)} + \sup_{t \in [0,T]} \sup_{x \in \mathbb{R}^d} \tfrac{\|v^\infty(t,x)\|}{\varphi(t,x)} < \infty, \tag{3.142}$$

and let $\varepsilon\colon [0,T]\times\mathbb{R}^d\to[0,\infty]^{d+1}$ *satisfy for all* $t\in[0,T]$, $x\in\mathbb{R}^d$ *that*

$$\begin{aligned}
&\varepsilon(t,x)\\
&=\mathbb{E}\Bigg[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)(F(v^\infty))(s,Y^{0,x,n}_{t,s})\begin{pmatrix}1\\ \frac{1}{s-t}\int_t^s(\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}))^{-1}D^{0,x,n}_{t,r}\,\mathrm{d}W^0_r\end{pmatrix}\\
&\qquad-\int_t^T(F(v^\infty))(r,Y^{0,x,n}_{t,r})\begin{pmatrix}1\\ \frac{1}{r-t}\int_t^r(\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1}D^{0,x,n}_{t,s}\,\mathrm{d}W^0_s\end{pmatrix}\mathrm{d}r\Bigg].
\end{aligned}\tag{3.143}$$

Then for all $m,k\in\mathbb{N}$, $t_0\in[0,T)$, $x\in\mathbb{R}^d$, $\nu_0\in\{1,2,\ldots,d+1\}$ *it holds that*

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t_0,x)-v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac12}\\
&\leqslant\sum_{j=0}^{k-1}\sum_{\substack{l_1,\ldots,l_{j+1}\in\mathbb{N}\\ l_1<\ldots<l_{j+1}=m}}\sum_{\substack{t_1,\ldots,t_{j+1}\in\mathbb{R}\\ t_0<t_1<\ldots<t_{j+1}\leqslant T}}\sum_{\nu_1,\ldots,\nu_{j+1}\in\{1,\ldots,d+1\}}\frac{2^j}{\sqrt{M^{m-j-l_1}}}\left[\prod_{i=1}^j L_{\nu_i}q^{Q,[t_{i-1},T]}(t_i)\right]\\
&\quad\cdot\Bigg\{\mathbb{1}_1(\nu_{j+1})\Bigg(\mathbb{1}_T(t_{j+1})\Bigg(\Bigg(\left(\mathbb{E}\left[\left|(\varepsilon(t_j,Y^{0,x,n}_{t_0,t_j}))_{\nu_j}\prod_{i=1}^j(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac12}\\
&\quad+\frac{1}{\sqrt{M^{l_1}}}\left(\mathbb{E}\left[\left|(g(Y^{0,x,n}_{t_0,T})-g(Y^{0,x,n}_{t_0,t_j}))\prod_{i=1}^{j+1}(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac12}\\
&\quad+\left(2K\sqrt{\frac1\alpha\Big(\frac{(T-t_0)}{n^2}+\frac1n\Big)}\Big(1+3K(T-t_0)\exp\big(5K(T-t_0)\big)\Big)\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right]\right.\\
&\quad+4K\sqrt{\frac1\alpha\Big(\frac{(T-t_0)^3}{n^2}+\frac{(T-t_0)^2}{n}\Big)}\Big(1+3K(T-t_0)\exp\big(5K(T-t_0)\big)\Big)\\
&\quad\cdot\left.\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}(|(F(0))(s,y)|+\mathcal{L}|\!|\!|v^\infty(s,y)|\!|\!|)\right]\right)\left(\mathbb{E}\left[\left|\prod_{i=1}^j(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)\Bigg)\\
&\quad+\frac{q^{Q,[t_j,T]}(t_{j+1})}{\sqrt{M^{l_1}}}\left(\mathbb{E}\left[\left|(F(0))(t_{j+1},Y^{0,x,n}_{t_0,t_{j+1}})\prod_{i=1}^{j+1}(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac12}\Bigg)\\
&\quad+\frac{L_{\nu_{j+1}}q^{Q,[t_j,T]}(t_{j+1})}{\sqrt{M^{l_1-1}}}\left(\mathbb{E}\left[\left|(v^\infty(t_{j+1},Y^{0,x,n}_{t_0,t_{j+1}}))_{\nu_{j+1}}\prod_{i=1}^{j+1}(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac12}\Bigg\}\\
&\quad+\sum_{\substack{l_1,\ldots,l_k\in\mathbb{N}\\ l_1<\ldots<l_k<m}}\sum_{\substack{t_1,\ldots,t_k\in\mathbb{R}\\ t_0<t_1<\ldots<t_k<T}}\sum_{\nu_1,\ldots,\nu_k\in\{1,\ldots,d+1\}}\frac{2^k}{\sqrt{M^{m-k-l_1}}}\left[\prod_{i=1}^k L_{\nu_i}q^{Q,[t_{i-1},T]}(t_i)\right]\\
&\quad\cdot\left(\mathbb{E}\left[\left|((V^0_{n,l_1,M,Q}-v^\infty)(t_k,Y^{0,x,n}_{t_0,t_k}))_{\nu_k}\prod_{i=1}^k(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac12}.
\end{aligned}\tag{3.144}$$

Proof of Lemma 3.2.3. First we derive the Monte Carlo error. Note that items (i) and (ii) of Lemma 3.1.9, and (3.17) imply that for all $m\in\mathbb{N}$, $t\in[0,T)$, $x\in\mathbb{R}^d$, $\nu\in\{1,2,\ldots,d+1\}$

it holds that

$$
\begin{aligned}
&\operatorname{Var}((V^0_{n,m,M,Q}(t,x))_\nu)\\
&= \frac{1}{M^m}\operatorname{Var}\big((g(Y^{0,x,n}_{t,T})-g(x))(Z^{0,x,n}_{t,T})_\nu\big)\\
&\quad+\sum_{l=0}^{m-1}\frac{1}{M^{m-l}}\operatorname{Var}\Bigg(\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\big((F(V^{(0,l,1)}_{n,l,M,Q}))\\
&\qquad\qquad-\mathbb{1}_{\mathbb{N}}(l)(F(V^{(0,-l,1)}_{n,l-1,M,Q}))\big)(s,Y^{0,x,n}_{t,s})(Z^{0,x,n}_{t,s})_\nu\Bigg)\\
&\leqslant \frac{1}{M^m}\mathbb{E}\Big[\big|(g(Y^{0,x,n}_{t,T})-g(x))(Z^{0,x,n}_{t,T})_\nu\big|^2\Big]\\
&\quad+\sum_{l=0}^{m-1}\frac{1}{M^{m-l}}\mathbb{E}\Bigg[\Bigg|\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\big((F(V^{(0,l,1)}_{n,l,M,Q}))\\
&\qquad\qquad-\mathbb{1}_{\mathbb{N}}(l)(F(V^{(0,-l,1)}_{n,l-1,M,Q}))\big)(s,Y^{0,x,n}_{t,s})(Z^{0,x,n}_{t,s})_\nu\Bigg|^2\Bigg].
\end{aligned}
\tag{3.145}
$$

This, the triangle inequality, and (3.9) show that for all $m\in\mathbb{N}$, $t\in[0,T)$, $x\in\mathbb{R}^d$, $\nu\in\{1,2,\ldots,d+1\}$ it holds that

$$
\begin{aligned}
&\Big(\mathbb{E}\Big[\big|(V^0_{n,m,M,Q}(t,x)-\mathbb{E}[V^0_{n,m,M,Q}(t,x)])_\nu\big|^2\Big]\Big)^{\frac12}=\Big(\operatorname{Var}\big((V^0_{n,m,M,Q}(t,x))_\nu\big)\Big)^{\frac12}\\
&\leqslant\frac{1}{\sqrt{M^m}}\Big(\mathbb{E}\Big[\big|(g(Y^{0,x,n}_{t,T})-g(x))(Z^{0,x,n}_{t,T})_\nu\big|^2\Big]\Big)^{\frac12}\\
&\quad+\sum_{l=0}^{m-1}\Bigg[\sum_{s\in(t,T)}\frac{q^{Q,[t,T]}(s)}{\sqrt{M^{m-l}}}\\
&\quad\cdot\Big(\mathbb{E}\Big[\big|\big((F(V^{(0,l,1)}_{n,l,M,Q}))-\mathbb{1}_{\mathbb{N}}(l)(F(V^{(0,-l,1)}_{n,l-1,M,Q}))\big)(s,Y^{0,x,n}_{t,s})(Z^{0,x,n}_{t,s})_\nu\big|^2\Big]\Big)^{\frac12}\Bigg]\\
&\leqslant\frac{1}{\sqrt{M^m}}\Big(\mathbb{E}\Big[\big|(g(Y^{0,x,n}_{t,T})-g(x))(Z^{0,x,n}_{t,T})_\nu\big|^2\Big]\Big)^{\frac12}\\
&\quad+\frac{1}{\sqrt{M^m}}\sum_{s\in(t,T)}q^{Q,[t,T]}(s)\Big(\mathbb{E}\Big[\big|(F(0))(s,Y^{0,x,n}_{t,s})(Z^{0,x,n}_{t,s})_\nu\big|^2\Big]\Big)^{\frac12}\\
&\quad+\sum_{l=1}^{m-1}\Bigg[\sum_{s\in(t,T)}\frac{q^{Q,[t,T]}(s)}{\sqrt{M^{m-l}}}\\
&\quad\cdot\Big(\mathbb{E}\Big[\big|\big((F(V^{(0,l,1)}_{n,l,M,Q}))-(F(V^{(0,-l,1)}_{n,l-1,M,Q}))\big)(s,Y^{0,x,n}_{t,s})\big|^2\big|(Z^{0,x,n}_{t,s})_\nu\big|^2\Big]\Big)^{\frac12}\\
&\leqslant\frac{1}{\sqrt{M^m}}\Big(\mathbb{E}\Big[\big|(g(Y^{0,x,n}_{t,T})-g(x))(Z^{0,x,n}_{t,T})_\nu\big|^2\Big]\Big)^{\frac12}\\
&\quad+\frac{1}{\sqrt{M^m}}\sum_{s\in(t,T)}q^{Q,[t,T]}(s)\Big(\mathbb{E}\Big[\big|(F(0))(s,Y^{0,x,n}_{t,s})(Z^{0,x,n}_{t,s})_\nu\big|^2\Big]\Big)^{\frac12}\\
&\quad+\sum_{l=1}^{m-1}\Bigg[\sum_{s\in(t,T)}\frac{q^{Q,[t,T]}(s)}{\sqrt{M^{m-l}}}\\
&\quad\cdot\Bigg(\mathbb{E}\Bigg[\Bigg(\sum_{\nu_1=1}^{d+1}L_{\nu_1}\big|((V^{(0,l,1)}_{n,l,M,Q}-V^{(0,-l,1)}_{n,l-1,M,Q})(s,Y^{0,x,n}_{t,s}))_{\nu_1}\big|\Bigg)^2\big|(Z^{0,x,n}_{t,s})_\nu\big|^2\Bigg]\Bigg)^{\frac12}.
\end{aligned}
\tag{3.146}
$$

The triangle inequality therefore demonstrates that for all $m \in \mathbb{N}$, $t \in [0,T)$, $x \in \mathbb{R}^d$, $\nu \in \{1,2,\dots,d+1\}$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t,x)-\mathbb{E}[V^0_{n,m,M,Q}(t,x)])_\nu\right|^2\right]\right)^{\frac{1}{2}}\\
&\leqslant \frac{1}{\sqrt{M^m}}\left(\mathbb{E}\left[\left|(g(Y^{0,x,n}_{t,T})-g(x))(Z^{0,x,n}_{t,T})_\nu\right|^2\right]\right)^{\frac{1}{2}}\\
&\quad+\frac{1}{\sqrt{M^m}}\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\left(\mathbb{E}\left[\left|(F(0))(s,Y^{0,x,n}_{t,s})(Z^{0,x,n}_{t,s})_\nu\right|^2\right]\right)^{\frac{1}{2}}\\
&\quad+\sum_{l=1}^{m-1}\sum_{s\in(t,T)}\sum_{\nu_1=1}^{d+1}\frac{L_{\nu_1}q^{Q,[t,T]}(s)}{\sqrt{M^{m-l}}}\left(\mathbb{E}\left[\left|((V^0_{n,l,M,Q}-v^\infty)(s,Y^{0,x,n}_{t,s}))_{\nu_1}\right|^2\left|(Z^{0,x,n}_{t,s})_\nu\right|^2\right]\right)^{\frac{1}{2}}\\
&\quad+\sum_{l=1}^{m-1}\sum_{s\in(t,T)}\sum_{\nu_1=1}^{d+1}\frac{L_{\nu_1}q^{Q,[t,T]}(s)}{\sqrt{M^{m-l}}}\left(\mathbb{E}\left[\left|((V^0_{n,l-1,M,Q}-v^\infty)(s,Y^{0,x,n}_{t,s}))_{\nu_1}\right|^2\left|(Z^{0,x,n}_{t,s})_\nu\right|^2\right]\right)^{\frac{1}{2}}\\
&=\frac{1}{\sqrt{M^m}}\left(\mathbb{E}\left[\left|(g(Y^{0,x,n}_{t,T})-g(x))(Z^{0,x,n}_{t,T})_\nu\right|^2\right]\right)^{\frac{1}{2}}\\
&\quad+\frac{1}{\sqrt{M^m}}\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\left(\mathbb{E}\left[\left|(F(0))(s,Y^{0,x,n}_{t,s})(Z^{0,x,n}_{t,s})_\nu\right|^2\right]\right)^{\frac{1}{2}}\\
&\quad+\sum_{l=0}^{m-1}\sum_{s\in(t,T)}\sum_{\nu_1=1}^{d+1}\frac{L_{\nu_1}q^{Q,[t,T]}(s)}{\sqrt{M^{m-l-1}}}\left(\frac{\mathbb{1}_{(0,m)}(l)}{\sqrt{M}}-\mathbb{1}_{(-1,m-1)}(l)\right)\\
&\quad\cdot\left(\mathbb{E}\left[\left|((V^0_{n,l,M,Q}-v^\infty)(s,Y^{0,x,n}_{t,s}))_{\nu_1}\right|^2\left|(Z^{0,x,n}_{t,s})_\nu\right|^2\right]\right)^{\frac{1}{2}}.
\end{aligned} \tag{3.147}$$

In the next step we analyse the time discretization error. Note that item (iii) of Lemma 3.1.9 ensures that for all $m \in \mathbb{N}$, $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
\mathbb{E}[V^0_{n,m,M,Q}(t,x)] &= \mathbb{E}[g(Y^{0,x,n}_{t,T})Z^{0,x,n}_{t,T}]\\
&\quad+\mathbb{E}\Bigg[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)(F(V^0_{n,m-1,M,Q}))(s,Y^{0,x,n}_{t,s})Z^{0,x,n}_{t,s}\Bigg].
\end{aligned} \tag{3.148}$$

In addition, observe that item (ii) of Lemma 3.1.6 and items (i) and (ii) of Lemma 3.2.1 demonstrate that for all $t \in [0,T)$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
v^\infty(t,x) &= \mathbb{E}\left[g(Y^{0,x,n}_{t,T})\begin{pmatrix}1\\ \frac{1}{T-t}\int_t^T(\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1}D^{0,x,n}_{t,s}\,\mathrm{d}W^0_s\end{pmatrix}\right]\\
&\quad+\mathbb{E}\left[\int_t^T (F(v^\infty))(r,Y^{0,x,n}_{t,r})\begin{pmatrix}1\\ \frac{1}{r-t}\int_t^r(\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1}D^{0,x,n}_{t,s}\,\mathrm{d}W^0_s\end{pmatrix}\mathrm{d}r\right].
\end{aligned} \tag{3.149}$$

Combining this with the triangle inequality and (3.148) shows that for all $m \in \mathbb{N}$, $t \in [0,T)$,

$x \in \mathbb{R}^d$, $nu \in \{1,2,\dots,d+1\}$ it holds that

$$\begin{aligned}
&\big|(\mathbb{E}[V^0_{n,m,M,Q}(t,x)] - v^\infty(t,x))_\nu\big| \\
&\leqslant \Bigg|\mathbb{E}\Bigg[g(Y^{0,x,n}_{t,T})\Bigg(Z^{0,x,n}_{t,T} - \Bigg(\frac{1}{\frac{1}{T-t}\int_t^T (\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1} D^{0,x,n}_{t,s}\,\mathrm{d}W^0_s}\Bigg)\Bigg)_\nu\Bigg]\Bigg| \\
&\quad + \Bigg|\mathbb{E}\Bigg[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)(F(V^0_{n,m-1,M,Q}))(s,Y^{0,x,n}_{t,s})(Z^{0,x,n}_{t,s})_\nu\Bigg] \\
&\quad - \mathbb{E}\Bigg[\int_t^T (F(v^\infty))(r,Y^{0,x,n}_{t,r})\Bigg(\frac{1}{\frac{1}{r-t}\int_t^r (\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1} D^{0,x,n}_{t,s}\,\mathrm{d}W^0_s}\Bigg)_\nu\,\mathrm{d}r\Bigg]\Bigg|.
\end{aligned} \tag{3.150}$$

Furthermore, observe that the triangle inequality and (3.9) imply that for all $m \in \mathbb{N}$, $t \in [0,T]$, $s \in (t,T]$, $x \in \mathbb{R}^d$, $\nu \in \{1,2,\dots,d+1\}$ it holds that

$$\begin{aligned}
&\Bigg|\mathbb{E}\Bigg[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)(F(V^0_{n,m-1,M,Q}))(s,Y^{0,x,n}_{t,s})(Z^{0,x,n}_{t,s})_\nu\Bigg] \\
&\qquad - \mathbb{E}\Bigg[\int_t^T (F(v^\infty))(r,Y^{0,x,n}_{t,r})\Bigg(\frac{1}{\frac{1}{r-t}\int_t^r (\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1} D^{0,x,n}_{t,s}\,\mathrm{d}W^0_s}\Bigg)_\nu\,\mathrm{d}r\Bigg]\Bigg| \\
&\leqslant \Bigg|\mathbb{E}\Bigg[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\big((F(V^0_{n,m-1,M,Q})) - (F(v^\infty))\big)(s,Y^{0,x,n}_{t,s})(Z^{0,x,n}_{t,s})_\nu\Bigg]\Bigg| \\
&\quad + \Bigg|\mathbb{E}\Bigg[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)(F(v^\infty))(s,Y^{0,x,n}_{t,s}) \\
&\quad \cdot\Bigg(Z^{0,x,n}_{t,s} - \Bigg(\frac{1}{\frac{1}{s-t}\int_t^s (\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}))^{-1} D^{0,x,n}_{t,r}\,\mathrm{d}W^0_r}\Bigg)\Bigg)_\nu\Bigg]\Bigg| \\
&\quad + \Bigg|\mathbb{E}\Bigg[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)(F(v^\infty))(s,Y^{0,x,n}_{t,s}) \\
&\quad \cdot\Bigg(\frac{1}{\frac{1}{s-t}\int_t^s (\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}))^{-1} D^{0,x,n}_{t,r}\,\mathrm{d}W^0_r}\Bigg)_\nu \\
&\quad - \int_t^T (F(v^\infty))(r,Y^{0,x,n}_{t,r})\Bigg(\frac{1}{\frac{1}{r-t}\int_t^r (\sigma(Y^{0,x,n}_{t,\{t,\lfloor s\rfloor_t\}}))^{-1} D^{0,x,n}_{t,s}\,\mathrm{d}W^0_s}\Bigg)_\nu\,\mathrm{d}r\Bigg]\Bigg|.
\end{aligned} \tag{3.151}$$

In addition, note that (3.9) and Hölder's inequality ensure that for all $t \in [0,T]$, $s \in (t,T]$, $x \in \mathbb{R}^d$, $\nu \in \{1,2,\dots,d+1\}$ it holds that

$$\begin{aligned}
&\Bigg|\mathbb{E}\Bigg[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\big((F(V^0_{n,m-1,M,Q})) - (F(v^\infty))\big)(s,Y^{0,x,n}_{t,s})(Z^{0,x,n}_{t,s})_\nu\Bigg]\Bigg| \\
&\leqslant \mathbb{E}\Bigg[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\big|\big((F(V^0_{n,m-1,M,Q})) - (F(v^\infty))\big)(s,Y^{0,x,n}_{t,s})\big|\,|(Z^{0,x,n}_{t,s})_\nu|\Bigg] \\
&\leqslant \sum_{s\in(t,T)}\sum_{\nu_1=1}^{d+1} L_{\nu_1} q^{Q,[t,T]}(s)\,\mathbb{E}\Big[|((V^0_{n,m-1,M,Q} - v^\infty)(s,Y^{0,x,n}_{t,s}))_{\nu_1}|\,|(Z^{0,x,n}_{t,s})_\nu|\Big] \\
&\leqslant \sum_{s\in(t,T)}\sum_{\nu=1}^{d+1} L_{\nu_1} q^{Q,[t,T]}(s)\Big(\mathbb{E}\Big[|((V^0_{n,m-1,M,Q} - v^\infty)(s,Y^{0,x,n}_{t,s}))_{\nu_1}|^2|(Z^{0,x,n}_{t,s})|^2\Big]\Big)^{\frac{1}{2}}.
\end{aligned} \tag{3.152}$$

In the next step we combine the bounds on the Monte Carlo error and on the time discretization error to analyse the global error. Observe that the triangle inequality, (3.143), (3.147), (3.150), (3.151), and (3.152) demonstrate that for all $m \in \mathbb{N}$, $t \in [0,T)$, $x \in \mathbb{R}^d$, $\nu \in \{1,2,\ldots,d+1\}$ it holds that

$$
\begin{aligned}
&\Big(\mathbb{E}\Big[\big|(V^0_{n,m,M,Q}(t,x) - v^\infty(t,x))_\nu\big|^2\Big]\Big)^{\frac{1}{2}}\\
&\leqslant \Big(\mathbb{E}\Big[\big|(V^0_{n,m,M,Q}(t,x) - \mathbb{E}[V^0_{n,m,M,Q}(t,x)])_\nu\big|^2\Big]\Big)^{\frac{1}{2}}\\
&\quad + \Big|(\mathbb{E}[V^0_{n,m,M,Q}(t,x)] - v^\infty(t,x))_\nu\Big|\\
&\leqslant (\varepsilon(t,x))_\nu + \frac{1}{\sqrt{M^m}}\Big(\mathbb{E}\Big[\big|(g(Y^{0,x,n}_{t,T}) - g(x))(Z^{0,x,n}_{t,T})_\nu\big|^2\Big]\Big)^{\frac{1}{2}}\\
&\quad + \frac{1}{\sqrt{M^m}}\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\Big(\mathbb{E}\Big[\big|(F(0))(s,Y^{0,x,n}_{t,s})(Z^{0,x,n}_{t,s})_\nu\big|^2\Big]\Big)^{\frac{1}{2}}\\
&\quad + \sum_{l=0}^{m-1}\sum_{s\in(t,T)}\sum_{\nu_1=1}^{d+1}\tfrac{L_{\nu_1}q^{Q,[t,T]}(s)}{\sqrt{M^{m-l-1}}}\Big(\tfrac{\mathbb{1}_{(0,m)}(l)}{\sqrt{M}} - \mathbb{1}_{(-1,m-1)}(l)\Big)\\
&\quad \cdot\Big(\mathbb{E}\Big[\big|((V^0_{n,l,M,Q} - v^\infty)(s,Y^{0,x,n}_{t,s}))_{\nu_1}\big|^2\big|(Z^{0,x,n}_{t,s})_\nu\big|^2\Big]\Big)^{\frac{1}{2}}\\
&\quad + \left|\mathbb{E}\left[g(Y^{0,x,n}_{t,T})\left(Z^{0,x,n}_{t,T} - \left(\frac{1}{\frac{1}{T-t}\int_t^T(\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1}D^{0,x,n}_{t,s}\,\mathrm{d}W^0_s}\right)\right)_\nu\right]\right|\\
&\quad + \sum_{s\in(t,T)}\sum_{\nu_1=1}^{d+1} L_{\nu_1}q^{Q,[t,T]}(s)\\
&\quad \cdot\Big(\mathbb{E}\Big[\big|((V^0_{n,m-1,M,Q} - v^\infty)(s,Y^{0,x,n}_{t,s}))_{\nu_1}\big|^2\big|(Z^{0,x,n}_{t,s})_\nu\big|^2\Big]\Big)^{\frac{1}{2}}\\
&\quad + \Bigg|\mathbb{E}\Bigg[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)(F(v^\infty))(s,Y^{0,x,n}_{t,s})\\
&\quad \cdot\left(Z^{0,x,n}_{t,s} - \left(\frac{1}{\frac{1}{s-t}\int_t^s(\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}))^{-1}D^{0,x,n}_{t,r}\,\mathrm{d}W^0_r}\right)\right)_\nu\Bigg]\Bigg|.
\end{aligned}
\tag{3.153}
$$

Items (i) and (ii) of Lemma 3.2.2 therefore prove that for all $t \in [0,T)$, $x \in \mathbb{R}^d$, $\nu \in$

$\{1, 2, \ldots, d+1\}$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t,x) - v^\infty(t,x))_\nu\right|^2\right]\right)^{\frac{1}{2}} \\
&\leqslant (\varepsilon(t,x))_\nu + \frac{1}{\sqrt{M^m}}\left(\mathbb{E}\left[\left|(g(Y^{0,x,n}_{t,T}) - g(x))(Z^{0,x,n}_{t,T})_\nu\right|^2\right]\right)^{\frac{1}{2}} \\
&\quad + \frac{1}{\sqrt{M^m}}\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\left(\mathbb{E}\left[\left|(F(0))(s,Y^{0,x,n}_{t,s})(Z^{0,x,n}_{t,s})_\nu\right|^2\right]\right)^{\frac{1}{2}} \\
&\quad + \sum_{l=1}^{m-1}\sum_{s\in(t,T)}\sum_{\nu_1=1}^{d+1}\tfrac{2L_{\nu_1}q^{Q,[t,T]}(s)}{\sqrt{M^{m-l-1}}} \\
&\quad \cdot \left(\mathbb{E}\left[\left|((V^0_{n,l,M,Q} - v^\infty)(s,Y^{0,x,n}_{t,s}))_{\nu_1}\right|^2\left|(Z^{0,x,n}_{t,s})_\nu\right|^2\right]\right)^{\frac{1}{2}} \\
&\quad + 2K\sqrt{\frac{1}{\alpha}\left(\frac{(T-t)}{n^2}+\frac{1}{n}\right)\Big(1+3K(T-t)\exp\big(5K(T-t)\big)\Big)}\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right] \\
&\quad + 4K\sqrt{\frac{1}{\alpha}\left(\frac{(T-t)^3}{n^2}+\frac{(T-t)^2}{n}\right)\Big(1+3K(T-t)\exp\big(5K(T-t)\big)\Big)} \\
&\quad \cdot\left[\sup_{s\in[t,T]}\sup_{y\in\mathbb{R}^d}\left(|(F(0))(s,y)| + \mathcal{L}|\!|\!|v^\infty(s,y)|\!|\!|\right)\right] \\
&\quad + \sum_{s\in(t,T)}\sum_{\nu_1=1}^{d+1}\frac{L_{\nu_1}q^{Q,[t,T](s)}}{\sqrt{M^{m-1}}}\left(\mathbb{E}\left[\left|(v^\infty(s,Y^{0,x,n}_{t,s}))_{\nu_1}\right|^2\left|(Z^{0,x,n}_{t,s})_\nu\right|^2\right]\right)^{\frac{1}{2}}.
\end{aligned} \tag{3.154}$$

Next we prove (3.144) by induction on $k \in \mathbb{N}$. Note that (3.154) establishes (3.144) in the base case $k = 1$. For the induction step assume that (3.144) holds for $k \in \mathbb{N}$. Item (v) of Lemma 3.1.4 ensures that for all $l_1 \in \mathbb{N}$, $t_0, t_1, \ldots, t_k \in [0,T)$, $x \in \mathbb{R}^d$ with $t_0 < t_1 < \ldots < t_k < T$, $\nu_0, \nu_1, \ldots, \nu_k \in \{1, 2, \ldots, d+1\}$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,l_1,M,Q}(t_k,Y^{0,x,n}_{t_0,t_k}) - v^\infty(t_k,Y^{0,x,n}_{t_0,t_k}))_{\nu_k}\right|^2\prod_{i=1}^{k}\left|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}} \\
&= \left(\mathbb{E}\left[\mathbb{E}\left[\left|(V^0_{n,l_1,M,Q} - v^\infty)(t_k,y)\right|^2\right]\Big|_{y=Y^{0,x,n}_{t_0,t_k}}\prod_{i=1}^{k}\left|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}}.
\end{aligned} \tag{3.155}$$

Combining this with (3.154) and the fact that for all $x \in \mathbb{R}^d$, $t \in [0,T)$, $s \in [t,T)$, $r \in [s,T)$ it holds that $Z^{0,x,n}_{t,s}$ and $Y^{0,x,n}_{s,r}$ are independent demonstrates that for all $l_1 \in \mathbb{N}$, $t_0, t_1, \ldots, t_k \in [0,T)$, $x \in \mathbb{R}^d$, $\nu_0, \nu_1, \ldots, \nu_k \in \{1, 2, \ldots, d+1\}$ with $t_0 < t_1 < \ldots < t_k < T$

it holds that

$$
\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,l_1,M,Q}(t_k,Y^{0,x,n}_{t_0,t_k})-v^\infty(t_k,Y^{0,x,n}_{t_0,t_k}))_{\nu_k}\right|^2\prod_{i=1}^{k}\left|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}}\\
&\leqslant\left(\mathbb{E}\left[\left|(\varepsilon(t_k,Y^{0,x,n}_{t_0,t_k}))_{\nu_k}\prod_{i=1}^{k}(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}}\\
&\quad+\left(\mathbb{E}\left[\left|\frac{g(Y^{0,x,n}_{t_0,T})-g(Y^{0,x,n}_{t_0,t_k})}{\sqrt{M^{l_1}}}(Z^{0,x,n}_{t_k,T})_{\nu_k}\prod_{i=1}^{k}(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}}\\
&\quad+\sum_{t_{k+1}\in(t_k,T)}\left(\mathbb{E}\left[\left|\frac{q^{Q,[t_k,T]}(t_{k+1})(F(0))(t_{k+1},Y^{0,x,n}_{t_0,t_{k+1}})}{\sqrt{M^{l_1}}}\prod_{i=1}^{k+1}(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}}\\
&\quad+\sum_{t_{k+1}\in(t_k,T)}\sum_{\nu_{k+1}=1}^{d+1}\frac{L_{\nu_{k+1}}q^{Q,[t_k,T]}(t_{k+1})}{\sqrt{M^{l_1-1}}}\\
&\quad\cdot\left(\mathbb{E}\left[\left|(v^\infty(t_{k+1},Y^{0,x,n}_{t_0,t_{k+1}}))_{\nu_{k+1}}\prod_{i=1}^{k+1}(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}}\\
&\quad+\Bigg(2K\sqrt{\tfrac{1}{\alpha}\left(\tfrac{(T-t_0)}{n^2}+\tfrac{1}{n}\right)}\Big(1+3K(T-t_0)\exp\big(5K(T-t_0)\big)\Big)\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right]\\
&\quad+4K\sqrt{\tfrac{1}{\alpha}\left(\tfrac{(T-t_0)^3}{n^2}+\tfrac{(T-t_0)^2}{n}\right)}\Big(1+3K(T-t_0)\exp\big(5K(T-t_0)\big)\Big)\\
&\quad\cdot\left[\sup_{s\in[t_k,T]}\sup_{y\in\mathbb{R}^d}\left(|(F(0))(s,y)|+\mathcal{L}|\!|\!|v^\infty(s,y)|\!|\!|\right)\right]\Bigg)\left(\mathbb{E}\left[\left|\prod_{i=1}^{k}(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}}\\
&\quad+\sum_{l_0=1}^{l_1-1}\sum_{t_{k+1}\in(t_k,T)}\sum_{\nu_{k+1}=1}^{d+1}\frac{2L_{\nu_{k+1}}q^{Q,[t_k,T]}(t_{k+1})}{\sqrt{M^{l_1-1-l_0}}}\\
&\quad\cdot\left(\mathbb{E}\left[\left|((V^0_{n,l_0,M,Q}-v^\infty)(t_{k+1},Y^{0,x,n}_{t_0,t_{k+1}}))_{\nu_{k+1}}\prod_{i=1}^{k+1}(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}}.
\end{aligned}
\tag{3.156}
$$

This and the induction hypothesis imply (3.144) for $k+1$. Induction therefore proves (3.144). The proof of Lemma 3.2.3 is thus complete. □

In the following theorem we apply the findings from Lemma 3.2.3 to the case $k \curvearrowleft m$ to achieve an error bound that is not recursive. Theorem 3.2.4 generalizes the results of [59, Theorem 4.4] to a more general setting of semilinear PDEs with gradient-dependent nonlinearities.

Theorem 3.2.4. *Assume Setting 3.1.1, let $n,m,Q\in\mathbb{N}$, $M\in\mathbb{N}\cap[2,\infty)$, $\nu_0\in\{1,2,\ldots,d+1\}$, $t_0\in[0,T)$, $x\in\mathbb{R}^d$, assume that*

$$
\sup_{x\in\mathbb{R}^d}\frac{|g(x)|}{\varphi(T,x)}+\sup_{t\in[0,T]}\sup_{x\in\mathbb{R}^d}\frac{|(F(0))(t,x)|}{\varphi(t,x)}+\sup_{t\in[0,T]}\sup_{x\in\mathbb{R}^d}\frac{|\!|\!|v^\infty(t,x)|\!|\!|}{\varphi(t,x)}<\infty,
\tag{3.157}
$$

let $C \in [1,\infty)$ *satisfy that*

$$\begin{aligned} C = \max\Big\{ &\sqrt{\pi T}, \mathcal{L}, \big(\sqrt{T} + (1 + \tfrac{1}{\sqrt{\alpha}}\sqrt{3KT}\exp\big(\tfrac{5}{2}KT\big))\big), \\ &K\sqrt{\tfrac{1}{\alpha}(1+3KT\exp(5KT))}, \big(2K\left(\gamma_x^{(4)}T\exp\left(6+32K^4T\right)\right)^{\frac{1}{4}} + \gamma_x^{(1)}\big) \\ &\cdot \Big[\sqrt{T} + \sqrt{\tfrac{6}{\alpha}}(1+\sqrt{6K(1+T)}\exp(40K^2(1+T)^2))\Big]\Big\}, \end{aligned} \tag{3.158}$$

and let $\varepsilon \colon [0,T]\times\mathbb{R}^d \to [0,\infty]^{d+1}$ *satisfy for all* $r \in [0,T]$, $y \in \mathbb{R}^d$ *that*

$$\begin{aligned} &\varepsilon(t,y) \\ &= \mathbb{E}\Bigg[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)(F(v^\infty))(s,Y^{0,y,n}_{t,s})\begin{pmatrix} 1 \\ \frac{1}{s-t}\int_t^s (\sigma(Y^{0,y,n}_{t,\max\{t,\lfloor r\rfloor_t\}}))^{-1}D^{0,y,n}_{t,r}\,\mathrm{d}W^0_r \end{pmatrix} \\ &\quad - \int_t^T (F(v^\infty))(r,Y^{0,y,n}_{t,r})\begin{pmatrix} 1 \\ \frac{1}{r-t}\int_t^r (\sigma(Y^{0,y,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1}D^{0,y,n}_{t,s}\,\mathrm{d}W^0_s \end{pmatrix}\mathrm{d}r\Bigg]. \end{aligned} \tag{3.159}$$

Then it holds that

$$\begin{aligned} &\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t_0,x) - v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac{1}{2}} \\ &\leqslant \left(1+14(8C^3)^{m-1}\right)\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|||\varepsilon(s,y)|||\right] + \left(1+6(4C^3)^{m-1}e^M\right)\frac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^{m-3}}} \\ &\quad + \left(1+14C(8C^3)^{m-1}\right)2C\sqrt{\left(\frac{T-t_0}{n^2}+\frac{1}{n}\right)}\Bigg(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right] \\ &\quad + 2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}\left(|(F(0))(s,y)| + C|||v^\infty(s,y)|||\right)\right]\Bigg) \\ &\quad + \frac{2C^2+(4C^3)^m e^M}{\sqrt{M^{m-3}}}\Bigg(\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|\right] \\ &\quad + \left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|||v^\infty(s,y)|||\right]\Bigg). \end{aligned} \tag{3.160}$$

Proof of Theorem 3.2.4. First note that Lemma 3.2.3 (applied with $k \curvearrowleft m$ in the notation of Lemma 3.2.3) and (3.158) demonstrate that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t_0,x)-v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac{1}{2}}\\
&\leqslant \sum_{j=0}^{m-1}\sum_{\substack{l_1,\ldots,l_{j+1}\in\mathbb{N}\\ l_1<\ldots<l_{j+1}=m}}\sum_{\substack{t_1,\ldots,t_{j+1}\in\mathbb{R}\\ t_0<t_1<\ldots<t_{j+1}\leqslant T}}\sum_{\nu_1,\ldots,\nu_{j+1}\in\{1,\ldots,d+1\}}\frac{2^j}{\sqrt{M^{m-j-l_1}}}\left[\prod_{i=1}^{j}L_{\nu_i}q^{Q,[t_{i-1},T]}(t_i)\right]\\
&\quad\cdot\Bigg\{\mathbb{1}_1(\nu_{j+1})\Bigg(\mathbb{1}_T(t_{j+1})\Bigg(\Bigg(\mathbb{E}\Bigg[\Bigg|(\varepsilon(t_j,Y^{0,x,n}_{t_0,t_j}))_{\nu_j}\prod_{i=1}^{j}(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\Bigg|^2\Bigg]\Bigg)^{\frac{1}{2}}\\
&\quad+\frac{1}{\sqrt{M^{l_1}}}\Bigg(\mathbb{E}\Bigg[\Bigg|(g(Y^{0,x,n}_{t_0,T})-g(Y^{0,x,n}_{t_0,t_j}))\prod_{i=1}^{j+1}(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\Bigg|^2\Bigg]\Bigg)^{\frac{1}{2}}\\
&\quad+2C\sqrt{\left(\frac{T-t_0}{n^2}+\frac{1}{n}\right)}\Bigg(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right]\\
&\quad+2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}(|(F(0))(s,y)|+\mathcal{L}|||v^\infty(s,y)|||)\right]\Bigg)\Bigg(\mathbb{E}\Bigg[\prod_{i=1}^{j}|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}|^2\Bigg]\Bigg)\Bigg)\\
&\quad+\frac{q^{Q,[t_j,T]}(t_{j+1})}{\sqrt{M^{l_1}}}\Bigg(\mathbb{E}\Bigg[\Bigg|(F(0))(t_{j+1},Y^{0,x,n}_{t_0,t_{j+1}})\prod_{i=1}^{j+1}(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\Bigg|^2\Bigg]\Bigg)^{\frac{1}{2}}\Bigg)\\
&\quad+\frac{L_{\nu_{j+1}}q^{Q,[t_j,T]}(t_{j+1})}{\sqrt{M^{l_1-1}}}\Bigg(\mathbb{E}\Bigg[\Bigg|(v^\infty(t_{j+1},Y^{0,x,n}_{t_0,t_{j+1}}))_{\nu_{j+1}}\prod_{i=1}^{j+1}(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\Bigg|^2\Bigg]\Bigg)^{\frac{1}{2}}\Bigg\}.
\end{aligned}\tag{3.161}$$

Item (v) of Lemma 3.1.4 hence shows that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t_0,x)-v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac{1}{2}}\\
&\leqslant \sum_{j=0}^{m-1}\sum_{\substack{l_1,\ldots,l_{j+1}\in\mathbb{N}\\ l_1<\ldots<l_{j+1}=m}}\sum_{\substack{t_1,\ldots,t_{j+1}\in\mathbb{R}\\ t_0<t_1<\ldots<t_{j+1}\leqslant T}}\sum_{\nu_1,\ldots,\nu_{j+1}\in\{1,\ldots,d+1\}}\frac{2^j}{\sqrt{M^{m-j-l_1}}}\left[\prod_{i=1}^{j}L_{\nu_i}q^{Q,[t_{i-1},T]}(t_i)\right]\\
&\quad\cdot\Bigg\{\mathbb{1}_1(\nu_{j+1})\Bigg(\mathbb{1}_T(t_{j+1})\Bigg(\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}(\varepsilon(s,y))_{\nu_j}\right]\prod_{i=1}^{j}\left(\mathbb{E}\left[\left|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}}\\
&\quad+\frac{(\mathbb{E}[|(g(Y^{0,x,n}_{t_0,T})-g(Y^{0,x,n}_{t_0,t_j}))(Z^{0,x,n}_{t_j,T})_{\nu_{i-1}}|^2])^{\frac{1}{2}}}{\sqrt{M^{l_1}}}\prod_{i=1}^{j}\left(\mathbb{E}\left[\left|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}}\\
&\quad+2C\sqrt{\left(\frac{T-t_0}{n^2}+\frac{1}{n}\right)}\Bigg(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right]\\
&\quad+2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}(|(F(0))(s,y)|+\mathcal{L}|||v^\infty(s,y)|||)\right]\Bigg)\prod_{i=1}^{j}\left(\mathbb{E}\left[\left|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}}\Bigg)\\
&\quad+\frac{q^{Q,[t_j,T]}(t_{j+1})}{\sqrt{M^{l_1}}}\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|\right]\prod_{i=1}^{j+1}\left(\mathbb{E}\left[\left|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}}\Bigg)\\
&\quad+\frac{L_{\nu_{j+1}}q^{Q,[t_j,T]}(t_{j+1})}{\sqrt{M^{l_1-1}}}\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}(v^\infty(s,y))_{\nu_{j+1}}\right]\prod_{i=1}^{j+1}\left(\mathbb{E}\left[\left|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\right|^2\right]\right)^{\frac{1}{2}}\Bigg\}.
\end{aligned}\tag{3.162}$$

Furthermore, note that the Cauchy-Schwarz inequality, item (ii) of Lemma 3.1.7, item (ii) of Lemma 3.1.8, (3.10), and (3.158) ensure that for all $t \in [0, T)$, $\nu \in \{1, 2, \ldots, d+1\}$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|(g(Y_{t,T}^{0,x,n}) - g(x))(Z_{t,T}^{0,x,n})_\nu\right|^2\right]\right)^{\frac{1}{2}} \leqslant \mathcal{L}\left(\mathbb{E}\left[\|Y_{t,T}^{0,x,n} - x\|^2 \|\!|Z_{t,T}^{0,x,n}|\!\|^2\right]\right)^{\frac{1}{2}} \\
&\leqslant \mathcal{L}\Big(\mathbb{E}\big[\|Y_{t,T}^{0,x,n} - x\|^4\big]\Big)^{\frac{1}{4}} \left(\mathbb{E}\left[\|\!|Z_{t,T}^{0,x,n}|\!\|^4\right]\right)^{\frac{1}{4}} \\
&\leqslant \mathcal{L}\sqrt{T-t}(\sqrt{T} + \sqrt{6})\left(2K\left(\gamma_x^{(4)}(T-t)\exp\left(6 + 32K^4 T\right)\right)^{\frac{1}{4}} + \gamma_x^{(1)}\right) \\
&\qquad \cdot \left[1 + \sqrt{\frac{6}{\alpha(T-t)}}\Big(1 + \sqrt{6K(1+T-t)}\exp\left(40K^2(1+T-t)^2\right)\Big)\right] \\
&= \mathcal{L}(\sqrt{T} + \sqrt{6})\left(2K\left(\gamma_x^{(4)} T\exp\left(6 + 32K^4 T\right)\right)^{\frac{1}{4}} + \gamma_x^{(1)}\right) \\
&\qquad \cdot \left[\sqrt{T} + \sqrt{\frac{6}{\alpha}}\Big(1 + \sqrt{6K(1+T)}\exp\left(40K^2(1+T)^2\right)\Big)\right] \\
&\leqslant C^2(\sqrt{T} + \sqrt{6}).
\end{aligned} \tag{3.163}$$

Combining this with (3.158) and (3.162) proves that

$$
\begin{aligned}
&\Big(\mathbb{E}\Big[\big|(V^0_{n,m,M,Q}(t_0,x)-v^\infty(t_0,x))_{\nu_0}\big|^2\Big]\Big)^{\frac12}\\
&\leqslant \bigg[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}(\varepsilon(s,y))_{\nu_0}\bigg]+\tfrac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^m}}+2C\sqrt{\left(\tfrac{T-t_0}{n^2}+\tfrac1n\right)}\bigg(\Big[\sup_{y\in\mathbb{R}^d}|g(y)|\Big]\\
&\quad+2T\bigg[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}\big(|(F(0))(s,y)|+\mathcal{L}|||v^\infty(s,y)|||\big)\bigg]\bigg)\\
&\quad+\tfrac{[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|]}{\sqrt{M^m}}\sum_{t_1\in(t_0,T]}q^{Q,[t_0,T]}(t_1)\left(\mathbb{E}\left[|(Z^{0,x,n}_{t_0,t_1})_{\nu_0}|^2\right]\right)^{\frac12}\\
&\quad+\tfrac{\sum_{\nu_1=1}^{d+1}L_{\nu_1}[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}(v^\infty(s,y))_{\nu_1}]}{\sqrt{M^{m-1}}}\sum_{t_1\in(t_0,T]}q^{Q,[t_0,T]}(t_1)\Big(\mathbb{E}\Big[\big|(Z^{0,x,n}_{t_0,t_1})_{\nu_0}\big|^2\Big]\Big)^{\frac12}\\
&\quad+\sum_{j=1}^{m-1}\sum_{\substack{l_1,\ldots,l_j\in\mathbb{N}\\ l_1<\ldots<l_j<m}}\sum_{\substack{t_1,\ldots,t_{j+1}\in\mathbb{R}\\ t_0<t_1<\ldots<t_{j+1}\leqslant T}}\sum_{\nu_1,\ldots,\nu_{j+1}\in\{1,\ldots,d+1\}}\tfrac{2^j}{\sqrt{M^{m-j-l_1}}}\\
&\quad\cdot\bigg[\prod_{i=1}^{j}L_{\nu_i}q^{Q,[t_{i-1},T]}(t_i)\bigg]\\
&\quad\cdot\bigg\{\mathbb{1}_1(\nu_{j+1})\bigg(\mathbb{1}_T(t_{j+1})\bigg(\bigg[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}(\varepsilon(s,y))_{\nu_j}\bigg]\prod_{i=1}^{j}\Big(\mathbb{E}\Big[\big|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\big|^2\Big]\Big)^{\frac12}\\
&\quad+\tfrac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^{l_1}}}\prod_{i=1}^{j}\Big(\mathbb{E}\Big[\big|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\big|^2\Big]\Big)^{\frac12}\\
&\quad+2C\sqrt{\left(\tfrac{T-t_0}{n^2}+\tfrac1n\right)}\bigg(\Big[\sup_{y\in\mathbb{R}^d}|g(y)|\Big]\\
&\quad+2T\bigg[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}\big(|(F(0))(s,y)|+\mathcal{L}|||v^\infty(s,y)|||\big)\bigg]\bigg)\prod_{i=1}^{j}\Big(\mathbb{E}\Big[\big|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\big|^2\Big]\Big)^{\frac12}\bigg)\\
&\quad+\tfrac{q^{Q,[t_j,T]}(t_{j+1})[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|]}{\sqrt{M^{l_1}}}\prod_{i=1}^{j+1}\left(\mathbb{E}\left[|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}|^2\right]\right)^{\frac12}\bigg)\\
&\quad+\tfrac{L_{\nu_{j+1}}q^{Q,[t_j,T]}(t_{j+1})[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(v^\infty(s,y))_{\nu_{j+1}}|]}{\sqrt{M^{l_1-1}}}\prod_{i=1}^{j+1}\Big(\mathbb{E}\Big[\big|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\big|^2\Big]\Big)^{\frac12}\bigg\}.
\end{aligned}
\tag{3.164}
$$

Next note that item (i) of Lemma 3.1.8 and (3.158) ensure that for all $j\in\mathbb{N}$, $t_1,t_2,\ldots,t_j\in[0,T)$ with $t_0<t_1<\ldots<t_j<T$ it holds that

$$
\begin{aligned}
&\prod_{i=1}^{j}\Big(\mathbb{E}\Big[\big|(Z^{0,x,n}_{t_{i-1},t_i})_{\nu_{i-1}}\big|^2\Big]\Big)^{\frac12}\leqslant\prod_{i=1}^{j}\left(1+\frac{1}{\sqrt{\alpha(t_i-t_{i-1})}}\left(1+\sqrt{3KT}\exp\left(\tfrac52KT\right)\right)\right)\\
&\leqslant\left(\sqrt{T-t_0}+\left(1+\frac{1}{\sqrt{\alpha}}\sqrt{3KT}\exp\left(\tfrac52KT\right)\right)\right)^j\prod_{i=1}^{j}\frac{1}{\sqrt{t_i-t_{i-1}}}\\
&\leqslant C^j\prod_{i=1}^{j}\frac{1}{\sqrt{t_i-t_{i-1}}}.
\end{aligned}
\tag{3.165}
$$

Next note that for all $j \in \mathbb{N}$ it holds that

$$\left[\sum_{\nu_1,\ldots\nu_j\in\{1,\ldots,d+1\}} \prod_{i=1}^{j} L_{\nu_i}\right] = \left(\sum_{\nu=1}^{d+1} L_\nu\right)^j = \mathcal{L}^j. \tag{3.166}$$

This, (3.164), and (3.165) show that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t_0,x) - v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac{1}{2}} \\
&\leqslant \left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}(\varepsilon(s,y))_{\nu_0}\right] + \tfrac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^m}} + 2C\sqrt{\left(\tfrac{T-t_0}{n^2}+\frac{1}{n}\right)}\Bigg(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right] \\
&\quad + 2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)| + \mathcal{L}|\!|\!|v^\infty(s,y)|\!|\!|\right]\Bigg) \\
&\quad + \tfrac{C[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|]}{\sqrt{M^m}} \sum_{t_1\in(t_0,T]} \tfrac{q^{Q,[t_0,T]}(t_1)}{\sqrt{t_1-t_0}} \\
&\quad + \tfrac{C\sum_{\nu_1=1}^{d+1} L_{\nu_1}[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(v^\infty(s,y))_{\nu_1}|]}{\sqrt{M^{m-1}}} \sum_{t_1\in(t_0,T]} \tfrac{q^{Q,[t_0,T]}(t_1)}{\sqrt{t_1-t_0}} \\
&\quad + \sum_{j=1}^{m-1} \sum_{\substack{l_1,\ldots,l_j\in\mathbb{N}\\ l_1<\ldots<l_j<m}} \Bigg\{ \tfrac{(2C\mathcal{L})^j[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|\!|\!|\varepsilon(s,y)|\!|\!|]}{\sqrt{M^{m-j-l_1}}} \\
&\quad \cdot \left[\sum_{\substack{t_1,\ldots,t_j\in\mathbb{R}\\ t<t_1<\ldots<t_j<T}} \prod_{i=1}^{j} \tfrac{q^{Q,[t_{i-1},T]}(t_i)}{\sqrt{t_i-t_{i-1}}}\right] \\
&\quad + \tfrac{(2C\mathcal{L})^j C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^{m-j}}}\left[\sum_{\substack{t_1,\ldots,t_j\in\mathbb{R}\\ t<t_1<\ldots<t_j<T}} \prod_{i=1}^{j} \tfrac{q^{Q,[t_{i-1},T]}(t_i)}{\sqrt{t_i-t_{i-1}}}\right] \\
&\quad + \tfrac{(2C\mathcal{L})^{j+1}}{\sqrt{M^{m-j-l_1}}}\left[\sum_{\substack{t_1,\ldots,t_j\in\mathbb{R}\\ t<t_1<\ldots<t_j<T}} \prod_{i=1}^{j} \tfrac{q^{Q,[t_{i-1},T]}(t_i)}{\sqrt{t_i-t_{i-1}}}\right]\sqrt{\left(\tfrac{T-t_0}{n^2}+\tfrac{1}{n}\right)}\Bigg(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right] \\
&\quad + 2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}\left(|(F(0))(s,y)| + \mathcal{L}|\!|\!|v^\infty(s,y)|\!|\!|\right)\right]\Bigg) \\
&\quad + \tfrac{(2C\mathcal{L})^j[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|]}{\sqrt{M^{m-j}}}\left[\sum_{\substack{t_1,\ldots,t_{j+1}\in\mathbb{R}\\ t<t_1<\ldots<t_{j+1}\leqslant T}} \prod_{i=1}^{j+1} \tfrac{q^{Q,[t_{i-1},T]}(t_i)}{\sqrt{t_i-t_{i-1}}}\right] \\
&\quad + \tfrac{(2C)^j\mathcal{L}^{j+1}[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|\!|\!|v^\infty(s,y)|\!|\!|]}{\sqrt{M^{m-j-1}}}\left[\sum_{\substack{t_1,\ldots,t_{j+1}\in\mathbb{R}\\ t<t_1<\ldots<t_{j+1}\leqslant T}} \prod_{i=1}^{j+1} \tfrac{q^{Q,[t_{i-1},T]}(t_i)}{\sqrt{t_i-t_{i-1}}}\right]\Bigg\}.
\end{aligned} \tag{3.167}$$

Combining this with [55, Lemma 3.3], the assumption that $\sum_{\nu=1}^{d+1} L_\nu = \mathcal{L}$, and the fact that $\Gamma(\frac{1}{2})=\sqrt{\pi}$ guarantees that

$$
\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t_0,x)-v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac{1}{2}}\\
&\leqslant\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}(\varepsilon(s,y))_{\nu_0}\right]+\frac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^m}}+2C\sqrt{\left(\frac{T-t_0}{n^2}+\frac{1}{n}\right)}\Bigg(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right]\\
&\quad+2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}\left(|(F(0))(s,y)|+\mathcal{L}|\!|\!|v^\infty(s,y)|\!|\!|\right)\right]\Bigg)\\
&\quad+\frac{2C\sqrt{T-t_0}}{\sqrt{M^{m-1}}}\Bigg(\frac{[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|]}{\sqrt{M}}\\
&\quad+\mathcal{L}\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|\!|\!|v^\infty(s,y)|\!|\!|\right]\Bigg)\\
&\quad+2\sum_{j=1}^{m-1}\sum_{\substack{l_1,\ldots,l_j\in\mathbb{N}\\ l_1<\ldots<l_j<m}}\Bigg\{\frac{(2C\mathcal{L}\sqrt{\pi(T-t_0)})^j[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|\!|\!|\varepsilon(s,y)|\!|\!|]}{\sqrt{M^{m-j-l_1}}\Gamma(\frac{j}{2})}\\
&\quad+\frac{(2C\mathcal{L}\sqrt{\pi(T-t_0)})^jC^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^{m-j}}\Gamma(\frac{j}{2})}\\
&\quad+\frac{(2C\mathcal{L})^{j+1}(\sqrt{\pi(T-t_0)})^j}{\sqrt{M^{m-j-l_1}}\Gamma(\frac{j}{2})}\sqrt{\left(\frac{T-t_0}{n^2}+\frac{1}{n}\right)}\Bigg(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right]\\
&\quad+2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}\left(|(F(0))(s,y)|+\mathcal{L}|\!|\!|v^\infty(s,y)|\!|\!|\right)\right]\Bigg)\\
&\quad+\frac{(2C\mathcal{L})^j(\sqrt{\pi(T-t_0)})^{j+1}[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|]}{\sqrt{M^{m-j}}\Gamma(\frac{j+1}{2})}\\
&\quad+\frac{(2C)^j(\mathcal{L}\sqrt{\pi(T-t_0)})^{j+1}[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|\!|\!|v^\infty(s,y)|\!|\!|]}{\sqrt{M^{m-j-1}}\Gamma(\frac{j+1}{2})}\Bigg\}.
\end{aligned}
\tag{3.168}
$$

This, [55, Lemma 3.4], and (3.158) demonstrate that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t_0,x)-v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac{1}{2}}\\
&\leqslant\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}(\varepsilon(s,y))_{\nu_0}\right]+\frac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^m}}+2C\sqrt{\left(\frac{T-t_0}{n^2}+\frac{1}{n}\right)}\Bigg(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right]\\
&\quad+2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}(|(F(0))(s,y)|+C|||v^\infty(s,y)|||)\right]\Bigg)\\
&\quad+\frac{2C\sqrt{T-t_0}}{\sqrt{M^{m-1}}}\Bigg(\frac{[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|]}{\sqrt{M}}\\
&\quad+C\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|||v^\infty(s,y)|||\right]\Bigg)\\
&\quad+\frac{2[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|||\varepsilon(s,y)|||]}{\sqrt{M^m}}\sum_{j=1}^{m-1}\frac{(2C^3\sqrt{M})^j}{\Gamma(\frac{j}{2})}\sum_{l=1}^{m-j}\sqrt{M}^l\binom{m-l-1}{j-1}\\
&\quad+\frac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^m}}\sum_{j=1}^{m-1}\frac{(2C^3\sqrt{M})^j}{\Gamma(\frac{j}{2})}\binom{m-1}{j}\\
&\quad+\frac{4C^2}{\sqrt{M^m}}\sqrt{\left(\frac{T-t_0}{n^2}+\frac{1}{n}\right)}\Bigg(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right]\\
&\quad+2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}(|(F(0))(s,y)|+C|||v^\infty(s,y)|||)\right]\Bigg)\\
&\quad\cdot\sum_{j=1}^{m-1}\frac{(2C^3\sqrt{M})^j}{\Gamma(\frac{j}{2})}\sum_{l=1}^{m-j}\sqrt{M}^l\binom{m-l-1}{j-1}\\
&\quad+\frac{[2C\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|]}{\sqrt{M^m}}\sum_{j=1}^{m-1}\frac{(2C^3\sqrt{M})^j}{\Gamma(\frac{j+1}{2})}\binom{m-1}{j}\\
&\quad+\frac{[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|||v^\infty(s,y)|||]}{\sqrt{M^m}}\sum_{j=1}^{m-1}\frac{(2C^3\sqrt{M})^{j+1}}{\Gamma(\frac{j+1}{2})}\binom{m-1}{j}.
\end{aligned}\tag{3.169}$$

Combining this with (3.158) ensures that

$$
\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t_0,x)-v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac{1}{2}}\\
&\leqslant \left(1+\frac{2}{\sqrt{M^m}}\sum_{j=1}^{m-1}\frac{(2C^3\sqrt{M})^j}{\Gamma(\frac{j}{2})}\sum_{l=1}^{m-j}\sqrt{M}^l\binom{m-l-1}{j-1}\right)\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|||\varepsilon(s,y)|||\right]\\
&\quad+\left(1+2\sum_{j=1}^{m-1}\frac{(2C^3\sqrt{M})^j}{\Gamma(\frac{j}{2})}\binom{m-1}{j}\right)\frac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^m}}\\
&\quad+\left(1+\frac{2C}{\sqrt{M^m}}\sum_{j=1}^{m-1}\frac{(2C^3\sqrt{M})^j}{\Gamma(\frac{j}{2})}\sum_{l=1}^{m-j}\sqrt{M}^l\binom{m-l-1}{j-1}\right)2C\sqrt{\left(\frac{T-t_0}{n^2}+\frac{1}{n}\right)}\\
&\quad\cdot\left(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right]+2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}\left(|(F(0))(s,y)|+C|||v^\infty(s,y)|||\right)\right]\right)\\
&\quad+2C\left(C+\sum_{j=1}^{m-1}\frac{(2C^3\sqrt{M})^j}{\Gamma(\frac{j+1}{2})}\binom{m-1}{j}\right)\frac{[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|]}{\sqrt{M^m}}\\
&\quad+2C^3\left(1+\sum_{j=1}^{m-1}\frac{(2C^3\sqrt{M})^j}{\Gamma(\frac{j+1}{2})}\binom{m-1}{j}\right)\frac{[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|||v^\infty(s,y)|||]}{\sqrt{M^{m-1}}}.
\end{aligned}
\tag{3.170}
$$

Throughout the rest of the proof let $\lfloor\cdot\rfloor\colon[0,\infty)\to[0,\infty)$ satisfy for all $a\in[0,\infty)$ that $\lfloor a\rfloor=\sup([0,a]\cap\mathbb{N}_0)$. Furthermore, observe that for all $z\in[0,\infty)$ it holds that

$$
\begin{aligned}
\sum_{j=1}^{m-1}\frac{z^j}{\Gamma(\frac{j+1}{2})}&\leqslant\sum_{j=1}^{m-1}\frac{z^j}{\Gamma(\lfloor\frac{j+1}{2}\rfloor)}=\sum_{l=1}^{\lfloor\frac{m}{2}\rfloor}\frac{z^{2l-1}}{\Gamma(l)}+\sum_{l=1}^{\lfloor\frac{m-1}{2}\rfloor}\frac{z^{2l}}{\Gamma(l)}\\
&=\sum_{l=0}^{\lfloor\frac{m}{2}\rfloor-1}\frac{z^{2l+1}}{l!}+\sum_{l=0}^{\lfloor\frac{m-1}{2}\rfloor-1}\frac{z^{2l+2}}{l!}\leqslant z(z+1)e^{z^2}.
\end{aligned}
\tag{3.171}
$$

Combining this with the fact that for all $j\in\{1,\dots,m-1\}$ it holds that $\binom{m-1}{j}\leqslant\sum_{k=0}^{m-1}\binom{m-1}{k}=2^{m-1}$ and with the fact that $\Gamma(\frac{1}{2})=\sqrt{\pi}$ ensures that for every $z\in[0,\infty)$ it holds that

$$
\begin{aligned}
\sum_{j=1}^{m-1}\frac{(z\sqrt{M})^j}{\Gamma(\frac{j+1}{2})}\binom{m-1}{j}&\leqslant(2z)^{m-1}\sum_{j=1}^{m-1}\frac{\sqrt{M}^j}{\Gamma(\frac{j+1}{2})}\\
&\leqslant(2z)^{m-1}\sqrt{M}(\sqrt{M}+1)e^M\leqslant 2(2z)^{m-1}Me^M
\end{aligned}
\tag{3.172}
$$

and

$$
\begin{aligned}
&\sum_{j=1}^{m-1}\frac{(z\sqrt{M})^j}{\Gamma(\frac{j}{2})}\binom{m-1}{j}\leqslant(2z)^{m-1}\sqrt{M}\left(\frac{1}{\sqrt{\pi}}+\sum_{l=1}^{m-1}\frac{\sqrt{M}^l}{\Gamma(\frac{l+1}{2})}\right)\\
&\leqslant(2z)^{m-1}\sqrt{M}\left(\frac{1}{\sqrt{\pi}}+\sqrt{M}(\sqrt{M}+1)e^M\right)\leqslant 3(2z)^{m-1}\sqrt{M^3}e^M.
\end{aligned}
\tag{3.173}
$$

Next note that for all $j\in\{1,2,\dots,m-1\}$ it holds that

$$
\begin{aligned}
&\sum_{l=1}^{m-j}\sqrt{M}^l\binom{m-l-1}{j-1}=\sum_{l=j-1}^{m-2}\sqrt{M^{m-l-1}}\binom{l}{j-1}\leqslant\sqrt{M^{m-1}}\sum_{l=j-1}^{\infty}\frac{1}{\sqrt{M^l}}\binom{l}{j-1}\\
&=\frac{\sqrt{M^{m-1}}(\frac{1}{\sqrt{M}})^{j-1}}{(1-\frac{1}{\sqrt{M}})^j}=\frac{\sqrt{M^{m-j}}}{(1-\frac{1}{\sqrt{M}})^j}.
\end{aligned}
\tag{3.174}
$$

Combining this with (3.171) implies that for all $z \in [0, \infty)$ it holds that

$$\begin{aligned}
&\sum_{j=1}^{m-1} \frac{(z\sqrt{M})^j}{\Gamma(\frac{j}{2})} \sum_{l=1}^{m-j} \sqrt{M^l} \binom{m-l-1}{j-1} \leqslant \sqrt{M^m} \sum_{j=1}^{m-1} \frac{z^j}{\Gamma(\frac{j}{2})(1-\frac{1}{\sqrt{M}})^j} \\
&\leqslant \sqrt{M^m} \frac{z^{m-1}}{(1-\frac{1}{\sqrt{2}})^{m-1}} \sum_{j=1}^{m-1} \frac{1}{\Gamma(\frac{j}{2})} = \sqrt{M^m} \frac{z^{m-1}}{(1-\frac{1}{\sqrt{2}})^{m-1}} \left(\frac{1}{\sqrt{\pi}} + \sum_{i=1}^{m-2} \frac{1}{\Gamma(\frac{i+1}{2})} \right) \\
&\leqslant (4z)^{m-1}\sqrt{M^m} \left(\frac{1}{\sqrt{\pi}} + 2e \right) \leqslant 7(4z)^{m-1}\sqrt{M^m}.
\end{aligned} \tag{3.175}$$

This, (3.169), (3.171), (3.172), and (3.174) imply that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t_0,x) - v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac{1}{2}} \\
&\leqslant \left(1 + 14(8C^3)^{m-1}\right) \left[\sup_{s\in[t_0,T]} \sup_{y\in\mathbb{R}^d} |||\varepsilon(s,y)|||\right] + \left(1 + 6(4C^3)^{m-1}e^M\right) \frac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^{m-3}}} \\
&\quad + \left(1 + 14C(8C^3)^{m-1}\right) 2C\sqrt{\left(\frac{T-t_0}{n^2} + \frac{1}{n}\right)} \left(\left[\sup_{y\in\mathbb{R}^d} |g(y)|\right] \right. \\
&\quad \left. + 2T\left[\sup_{s\in[t_0,T]} \sup_{y\in\mathbb{R}^d} \left(|(F(0))(s,y)| + C|||v^\infty(s,y)|||\right)\right]\right) \\
&\quad + \left(2C^2 + (4C^3)^m e^M\right) \frac{[\sup_{s\in[t_0,T]} \sup_{y\in\mathbb{R}^d} |(F(0))(s,y)|]}{\sqrt{M^{m-3}}} \\
&\quad + \left(2C^3 + (4C^3)^m e^M\right) \frac{[\sup_{s\in[t_0,T]} \sup_{y\in\mathbb{R}^d} |||v^\infty(s,y)|||]}{\sqrt{M^{m-3}}}.
\end{aligned} \tag{3.176}$$

The proof of Theorem 3.2.4 is thus complete. □

The upper bound for the approximation error given in Theorem 3.2.4 still contains the unknown error term ε arising from the GauSS-Legendre quadrature rule which we intend to determine now. The following lemma therefore calculates this error term under the additional assumption that the solution is smooth.

Lemma 3.2.5. *Assume Setting 3.1.1, let $n, Q \in \mathbb{N}$, $t \in [0,T)$, $x \in \mathbb{R}^d$, assume that $u^\infty \in C^\infty([0,T] \times \mathbb{R}^d, \mathbb{R})$, and assume for all $k \in \mathbb{N}_0$ that*

$$\sup_{s\in[0,T]} \sup_{y\in\mathbb{R}^d} \frac{((\frac{\partial u^\infty}{\partial t}) + \langle \mu, (\nabla_x u^\infty)\rangle + \frac{1}{2}\operatorname{Tr}(\sigma\sigma^*(\operatorname{Hess}_x u^\infty)))^k(s,y)}{\varphi(s,y)} < \infty. \tag{3.177}$$

Then there exists $\xi = (\xi_1, \dots, \xi_{d+1}) \in [t,T]^{d+1}$ which satisfies for all $\nu \in \{1, 2, \dots, d+1\}$ that

$$\begin{aligned}
&\mathbb{E}\Bigg[\sum_{s\in(t,T)} q^{Q,[t,T]}(s)(F(v^\infty))(s, Y^{0,x,n}_{t,s}) \left(\frac{1}{\frac{1}{s-t}\int_t^s (\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}))^{-1} D^{0,x,n}_{t,r}\, dW^0_r} \right)_\nu \\
&\quad - \int_t^T (F(v^\infty))(r, Y^{0,x,n}_{t,r}) \left(\frac{1}{\frac{1}{r-t}\int_t^r (\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1} D^{0,x,n}_{t,s}\, dW^0_s} \right)_\nu dr\Bigg] \\
&= (1, \nabla_x)_\nu\, \mathbb{E}\Bigg[\Big((\tfrac{\partial u^\infty}{\partial t}) + \langle \mu, (\nabla_x u^\infty)\rangle \\
&\quad + \tfrac{1}{2}\operatorname{Tr}(\sigma\sigma^*(\operatorname{Hess}_x u^\infty))\Big)^{2Q+1}(\xi_\nu, Y^{0,x,n}_{t,\xi_\nu})\Bigg] \frac{[Q!]^4(T-t)^{2Q+1}}{(2Q+1)[(2Q)!]^3}.
\end{aligned} \tag{3.178}$$

Proof of Lemma 3.2.5. Throughout this proof let $\Psi_k\colon [0,T]\times\mathbb{R}^d\to\mathbb{R}$, $k\in\mathbb{N}_0$, satisfy for all $k\in\mathbb{N}_0$, $s\in[0,T]$, $y\in\mathbb{R}^d$ that

$$\Psi_k(s,y) = \big((\tfrac{\partial u^\infty}{\partial t}) + \langle \mu, (\nabla_x u^\infty)\rangle + \tfrac{1}{2}\operatorname{Tr}(\sigma\sigma^*(\operatorname{Hess}_x u^\infty))\big)^k(s,y). \tag{3.179}$$

First note that the dominated convergence theorem and (3.177) show that for every $k\in\mathbb{N}_0$ it holds that

$$\big([t,T]\ni s\mapsto \mathbb{E}[\Psi_k(s,Y^{0,x,n}_{t,s})]\in\mathbb{R}\big)\in C([t,T],\mathbb{R}). \tag{3.180}$$

Furthermore, observe that Itô's formula and the assumption that $u^\infty\in C^\infty([0,T)\times\mathbb{R}^d,\mathbb{R})$ ensure that for all $k\in\mathbb{N}_0$, $s\in[t,T]$ it holds a.s. that

$$\begin{aligned}
&\Psi_k(s,Y^{0,x,n}_{t,s}) - \Psi_k(t,x)\\
&= \int_t^s (\tfrac{\partial \Psi_k}{\partial t})(r,Y^{0,x,n}_{t,r})\,\mathrm{d}r + \int_t^s (\nabla_x\Psi_k)(r,Y^{0,x,n}_{t,r})\,\mathrm{d}Y^{0,x,n}_{t,r}\\
&\quad + \tfrac{1}{2}\sum_{i,j=1}^d \int_t^s (\tfrac{\partial^2\Psi_k}{\partial x_i\partial x_j})(r,Y^{0,x,n}_{t,r})\,\mathrm{d}[(Y^{0,x,n}_t)_i,(Y^{0,x,n}_t)_j]_r\\
&= \int_t^s \Big[(\tfrac{\partial \Psi_k}{\partial t})(r,Y^{0,x,n}_{t,r}) + \langle \mu(Y^{0,x,n}_{t,r}), (\nabla_x\Psi_k)(r,Y^{0,x,n}_{t,r})\rangle\\
&\quad + \tfrac{1}{2}\operatorname{Tr}(\sigma(Y^{0,x,n}_{t,r})[\sigma(Y^{0,x,n}_{t,r})]^*(\operatorname{Hess}_x\Psi_k)\Big]\,\mathrm{d}r\\
&\quad + \int_t^s (\nabla_x\Psi_k)(r,Y^{0,x,n}_{t,r})\sigma(Y^{0,x,n}_{t,r})\,\mathrm{d}W^0_r.
\end{aligned} \tag{3.181}$$

This and (3.177) imply that for all $k\in\mathbb{N}$ it holds that

$$\mathbb{E}\bigg[\sup_{s\in[t,T]}\bigg|\int_t^s (\nabla_x\Psi_k)(r,Y^{0,x,n}_{t,r})\sigma(Y^{0,x,n}_{t,r})\,\mathrm{d}W^0_r\bigg|\bigg] < \infty. \tag{3.182}$$

Hence, we obtain that for all $s\in(t,T]$ it holds that

$$\int_t^s (\nabla_x\Psi_k)(r,Y^{0,x,n}_{t,r})\sigma(Y^{0,x,n}_{t,r})\,\mathrm{d}W^0_r = 0. \tag{3.183}$$

Combining this with Fubini's theorem, (3.179), and (3.181) demonstrates that for all $k\in\mathbb{N}_0$, $s\in[t,T]$ it holds that

$$\mathbb{E}\big[\Psi_k(s,Y^{0,x,n}_{t,s})\big] - \Psi_k(t,x) = \int_t^s \mathbb{E}\big[\Psi_{k+1}(r,Y^{0,x,n}_{t,r})\big]\,\mathrm{d}r. \tag{3.184}$$

Moreover, observe that (3.184) (applied with $k=1$) and (3.180) (applied with $k=2$) ensure that

$$([t,T]\ni s\mapsto \mathbb{E}[\Psi_1(s,Y^{0,x,n}_{t,s})]\in\mathbb{R})\in C^1([t,T],\mathbb{R}). \tag{3.185}$$

Induction, (3.180) and (3.184) therefore prove that $([t,T]\ni s\mapsto\mathbb{E}[\Psi_1(s,Y^{0,x,n}_{t,s})]\in\mathbb{R})\in C^\infty([t,T],\mathbb{R})$ and that for all $k\in\mathbb{N}$, $s\in[t,T]$ it holds that

$$\frac{\partial^k}{\partial t^k}\mathbb{E}\big[\Psi_1(s,Y^{0,x,n}_{t,s})\big] = \mathbb{E}\big[\Psi_{k+1}(s,Y^{0,x,n}_{t,s})\big]. \tag{3.186}$$

Combining this with the error presentation for the GauSS-Legendre quadrature rule (see, e.g., [28, Display (2.7.12)]), (3.11), and (3.179) demonstrates that there exists $\zeta\in[t,T]$

which satisfies

$$\begin{aligned}&\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\mathbb{E}\big[(F(v^\infty))(s,Y^{0,x,n}_{t,s})\big]-\int_t^T \mathbb{E}\big[(F(v^\infty))(r,Y^{0,x,n}_{t,r})\big]\,\mathrm{d}r\\&=\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\mathbb{E}\big[\Psi_1(s,Y^{0,x,n}_{t,s})\big]-\int_t^T \mathbb{E}\big[\Psi_1(r,Y^{0,x,n}_{t,r})\big]\,\mathrm{d}r\\&=\Big(\frac{\partial^{2Q}}{\partial t^{2Q}}\mathbb{E}\big[\Psi_1(s,Y^{0,x,n}_{t,s})\big]\Big)\Big|_{s=\zeta}\frac{[Q!]^4(T-t)^{2Q+1}}{(2Q+1)[(2Q)!]^3}\\&=\mathbb{E}\big[\Psi_{2Q+1}(\zeta,Y^{0,x,n}_{t,\zeta})\big]\frac{[Q!]^4(T-t)^{2Q+1}}{(2Q+1)[(2Q)!]^3}.\end{aligned}\tag{3.187}$$

In addition, observe that the error presentation for the GauSS-Legendre quadrature rule (see, e.g., [28, Display (2.7.12)]), item (ii) of Lemma 3.2.1, (3.11), and (3.186) demonstrate that there exists $\xi=(\xi_1,\xi_2,\ldots,\xi_d)\in\mathbb{R}^d$ which satisfies for all $\nu\in\{1,2,\ldots,d\}$ that

$$\begin{aligned}&\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\mathbb{E}\bigg[(F(v^\infty))(s,Y^{0,x,n}_{t,s})\bigg(\frac{1}{s-t}\int_t^s(\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor r\rfloor_t\}}))^{-1}D^{0,x,n}_{t,r}\,\mathrm{d}W^0_r\bigg)_\nu\bigg]\\&\quad-\int_t^T\mathbb{E}\bigg[(F(v^\infty))(r,Y^{0,x,n}_{t,r})\bigg(\frac{1}{r-t}\int_t^r(\sigma(Y^{0,x,n}_{t,\max\{t,\lfloor s\rfloor_t\}}))^{-1}D^{0,x,n}_{t,s}\,\mathrm{d}W^0_s\bigg)_\nu\bigg]\,\mathrm{d}r\\&=\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\frac{\partial}{\partial x_\nu}\mathbb{E}\big[(F(v^\infty))(s,Y^{0,x,n}_{t,s})\big]-\int_t^T\frac{\partial}{\partial x_\nu}\mathbb{E}\big[(F(v^\infty))(r,Y^{0,x,n}_{t,r})\big]\,\mathrm{d}r\\&=\sum_{s\in(t,T)} q^{Q,[t,T]}(s)\frac{\partial}{\partial x_\nu}\mathbb{E}\big[\Psi_1(s,Y^{0,x,n}_{t,s})\big]-\int_t^T\frac{\partial}{\partial x_\nu}\mathbb{E}\big[\Psi_1(r,Y^{0,x,n}_{t,r})\big]\,\mathrm{d}r\\&=\Big(\frac{\partial^{2Q}}{\partial t^{2Q}}\frac{\partial}{\partial x_\nu}\mathbb{E}\big[\Psi_1(s,Y^{0,x,n}_{t,s})\big]\Big)\Big|_{s=\xi_\nu}\frac{[Q!]^4(T-t)^{2Q+1}}{(2Q+1)[(2Q)!]^3}\\&=\frac{\partial}{\partial x_\nu}\mathbb{E}\big[\Psi_{2Q+1}(\xi_\nu,Y^{0,x,n}_{t,\xi_\nu})\big]\frac{[Q!]^4(T-t)^{2Q+1}}{(2Q+1)[(2Q)!]^3}.\end{aligned}\tag{3.188}$$

Combining this with (3.179) and (3.187) proves (3.178). The proof of Lemma 3.2.5 is thus complete. □

Theorem 3.2.4 and Lemma 3.2.5 lead together to the following Corollary which provides an upper bound for the approximation error under the assumption of a smooth solution. The proof of Corollary 3.2.6 uses ideas from [55, Corollary 4.6].

Corollary 3.2.6. *Assume Setting 3.1.1, let $n,m,Q\in\mathbb{N}$, $M\in[2,\infty)$, $a\in[0,1]$, $\nu_0\in\{1,2,\ldots,d+1\}$, $t_0\in[0,T)$, $x\in\mathbb{R}^d$ assume that $u^\infty\in C^\infty([0,T]\times\mathbb{R}^d,\mathbb{R})$, and let $C\in[1,\infty)$ satisfy that*

$$\begin{aligned}C=\max\Big\{&\sqrt{\pi T},\mathcal{L},\big(\sqrt{T}+(1+\tfrac{1}{\sqrt{\alpha}}\sqrt{3KT}\exp\big(\tfrac{5}{2}KT\big))\big),\\&K\sqrt{\tfrac{1}{\alpha}(1+3KT\exp(5KT))},\big(2K\big(\gamma_x^{(4)}T\exp\big(6+32K^4T\big)\big)^{\frac14}+\gamma_x^{(1)}\big)\\&\cdot\Big[\sqrt{T}+\sqrt{\tfrac{6}{\alpha}}\big(1+\sqrt{6K(1+T)}\exp(40K^2(1+T)^2)\big)\Big]\Big\},\end{aligned}\tag{3.189}$$

Then it holds that

$$
\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t_0,x)-v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac{1}{2}}\\
&\leqslant \left(1+6(4C^3)^{m-1}e^M\right)\frac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^{m-3}}}\\
&\quad+\left(1+14C(8C^3)^{m-1}\right)2C\sqrt{\left(\frac{T-t_0}{n^2}+\frac{1}{n}\right)}\left(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right]\right.\\
&\quad\left.+2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}\left(|(F(0))(s,y)|+C|\!|\!|v^\infty(s,y)|\!|\!|\right)\right]\right)\\
&\quad+\frac{2C^3+(4C^3)^m e^M}{\sqrt{M^{m-3}}}\left(\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|\right]\right.\\
&\quad\left.+\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|\!|\!|v^\infty(s,y)|\!|\!|\right]\right)\\
&\quad+\frac{(1+14(8C^3)^{m-1})T^{2Q+1}}{Q^{2aQ}}\\
&\quad\cdot\left[\sup_{k\in\mathbb{N}}\sup_{s\in[0,T]}\sup_{y\in\mathbb{R}^d}\frac{\left|\!\left|\!\left|(1,\nabla_x)((\frac{\partial u^\infty}{\partial t})+\langle\mu,(\nabla_x u^\infty)\rangle+\frac{1}{2}\operatorname{Tr}(\sigma\sigma^*(\operatorname{Hess}_x u^\infty)))^k(s,y)\right|\!\right|\!\right|}{(k!)^{1-a}}\right].
\end{aligned}
\tag{3.190}
$$

Proof of Corollary 3.2.6. To prove (3.190) we assume w.l.o.g. that the right-hand side of (3.190) is finite. Note that [90, Displays (1)-(2)] demonstrates that for all $k\in\mathbb{N}$ it holds that

$$
\sqrt{2\pi k}\left(\frac{k}{e}\right)^k\leqslant k!\leqslant\sqrt{2\pi k}\left(\frac{k}{e}\right)^k e^{\frac{1}{12}}.
\tag{3.191}
$$

Combining this with the fact that for all $k\in\mathbb{N}$ it holds that $\pi e^{\frac{1}{3}}k\leqslant 8^k$ implies that for all $k\in\mathbb{N}$ it holds that

$$
\begin{aligned}
&\frac{k^{2ak}((2k+1)!)^{1-a}[k!]^4}{(2k+1)[(2k)!]^3}\leqslant\frac{k^{2ak}[k!]^4}{[(2k)!]^{2+a}}\leqslant\frac{k^{2ak}[\sqrt{2\pi}k^{k+\frac{1}{2}}e^{-k+\frac{1}{12}}]^4}{[\sqrt{2\pi}(2k)^{2k+\frac{1}{2}}e^{-2k}]^{2+a}}\\
&=(\sqrt{2\pi})^{2-a}k^{1-\frac{a}{2}}e^{\frac{1}{3}}2^{-2(2k+\frac{1}{2})}\left(\frac{e^{2k}}{2^{2k+\frac{1}{2}}}\right)^a\leqslant 2\pi k e^{\frac{1}{3}}2^{-4k-1}e^{2k}2^{-2k}\\
&=\pi e^{\frac{1}{3}}k\left(\frac{e^2}{64}\right)^k\leqslant\pi e^{\frac{1}{3}}k8^{-k}\leqslant 1.
\end{aligned}
\tag{3.192}
$$

Hence, we obtain for all $z\in[0,\infty)$ that

$$
\begin{aligned}
&\frac{z[Q!]^4}{(2Q+1)[(2Q)!]^3}\leqslant\frac{1}{Q^{2aQ}}\left[\sup_{k\in\mathbb{N}}\frac{zk^{2ak}[k!]^4}{(2k+1)[(2k)!]^3}\right]\\
&\leqslant\frac{1}{Q^{2aQ}}\left[\sup_{l\in\mathbb{N}}\frac{l^{2al}(l!)^{1-a}[l!]^4}{(2l+1)[(2l)!]^3}\right]\left[\sup_{k\in\mathbb{N}}\frac{z}{(k!)^{1-a}}\right]\\
&\leqslant\frac{1}{Q^{2aQ}}\left[\sup_{l\in\mathbb{N}}\frac{l^{2al}((2l+1)!)^{1-a}[l!]^4}{(2l+1)[(2l)!]^3}\right]\left[\sup_{k\in\mathbb{N}}\frac{z}{(k!)^{1-a}}\right]\leqslant\frac{1}{Q^{2aQ}}\left[\sup_{k\in\mathbb{N}}\frac{z}{(k!)^{1-a}}\right].
\end{aligned}
\tag{3.193}
$$

Next observe that Theorem 3.2.4 and (3.158) show that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t_0,x)-v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac{1}{2}} \\
&\leqslant \left(1+3(4C^3)^{m-1}e^M\right)\frac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^{m-3}}} \\
&\quad + \left(1+14(8C^3)^{m-1}\right)2C\sqrt{\left(\frac{T-t_0}{n^2}+\frac{1}{n}\right)}\Bigg(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right] \\
&\quad + 2C\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}\left(|(F(0))(s,y)|+C|\!|\!|v^\infty(s,y)|\!|\!|\right)\right]\Bigg) \\
&\quad + \left(2C^3+(4C^3)^m e^M\right)\Bigg(\frac{[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|]}{\sqrt{M^{m-3}}} \\
&\quad + \frac{[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|\!|\!|v^\infty(s,y)|\!|\!|]}{\sqrt{M^{m-3}}}\Bigg) \\
&+ \left(1+14(8C^3)^{m-1}\right)\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|\!|\!|\varepsilon(s,y)|\!|\!|\right].
\end{aligned} \tag{3.194}$$

Combining this with Lemma 3.2.5 demonstrates that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t_0,x)-v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac{1}{2}} \\
&\leqslant \left(1+6(4C^3)^{m-1}e^M\right)\frac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^{m-3}}} \\
&\quad + \left(1+14C(8C^3)^{m-1}\right)2C\sqrt{\left(\frac{T-t_0}{n^2}+\frac{1}{n}\right)}\Bigg(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right] \\
&\quad + 2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}\left(|(F(0))(s,y)|+C|\!|\!|v^\infty(s,y)|\!|\!|\right)\right]\Bigg) \\
&\quad + \frac{2C^3+(4C^3)^m e^M}{\sqrt{M^{m-3}}}\Bigg(\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|\right] \\
&\quad + \left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|\!|\!|v^\infty(s,y)|\!|\!|\right]\Bigg) \\
&\quad + \left(1+14(8C^3)^{m-1}\right)\Bigg(\Bigg[\sup_{s\in[0,T]}\sup_{y\in\mathbb{R}^d}\Big\|(1,\nabla_x)\big((\tfrac{\partial u^\infty}{\partial t})+\langle\mu,(\nabla_x u^\infty)\rangle \\
&\quad + \tfrac{1}{2}\operatorname{Tr}(\sigma\sigma^*(\operatorname{Hess}_x u^\infty))\big)^{2Q+1}(s,y)\Big\|\frac{[Q!]^4T^{2Q+1}}{(2Q+1)[(2Q)!]^3}\Bigg].
\end{aligned} \tag{3.195}$$

This and (3.193) show that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{n,m,M,Q}(t_0,x)-v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac{1}{2}} \\
&\leqslant \left(1+6(4C^3)^{m-1}e^M\right)\frac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{M^{m-3}}} \\
&\quad+\left(1+14C(8C^3)^{m-1}\right)2C\sqrt{\left(\frac{T-t_0}{n^2}+\frac{1}{n}\right)}\left(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right]\right. \\
&\quad\left.+2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}\left(|(F(0))(s,y)|+C|\!|\!|v^\infty(s,y)|\!|\!|\right)\right]\right) \\
&\quad+\frac{2C^3+(4C^3)^m e^M}{\sqrt{M^{m-3}}}\left(\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|\right]\right. \\
&\quad\left.+\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|\!|\!|v^\infty(s,y)|\!|\!|\right]\right) \\
&\quad+\frac{(1+14(8C^3)^{m-1})T^{2Q+1}}{Q^{2aQ}} \\
&\quad\cdot\left[\sup_{k\in\mathbb{N}}\sup_{s\in[0,T]}\sup_{y\in\mathbb{R}^d}\frac{\left\|(1,\nabla_x)\big((\frac{\partial u^\infty}{\partial t})+\langle\mu,(\nabla_x u^\infty)\rangle+\frac{1}{2}\operatorname{Tr}(\sigma\sigma^*(\operatorname{Hess}_x u^\infty))\big)^k(s,y)\right\|}{(k!)^{1-a}}\right].
\end{aligned} \tag{3.196}$$

The proof of Corollary 3.2.6 is thus complete. □

The following corollary is an application of Corollary 3.2.6 to the special case that $m = M = Q$, $n = m^m$, and $a = \frac{1}{4}$. Note that in the case $a = \frac{1}{4}$ it holds that $\sqrt{M^{-m}}$ and Q^{-2aQ} are equal.

Corollary 3.2.7. *Assume Setting 3.1.1, let $N \in \mathbb{N} \cap [2,\infty)$, $\nu_0 \in \{1,2,\ldots,d+1\}$, $t_0 \in [0,T)$, $x \in \mathbb{R}^d$, $\nu_0 \in \{1,2,\ldots,d+1\}$, assume that $u^\infty \in C^\infty([0,T]\times\mathbb{R}^d,\mathbb{R})$, and let $C \in [0,\infty)$ satisfy that*

$$\begin{aligned}
C = \max\Big\{&\sqrt{\pi T}, \mathcal{L}, \left(\sqrt{T}+(1+\tfrac{1}{\sqrt{\alpha}}\sqrt{3KT}\exp\left(\tfrac{5}{2}KT\right)\right)\big), \\
&K\sqrt{\tfrac{1}{\alpha}\left(1+3KT\exp(5KT)\right)}, \left(2K\left(\gamma_x^{(4)}T\exp\left(6+32K^4T\right)\right)^{\frac{1}{4}}+\gamma_x^{(1)}\right) \\
&\cdot\left[\sqrt{T}+\sqrt{\tfrac{6}{\alpha}}\left(1+\sqrt{6K(1+T)}\exp(40K^2(1+T)^2)\right)\right]\Big\},
\end{aligned} \tag{3.197}$$

Then it holds that

$$
\begin{aligned}
&\left(\mathbb{E}\left[\left|(V^0_{N^N,N,N,N}(t_0,x) - v^\infty(t_0,x))_{\nu_0}\right|^2\right]\right)^{\frac{1}{2}} \\
&\leqslant \left(1+6(4C^3)^{N-1}e^N\right)\frac{C^2(\sqrt{T}+\sqrt{6})}{\sqrt{N^{N-3}}} \\
&\quad + \left(1+14C(8C^3)^{N-1}\right)2C\sqrt{\left(\frac{T-t_0}{N^{2N}}+\frac{1}{N^N}\right)}\Bigg(\left[\sup_{y\in\mathbb{R}^d}|g(y)|\right] \\
&\quad + 2T\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}\left(|(F(0))(s,y)| + C|\!|\!|v^\infty(s,y)|\!|\!|\right)\right]\Bigg) \\
&\quad + \frac{2C^3+(4C^3)^N e^N}{\sqrt{N^{N-3}}}\Bigg(\left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|(F(0))(s,y)|\right] \\
&\quad + \left[\sup_{s\in[t_0,T]}\sup_{y\in\mathbb{R}^d}|\!|\!|v^\infty(s,y)|\!|\!|\right]\Bigg) \\
&\quad + \frac{(1+14(8C^3)^{N-1})T^{2N+1}}{\sqrt{N^N}} \\
&\quad \cdot\left[\sup_{k\in\mathbb{N}}\sup_{s\in[0,T]}\sup_{y\in\mathbb{R}^d}\frac{\left\|(1,\nabla_x)((\frac{\partial u^\infty}{\partial t})+\langle\mu,(\nabla_x u^\infty)\rangle+\frac{1}{2}\operatorname{Tr}(\sigma\sigma^*(\operatorname{Hess}_x u^\infty)))^k(s,y)\right\|}{(k!)^{\frac{3}{4}}}\right].
\end{aligned}
\tag{3.198}
$$

Proof of Corollary 3.2.7. Note that Corollary 3.2.6 (applied with $a \curvearrowleft \frac{1}{4}$, $n \curvearrowleft N^N$, $m \curvearrowleft N$, $M \curvearrowleft N$, $Q \curvearrowleft N$ in the notation of Corollary 3.2.6) implies (3.198). The proof of Corollary 3.2.7 is thus complete. ☐

The following corollary proves that if the constant in (3.197) below is finite then the computational complexity of the approximation scheme in (3.17) is bounded by $O(d\varepsilon^{-(6+\delta)})$ for any $\delta \in (0,\infty)$ where d denotes the dimensionality of the problem and ε the prescribed accuracy. In the corollary for every $n,m,M,Q \in \mathbb{N}$ we think of $RN_{n,m,M,Q}$ as the number of realizations of a scalar standard normal random variable required to compute one realization of the random variable $V^0_{n,m,M,Q}(0,0)\colon \Omega \to \mathbb{R}$ and we think of $FE_{n,m,M,Q}$ as the number of function evaluations of f and g required to compute on realization of the random variable $V^0_{n,m,M,Q}(0,0)\colon \Omega \to \mathbb{R}$.

Corollary 3.2.8. *Assume Setting 3.1.1, let $\delta \in (0,\infty)$, assume that $u^\infty \in C^\infty([0,T]\times\mathbb{R}^d,\mathbb{R})$, let $C \in [0,\infty)$ satisfy that*

$$
\begin{aligned}
C = \max\Bigg\{&\sqrt{\pi T}, \mathcal{L}, \left(\sqrt{T}+(1+\tfrac{1}{\sqrt{\alpha}}\sqrt{3KT}\exp\left(\tfrac{5}{2}KT\right))\right), \\
&K\sqrt{\tfrac{1}{\alpha}(1+3KT\exp(5KT))}, \Big(2K\big([\sup_{x\in\mathbb{R}^d}\gamma_x^{(4)}]T\exp\left(6+32K^4T\right)\big)^{\frac{1}{4}} \\
&+[\sup_{x\in\mathbb{R}^d}\gamma_x^{(1)}]\Big)\left[\sqrt{T}+\sqrt{\tfrac{6}{\alpha}}(1+\sqrt{6K(1+T)}\exp(40K^2(1+T)^2))\right], \\
&\sup\nolimits_{k\in\mathbb{N}_0}\sup\nolimits_{s\in[0,T]}\sup\nolimits_{y\in\mathbb{R}^d}\frac{\left\|(1,\nabla_x)((\frac{\partial u^\infty}{\partial t})+\langle\mu,(\nabla_x u^\infty)\rangle+\frac{1}{2}\operatorname{Tr}(\sigma\sigma^*(\operatorname{Hess}_x u^\infty)))^k(s,y)\right\|}{(k!)^{\frac{3}{4}}}, \\
&\sup\nolimits_{y\in\mathbb{R}^d}|g(y)|, \sup\nolimits_{s\in[0,T)}\sup\nolimits_{y\in\mathbb{R}^d}|(F(0))(s,y)|\Bigg\},
\end{aligned}
\tag{3.199}
$$

let $(RN_{n,m,M,Q})_{n,m,M,Q\in\mathbb{Z}} \subseteq \mathbb{N}_0$ satisfy for all $n,m,M,Q \in \mathbb{N}$ that $RN_{n,0,M,Q} = 0$ and

$$
RN_{n,m,M,Q} \leqslant dnM^m + \sum_{l=0}^{m-1}\left[QM^{m-l}(dn + RN_{n,l,M,Q} + \mathbb{1}_{\mathbb{N}}(l) + RN_{n,l-1,M,Q})\right], \tag{3.200}
$$

and let $(FE_{n,m,M,Q})_{n,m,M,Q\in\mathbb{Z}} \subseteq \mathbb{N}_0$ *satisfy for all* $n, m, M, Q \in \mathbb{N}$ *that* $FE_{n,0,M,Q} = 0$ *and*

$$FE_{n,m,M,Q} \leqslant M^m + \sum_{l=0}^{m-1} [QM^{m-l}(1 + FE_{n,l,M,Q} + \mathbb{1}_{\mathbb{N}}(l) \cdot FE_{n,l-1,M,Q})]. \tag{3.201}$$

Then it holds for all $N \in \mathbb{N} \cap [2, \infty)$ *that*

$$\begin{aligned} &RN_{N^N,N,N,N} + FE_{N^N,N,N,N} \\ &\leqslant \left[\sup_{t\in[0,T)} \sup_{x\in\mathbb{R}^d} \|V^0_{N^N,N,N,N}(t,x) - v^\infty(t,x)\|_{L^2(\mathbb{P};\mathbb{R}^d)} \right]^{-(6+\delta)} \\ &\quad \cdot 16d \sum_{\mathfrak{n}\in\mathbb{N}} \sqrt{\mathfrak{n}^{-\delta\mathfrak{n}}} \Big(56C^{4\mathfrak{n}} 2^{5\mathfrak{n}} T^{3\mathfrak{n}} \Big)^{(6+\delta)}. \end{aligned} \tag{3.202}$$

Proof of Corollary 3.2.8. First observe that [36, Lemma 3.15] and [36, Lemma 3.16] (applied for all $N \in \mathbb{N}$ with $d \curvearrowleft dN^N$ in the notation of [36, Lemma 3.15] and [36, Lemma 3.16]) demonstrate that for all $N \in \mathbb{N}$ it holds that $RN_{N^N,N,N,N} \leqslant 8dN^{3N}$ and $FE_{N^N,N,N,N} \leqslant 8N^{3N}$. Combining this with Corollary 3.2.7 and (3.197) shows that for all $N \in \mathbb{N}$ it holds that

$$\begin{aligned} &(RN_{N^N,N,N,N} + FE_{N^N,N,N,N}) \\ &\quad \cdot \left[\sup_{t\in[0,T)} \sup_{x\in\mathbb{R}^d} \left\| V^0_{N^N,N,N,N}(t,x) - v^\infty(t,x) \right\|_{L^2(\mathbb{P};\mathbb{R}^{d+1})} \right]^{(6+\delta)} \\ &\leqslant 8(d+1)N^{3N} \Bigg(\frac{(1 + 6(4C^3)^{N-1}e^N)C^2(\sqrt{T} + \sqrt{6})}{\sqrt{N^{N-3}}} \\ &\quad + \frac{(1 + 14(8C^3)^{N-1}2C^3)(C + 4C^2T)}{\sqrt{N^N}} + \frac{(2C^3 + (4C^3)^N e^N)2C}{\sqrt{N^{N-3}}} \\ &\quad + \frac{(1 + 14(8C^3)^{N-1})T^{2N+1}C}{\sqrt{N^N}} \Bigg)^{(6+\delta)} \\ &\leqslant 8(d+1)N^{3N} \left(\frac{56C^{3N+1}2^{5N}(T+1)^{2N+1}}{\sqrt{N^N}} \right)^{(6+\delta)} \\ &\leqslant 16d \sum_{\mathfrak{n}\in\mathbb{N}} \frac{\Big(56C^{4\mathfrak{n}}2^{5\mathfrak{n}}(T+1)^{3\mathfrak{n}}\Big)^{(6+\delta)}}{\sqrt{\mathfrak{n}^{\delta\mathfrak{n}}}} < \infty. \end{aligned} \tag{3.203}$$

The proof of Corollary 3.2.8 is thus complete. □

Chapter 4

Convergence proof for stochastic gradient descent (SGD) in the training of deep neural networks with ReLU activation for constant target functions

In this chapter we focus on a different approach to calculate solutions of high-dimensional semilinear PDEs with gradient-dependent nonlinearities: reformulating the PDEs as stochastic learning problems and then solving them by using artificial neural networks (ANNs). With this goal, we prove in this chapter the convergence of the expected risk of certain SGD processes in the training of fully-connected feedforward deep artificial neural networks (DNNs) with rectified linear unit (ReLU) activation function as the number of SGD steps goes to infinity. In our approach we consider the plain vanilla SGD optimization method. One of the main challenges in our convergence analysis arises from the fact that the ReLU activation function

$$\mathbb{R} \ni x \mapsto \max\{x, 0\} \in \mathbb{R} \tag{4.1}$$

is not differentiable and therefore does not admit a gradient. We overcome this issue by defining a suitable sequence of approximating functions which are appropriate continuously differentiable, converge to the ReLU function, and whose derivatives converge to the left derivative of the ReLU function. With this technique, we can prove that the expected risk of the SGD process converges to zero as the number of steps goes to infinity provided that the target function is constant and the step sizes are sufficiently small but not L^1-summable.

To illustrate the setting of deep neural networks considered in this chapter we offer in Figure 4.1 a graphic visualization of a deep neural network with 3 hidden layers. The mathematical description of such deep neural networks requires a lot of terminology (cf. Setting 4.1.1). To better understand the appearing objects, we provide the following explanations. We denote by $L \in \mathbb{N}$ the number of affine linear transformations in the considered DNN. In this sense, $L-1$ denotes the number of hidden layers and $L+1$ denotes the number of overall layers including the input and output layer of the considered DNN. We denote by $\ell_0, \ell_1, \ldots, \ell_L \in \mathbb{N}$ the dimensions of the layers which present the number of neurons of the layers of the considered DNN. We call $[a, b]^{\ell_0} \ni x \mapsto \mathcal{N}_\infty^{L,\theta}(x) \in \mathbb{R}^{\ell_L}$, $\theta = (\theta_1, \ldots, \theta_\mathfrak{d}) \in \mathbb{R}^\mathfrak{d}$, the realization function of the considered deep ReLU ANNs (see (4.5)), $\mathcal{L}_\infty \colon \mathbb{R}^\mathfrak{d} \to \mathbb{R}$ the risk function (see (4.6)), $\mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_\mathfrak{d}) \colon \mathbb{R}^\mathfrak{d} \to \mathbb{R}^\mathfrak{d}$ the generalized

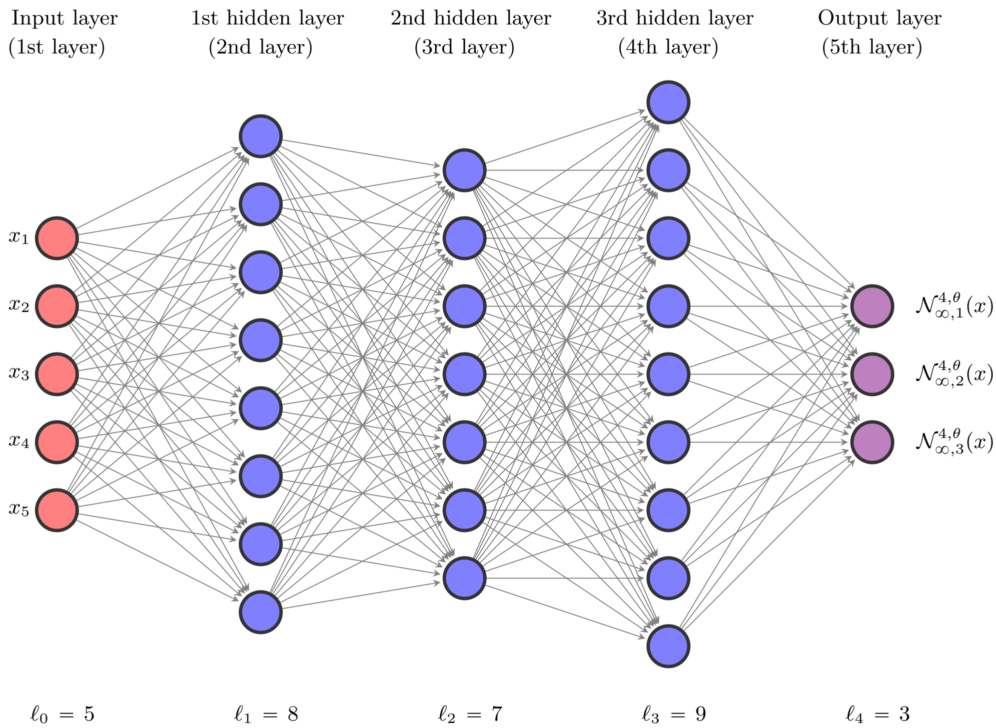

Figure 4.1: Graphical illustration of the used deep ANN architecture in the case of a simple example deep ANN with 3 hidden layers (corresponding to $L = 4$ affine linear transformations), with 5 neurons on the input layer (corresponding to $\ell_0 = 5$), 8 neurons on the 1st hidden layer (corresponding to $\ell_1 = 8$), 7 neurons on the 2nd hidden layer (corresponding to $\ell_2 = 7$), 9 neurons on the 3rd hidden layer (corresponding to $\ell_3 = 9$), and 3 neurons on the output layer (corresponding to $\ell_4 = 3$). In this situation we have for every ANN parameter vector $\theta \in \mathbb{R}^{\mathfrak{d}} = \mathbb{R}^{213}$ that the realization function $\mathbb{R}^5 \ni x \mapsto \mathcal{N}^{4,\theta}_{\infty}(x) \in \mathbb{R}^3$ of the considered deep ANN maps the 5-dimensional input vector $x = (x_1, x_2, x_3, x_4, x_5) \in [a,b]^5$ to the 3-dimensional output vector $\mathcal{N}^{4,\theta}_{\infty}(x) = (\mathcal{N}^{4,\theta}_{\infty,1}(x), \mathcal{N}^{4,\theta}_{\infty,2}(x), \mathcal{N}^{4,\theta}_{\infty,3}(x))$.

gradient function associated to the risk function, and $V: \mathbb{R}^{\mathfrak{d}} \to \mathbb{R}$ the Lyapunov function which we will use for the mathematical analysis of GD and SGD optimization methods in the training of DNNs (see (4.7)). We denote by $f: [a,b]^{\ell_0} \to \mathbb{R}^{\ell_L}$ the target function which describes the relationship between the input data and the output data. The main purpose of the DNN is to approximatively learn this target function. In Section 4.4 we assume that the target function is a constant function represented by the ℓ_L-dimensional vector $\xi \in R^{\ell_L}$. This assumption is fundamental for the procedure we used in the proofs below. In general, in settings where the target function is not constant, there exist non-global local minimum points in the risk landscape whose domain of attraction admit a strictly positive probability. In this case, the conclusions made in our analysis therefore do not hold true.

This Chapter is structured as follows. In Section 4.1 several regularity properties and representation results for the risk function and its generalized gradient function are proven in the setting of general measurable target functions. These results are used in Section 4.2 to show that - in the case of a constant target function - the risks of GF processes converge to 0 with convergence rate 1. In Section 4.3 the findings from Section 4.1 and 4.2 are applied to show that - under the assumption of a constant target function and L^1-summable step sizes - the risks of GD processes converge to zero. These results are then extended in Section 4.4 proving that, in the setting of a constant target function and L^1-summable but sufficiently small step sizes, the expectations of risks of SGD processes converge to zero. In Section 4.5 we offer two numerical examples to display practical applications of the theoretical findings in Theorem 4.4.11.

4.1 Properties of the risk function associated to deep artificial neural networks (ANNs)

The main goal of this section is to introduce and analyse the risk function and its generalized gradient function in the setting of deep ANNs with an arbitrary number of hidden layers and a general measurable target function. We start by giving a mathematical description of the considered DNN in Section 4.1.1. In Section Subsection 4.1.2 we approximate the considered risk function by approximating the ReLU activation function. One of this section's main results is Theorem 4.1.9 in Section 4.1.3 which establishes an explicit representation for the generalized gradients of the risk function. Section 4.1.4 ensures that the considered risk and realization function are Lipschitz continuous. Another main result of this section is Theorem 4.1.11 in Subsection 4.1.5 which proves an explicit polynomial growth estimate for the generalized gradient functions. In Subsections 4.1.6 and 4.1.7 we analyse the convexity properties of the considered risk function. In particular, we establish in Corollary 4.1.19 that the risk function is convex if and only if the product of the total mass and the number of hidden layers vanishes. This implies that for every measurable target function the risk function is convex provided that the underlying measure is not the zero measure.

4.1.1 Mathematical framework for deep ANNs with ReLU activation

Throughout this chapter we frequently use the following mathematical description of DNNs.

Setting 4.1.1. *Let $L, \mathfrak{d} \in \mathbb{N}$, $(\ell_k)_{k \in \mathbb{N}_0} \subseteq \mathbb{N}$, $a, \mathbf{a} \in \mathbb{R}$, $b \in (a, \infty)$, $\mathscr{A} \in (0, \infty)$, $\mathscr{B} \in (\mathscr{A}, \infty)$ satisfy $\mathfrak{d} = \sum_{k=1}^{L} \ell_k(\ell_{k-1} + 1)$ and $\mathbf{a} = \max\{|a|, |b|, 1\}$, for every $\theta = (\theta_1, \dots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$*

let $\mathfrak{w}^{k,\theta} = (\mathfrak{w}^{k,\theta}_{i,j})_{(i,j)\in\{1,\dots,\ell_k\}\times\{1,\dots,\ell_{k-1}\}} \in \mathbb{R}^{\ell_k\times\ell_{k-1}}$, $k \in \mathbb{N}$, *and* $\mathfrak{b}^{k,\theta} = (\mathfrak{b}^{k,\theta}_1, \dots, \mathfrak{b}^{k,\theta}_{\ell_k}) \in \mathbb{R}^{\ell_k}$, $k \in \mathbb{N}$, *satisfy for all* $k \in \{1, \dots, L\}$, $i \in \{1, \dots, \ell_k\}$, $j \in \{1, \dots, \ell_{k-1}\}$ *that*

$$\mathfrak{w}^{k,\theta}_{i,j} = \theta_{(i-1)\ell_{k-1}+j+\sum_{h=1}^{k-1}\ell_h(\ell_{h-1}+1)} \qquad \text{and} \qquad \mathfrak{b}^{k,\theta}_i = \theta_{\ell_k\ell_{k-1}+i+\sum_{h=1}^{k-1}\ell_h(\ell_{h-1}+1)}, \tag{4.2}$$

for every $k \in \mathbb{N}$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ *let* $\mathcal{A}^\theta_k = (\mathcal{A}^\theta_{k,1}, \dots, \mathcal{A}^\theta_{k,\ell_k}) \colon \mathbb{R}^{\ell_{k-1}} \to \mathbb{R}^{\ell_k}$ *satisfy for all* $x \in \mathbb{R}^{\ell_{k-1}}$ *that* $\mathcal{A}^\theta_k(x) = \mathfrak{b}^{k,\theta} + \mathfrak{w}^{k,\theta}x$, *let* $\mathscr{R}_r \colon \mathbb{R} \to \mathbb{R}$, $r \in [1, \infty]$, *satisfy for all* $r \in [1, \infty)$, $x \in (-\infty, \mathscr{A}r^{-1}]$, $y \in \mathbb{R}$, $z \in [\mathscr{B}r^{-1}, \infty)$ *that*

$$\mathscr{R}_r \in C^1(\mathbb{R}, \mathbb{R}),\ \mathscr{R}_r(x) = 0,\ 0 \leqslant \mathscr{R}_r(y) \leqslant \mathscr{R}_\infty(y) = \max\{y, 0\}, \text{ and } \mathscr{R}_r(z) = z, \tag{4.3}$$

assume $\sup_{r\in[1,\infty)}\sup_{x\in\mathbb{R}} |(\mathscr{R}_r)'(x)| < \infty$, *let* $\|\cdot\| \colon (\cup_{n\in\mathbb{N}}\mathbb{R}^n) \to \mathbb{R}$, $\langle\cdot,\cdot\rangle \colon (\cup_{n\in\mathbb{N}}(\mathbb{R}^n \times \mathbb{R}^n)) \to \mathbb{R}$, *and* $\mathfrak{M}_r \colon (\cup_{n\in\mathbb{N}}\mathbb{R}^n) \to (\cup_{n\in\mathbb{N}}\mathbb{R}^n)$, $r \in [1, \infty]$, *satisfy for all* $r \in [1, \infty]$, $n \in \mathbb{N}$, $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ *that*

$$\|x\| = (\textstyle\sum_{i=1}^n |x_i|^2)^{1/2},\ \langle x, y\rangle = \sum_{i=1}^n x_iy_i, \text{ and } \mathfrak{M}_r(x) = (\mathscr{R}_r(x_1), \dots, \mathscr{R}_r(x_n)), \tag{4.4}$$

for every $\theta \in \mathbb{R}^{\mathfrak{d}}$ *let* $\mathcal{N}^{k,\theta}_r = (\mathcal{N}^{k,\theta}_{r,1}, \dots, \mathcal{N}^{k,\theta}_{r,\ell_k}) \colon \mathbb{R}^{\ell_0} \to \mathbb{R}^{\ell_k}$, $r \in [1, \infty]$, $k \in \mathbb{N}$, *and* $\mathcal{X}^{k,\theta}_i \subseteq \mathbb{R}^{\ell_0}$, $k, i \in \mathbb{N}$, *satisfy for all* $r \in [1, \infty]$, $k \in \mathbb{N}$, $i \in \{1, \dots, \ell_k\}$, $x \in \mathbb{R}^{\ell_0}$ *that*

$$\mathcal{N}^{1,\theta}_r(x) = \mathcal{A}^\theta_1(x), \qquad \mathcal{N}^{k+1,\theta}_r(x) = \mathcal{A}^\theta_{k+1}(\mathfrak{M}_{r^{1/k}}(\mathcal{N}^{k,\theta}_r(x))), \tag{4.5}$$

and $\mathcal{X}^{k,\theta}_i = \{y \in [a, b]^{\ell_0} \colon \mathcal{N}^{k,\theta}_{\infty,i}(y) > 0\}$, *let* $\mu \colon \mathcal{B}([a, b]^{\ell_0}) \to [0, \infty]$ *be a measure, let* $\mathfrak{m} \in \mathbb{R}$ *satisfy* $\mathfrak{m} = \mu([a, b]^{\ell_0})$, *let* $f = (f_1, \dots, f_{\ell_L}) \colon [a, b]^{\ell_0} \to \mathbb{R}^{\ell_L}$ *be measurable, for every* $r \in [1, \infty]$ *let* $\mathcal{L}_r \colon \mathbb{R}^{\mathfrak{d}} \to \mathbb{R}$ *satisfy for all* $\theta \in \mathbb{R}^{\mathfrak{d}}$ *that*

$$\mathcal{L}_r(\theta) = \textstyle\int_{[a,b]^{\ell_0}} \|\mathcal{N}^{L,\theta}_r(x) - f(x)\|^2\, \mu(\mathrm{d}x), \tag{4.6}$$

let $\mathcal{G} = (\mathcal{G}_1, \dots, \mathcal{G}_{\mathfrak{d}}) \colon \mathbb{R}^{\mathfrak{d}} \to \mathbb{R}^{\mathfrak{d}}$ *satisfy for all* $\theta \in \{\vartheta \in \mathbb{R}^{\mathfrak{d}} \colon ((\nabla\mathcal{L}_r)(\vartheta))_{r\in[1,\infty)} \text{ is convergent}\}$ *that* $\mathcal{G}(\theta) = \lim_{r\to\infty}(\nabla\mathcal{L}_r)(\theta)$, *and let* $V \colon \mathbb{R}^{\mathfrak{d}} \to \mathbb{R}$ *satisfy for all* $\theta \in \mathbb{R}^{\mathfrak{d}}$ *that*

$$V(\theta) = \big[\textstyle\sum_{k=1}^L (k\|\mathfrak{b}^{k,\theta}\|^2 + \sum_{i=1}^{\ell_k}\sum_{j=1}^{\ell_{k-1}} |\mathfrak{w}^{k,\theta}_{i,j}|^2)\big] - 2L\langle f(0), \mathfrak{b}^{L,\theta}\rangle. \tag{4.7}$$

4.1.2 Approximations of the realization functions of the considered deep ANNs

The following lemma presents elementary results for the realization functions of the DNN defined in Setting 4.1.1.

Lemma 4.1.2. *Assume Setting 4.1.1 and let* $\theta \in \mathbb{R}^{\mathfrak{d}}$, $r \in [1, \infty]$. *Then*

(i) it holds for all $k \in \mathbb{N}$, $i \in \{1, \dots, \ell_k\}$, $x = (x_1, \dots, x_{\ell_{k-1}}) \in \mathbb{R}^{\ell_{k-1}}$ *that*

$$\mathcal{A}^\theta_{k,i}(x) = \mathfrak{b}^{k,\theta}_i + \sum_{j=1}^{\ell_{k-1}} \mathfrak{w}^{k,\theta}_{i,j}x_j, \tag{4.8}$$

(ii) it holds for all $i \in \{1, \dots, \ell_1\}$, $x = (x_1, \dots, x_{\ell_0}) \in \mathbb{R}^{\ell_0}$ *that*

$$\mathcal{N}^{1,\theta}_{r,i}(x) = \mathfrak{b}^{1,\theta}_i + \sum_{j=1}^{\ell_0} \mathfrak{w}^{1,\theta}_{i,j}x_j, \tag{4.9}$$

and

(iii) *it holds for all* $k \in \mathbb{N}$, $i \in \{1, \ldots, \ell_{k+1}\}$, $x \in \mathbb{R}^{\ell_0}$ *that*

$$\mathcal{N}_{r,i}^{k+1,\theta}(x) = \mathfrak{b}_i^{k+1,\theta} + \sum_{j=1}^{\ell_k} \mathfrak{w}_{i,j}^{k+1,\theta} \mathscr{R}_{r^{1/k}}(\mathcal{N}_{r,j}^{k,\theta}(x)). \tag{4.10}$$

Proof of Lemma 4.1.2. Observe that (4.5) and the assumption that for all $k \in \mathbb{N}$, $x \in \mathbb{R}^{\ell_{k-1}}$ it holds that $\mathcal{A}_k^\theta(x) = \mathfrak{b}^{k,\theta} + \mathfrak{w}^{k,\theta}x$ establish (4.8), (4.9), and (4.10). The proof of Lemma 4.1.2 is thus complete. □

Next we study the approximations of the ReLU activation function and establish the following convergence rate result.

Lemma 4.1.3. *Assume Setting 4.1.1. Then*

(i) *it holds for all* $x \in \mathbb{R}$ *that*

$$\limsup_{r\to\infty}(|\mathscr{R}_r(x) - \mathscr{R}_\infty(x)| + |(\mathscr{R}_r)'(x) - \mathbb{1}_{(0,\infty)}(x)|) = 0 \tag{4.11}$$

and

(ii) *it holds for all* $r \in [1,\infty)$, $x \in \mathbb{R}$ *that* $|\mathscr{R}_r(x) - \mathscr{R}_\infty(x)| \leqslant \mathscr{B}r^{-1}$.

Proof of Lemma 4.1.3. Note that (4.3) ensures that for all $r \in [1,\infty)$, $x \in (-\infty,0]$ it holds that $\mathscr{R}_r(x) = 0 = \max\{x,0\} = \mathscr{R}_\infty(x)$. Therefore, we obtain for all $r \in [1,\infty)$, $x \in (-\infty,0]$ that

$$|\mathscr{R}_r(x) - \mathscr{R}_\infty(x)| + |(\mathscr{R}_r)'(x) - \mathbb{1}_{(0,\infty)}(x)| = 0. \tag{4.12}$$

Furthermore, observe that (4.3) proves that for all $x \in (0,\infty)$ there exists $R \in [1,\infty)$ such that for all $r \in [R,\infty)$, $y \in (x/2,\infty)$ it holds that $\mathscr{R}_r(y) = \mathscr{R}_\infty(y)$. Hence, we obtain for all $x \in (0,\infty)$ that

$$\limsup_{r\to\infty}\big(|\mathscr{R}_r(x) - \mathscr{R}_\infty(x)| + |(\mathscr{R}_r)'(x) - (\mathscr{R}_\infty)'(x)|\big) = 0. \tag{4.13}$$

Combining this with (4.12) establishes item (i). Note that (4.3) shows that for all $r \in [1,\infty)$, $y \in (0,\mathscr{B}r^{-1})$ it holds that

$$|\mathscr{R}_r(y) - \mathscr{R}_\infty(y)| = \mathscr{R}_\infty(y) - \mathscr{R}_r(y) \leqslant \mathscr{R}_\infty(y) \leqslant y \leqslant \mathscr{B}r^{-1}. \tag{4.14}$$

Moreover, observe that (4.3) proves that for all $r \in [1,\infty)$, $x \in (-\infty,0] \cup [\mathscr{B}r^{-1},\infty)$ it holds that $\mathscr{R}_r(x) = \mathscr{R}_\infty(x)$. Combining this with (4.14) establishes item (ii). The proof of Lemma 4.1.3 is thus complete. □

The following proposition provides examples of suitable approximating functions of the ReLU function for deep ANNs.

Lemma 4.1.4. *Let* $\mathscr{A} \in (0,\infty)$, $\mathscr{B} \in (\mathscr{A},\infty)$, $\eta \in C^1(\mathbb{R},\mathbb{R})$ *satisfy for all* $x \in (-\infty,0]$, $y \in \mathbb{R}$, $z \in [1,\infty)$ *that* $\eta(x) = 0 \leqslant \eta(y) \leqslant 1 = \eta(z)$ *and for every* $r \in [1,\infty)$ *let* $\mathscr{R}_r\colon \mathbb{R} \to \mathbb{R}$ *satisfy for all* $x \in \mathbb{R}$ *that* $\mathscr{R}_r(x) = \max\{x,0\}\eta(\frac{rx-\mathscr{A}}{\mathscr{B}-\mathscr{A}})$. *Then*

(i) *it holds for all* $r \in [1,\infty)$, $x \in (-\infty,\mathscr{A}r^{-1}]$ *that* $\mathscr{R}_r(x) = 0$,

(ii) *it holds for all* $r \in [1,\infty)$, $x \in [\mathscr{B}r^{-1},\infty)$ *that* $\mathscr{R}_r(x) = x$,

(iii) *it holds for all* $r \in [1,\infty)$, $x \in \mathbb{R}$ *that* $0 \leqslant \mathscr{R}_r(x) \leqslant \max\{x,0\}$,

(iv) *it holds for all* $r \in [1,\infty)$ *that* $\mathscr{R}_r \in C^1(\mathbb{R},\mathbb{R})$, *and*

(v) *it holds that* $\sup_{r\in[1,\infty)}\sup_{x\in\mathbb{R}}|(\mathscr{R}_r)'(x)| < \infty$.

Proof of Lemma 4.1.4. Note that the assumption that for all $x \in (-\infty, 0]$ it holds that $\eta(x) = 0$ establishes item (i). Observe that the assumption that for all $x \in [1,\infty)$ it holds that $\eta(x) = 1$ proves item (ii). Note the assumption that for all $x \in \mathbb{R}$ it holds that $0 \leqslant \eta(x) \leqslant 1$ establishes item (iii). Observe that item (i), the fact that for all $r \in [1,\infty)$, $x \in (0,\infty)$ it holds that $\mathscr{R}_r(x) = x\eta(\frac{rx-\mathscr{A}}{\mathscr{B}-\mathscr{A}})$, and the assumption that $\eta \in C^1(\mathbb{R},\mathbb{R})$ establish item (iv). Note that the chain rule implies for all $r \in [1,\infty)$, $x \in (0, \mathscr{B}r^{-1}]$ that

$$|(\mathscr{R}_r)'(x)| = \big|\eta(\tfrac{rx-\mathscr{A}}{\mathscr{B}-\mathscr{A}}) + x\big[\tfrac{r}{\mathscr{B}-\mathscr{A}}\big]\eta'(\tfrac{rx-\mathscr{A}}{\mathscr{B}-\mathscr{A}})\big| \leqslant 1 + \big[\tfrac{\mathscr{B}}{\mathscr{B}-\mathscr{A}}\big]\big[\sup_{y\in\mathbb{R}}|\eta'(y)|\big]. \tag{4.15}$$

Combining this with items (i) and (ii) proves that for all $r \in [1,\infty)$, $x \in \mathbb{R}$ it holds that

$$\begin{aligned} |(\mathscr{R}_r)'(x)| &\leqslant \max\{0, 1, 1 + \big[\tfrac{\mathscr{B}}{\mathscr{B}-\mathscr{A}}\big]\big[\sup_{y\in\mathbb{R}}|\eta'(y)|\big]\} \\ &\leqslant 1 + \big[\tfrac{\mathscr{B}}{\mathscr{B}-\mathscr{A}}\big]\big[\sup_{y\in\mathbb{R}}|\eta'(y)|\big] < \infty. \end{aligned} \tag{4.16}$$

This establishes item (v). The proof of Lemma 4.1.4 is thus complete. □

The next lemma establishes a successful approximation of realization functions in the setting of deep ReLU ANNs. Proposition 4.1.5 is a generalization of [63, Proposition 2.2].

Proposition 4.1.5. *Assume Setting 4.1.1, for every $r \in [1,\infty]$, $k \in \mathbb{N}$ let $\mathcal{M}_{r,k}\colon \mathbb{R}^{\ell_k} \to \mathbb{R}^{\ell_k}$ satisfy for all $x \in \mathbb{R}^{\ell_k}$ that $\mathcal{M}_{r,k}(x) = \mathfrak{M}_r(x)$, and let $k \in \{1,\dots,L\}$, $\theta \in \mathbb{R}^{\mathfrak{d}}$. Then*

(i) *it holds for all $r \in [1,\infty]$ that*

$$\mathcal{N}_r^{k,\theta} = \begin{cases} \mathcal{A}_k^\theta & : k = 1 \\ \mathcal{A}_k^\theta \circ \mathcal{M}_{r^{1/(k-1)},k-1} \circ \mathcal{A}_{k-1}^\theta \circ \ldots \circ \mathcal{M}_{r^{1/2},1} \circ \mathcal{A}_2^\theta \circ \mathcal{M}_{r,1} \circ \mathcal{A}_1^\theta & : k > 1, \end{cases} \tag{4.17}$$

(ii) *it holds that*

$$\begin{aligned} &\sup_{r\in[1,\infty)}\sup_{x\in\mathbb{R}^{\ell_0}}\max_{i\in\{1,\dots,\ell_k\}}\big(r^{1/k}|\mathscr{R}_{r^{1/k}}(\mathcal{N}_{r,i}^{k,\theta}(x)) - \mathscr{R}_\infty(\mathcal{N}_{\infty,i}^{k,\theta}(x))|\big) \\ &\leqslant \mathscr{B}\big[\textstyle\sum_{j=0}^{k-1}\big(\sup_{r\in[1,\infty)}\sup_{x\in\mathbb{R}}|(\mathscr{R}_r)'(x)|\big)^j\big(\sum_{\mathfrak{v}=1}^{\mathfrak{d}}|\theta_{\mathfrak{v}}|\big)^j\big] < \infty, \end{aligned} \tag{4.18}$$

(iii) *it holds that*

$$\begin{aligned} &\sup_{r\in[1,\infty)}\sup_{x\in\mathbb{R}^{\ell_0}}\max_{i\in\{1,\dots,\ell_k\}}\big(r^{1/(\max\{k-1,1\})}|\mathcal{N}_{r,i}^{k,\theta}(x) - \mathcal{N}_{\infty,i}^{k,\theta}(x)|\big) \\ &\leqslant \mathscr{B}\big[\textstyle\sum_{j=1}^{k}\big(\sup_{r\in[1,\infty)}\sup_{x\in\mathbb{R}}|(\mathscr{R}_r)'(x)|\big)^{j-1}\big(\sum_{\mathfrak{v}=1}^{\mathfrak{d}}|\theta_{\mathfrak{v}}|\big)^j\big]\mathbb{1}_{(1,\infty)}(k) < \infty, \end{aligned} \tag{4.19}$$

(iv) *it holds for all $i \in \{1,\dots,\ell_k\}$, $x \in \mathbb{R}^{\ell_0}$ that*

$$\limsup_{r\to\infty}\big(|\mathcal{N}_{r,i}^{k,\theta}(x) - \mathcal{N}_{\infty,i}^{k,\theta}(x)| + |\mathscr{R}_{r^{1/k}}(\mathcal{N}_{r,i}^{k,\theta}(x)) - \mathscr{R}_\infty(\mathcal{N}_{\infty,i}^{k,\theta}(x))|\big) = 0, \tag{4.20}$$

and

(v) *it holds for all $i \in \{1,\dots,\ell_k\}$, $x \in \mathbb{R}^{\ell_0}$ that*

$$\limsup_{r\to\infty}|(\mathscr{R}_{r^{1/k}})'(\mathcal{N}_{r,i}^{k,\theta}(x)) - \mathbb{1}_{\mathcal{X}_i^{k,\theta}}(x)| = 0. \tag{4.21}$$

Proof of Proposition 4.1.5. Throughout this proof let $\mathbf{L} \in \mathbb{R}$ satisfy

$$\mathbf{L} = \sup_{r\in[1,\infty)} \sup_{x\in\mathbb{R}} |(\mathscr{R}_r)'(x)|, \tag{4.22}$$

let $c \in \mathbb{R}$ satisfy $c = \mathbf{L}[\sum_{j=1}^{\mathfrak{d}} |\theta_j|]$, and for every $K \in \{1, \dots, L\}$, $r \in [1, \infty]$ let $e_{K,r} \in \mathbb{R}$ satisfy

$$e_{K,r} = \sup_{x\in\mathbb{R}^{\ell_0}} \max_{i\in\{1,\dots,\ell_K\}} |\mathscr{R}_{r^{1/K}}(\mathcal{N}^{K,\theta}_{r,i}(x)) - \mathscr{R}_\infty(\mathcal{N}^{K,\theta}_{\infty,i}(x))|. \tag{4.23}$$

Observe that (4.22) and the fundamental theorem of calculus ensure that for all $r \in [1, \infty)$, $x, y \in \mathbb{R}$ it holds that

$$|\mathscr{R}_r(x) - \mathscr{R}_r(y)| \leqslant \mathbf{L}|x - y|. \tag{4.24}$$

Furthermore, note that (4.5) establishes item (i). Observe that Lemma 4.1.3 assures that for all $r \in [1, \infty)$ it holds that

$$\sup_{y\in\mathbb{R}} |\mathscr{R}_r(y) - \mathscr{R}_\infty(y)| \leqslant \mathscr{B}r^{-1}. \tag{4.25}$$

Item (ii) of Lemma 4.1.2 therefore ensures that for all $r \in [1, \infty)$, $i \in \{1, \dots, \ell_1\}$, $x \in \mathbb{R}^{\ell_0}$ it holds that

$$|\mathscr{R}_r(\mathcal{N}^{1,\theta}_{r,i}(x)) - \mathscr{R}_\infty(\mathcal{N}^{1,\theta}_{\infty,i}(x))| = |\mathscr{R}_r(\mathcal{N}^{1,\theta}_{\infty,i}(x)) - \mathscr{R}_\infty(\mathcal{N}^{1,\theta}_{\infty,i}(x))| \leqslant \mathscr{B}r^{-1}. \tag{4.26}$$

Combining this with (4.23) proves for all $r \in [1, \infty)$ that

$$e_{1,r} \leqslant \mathscr{B}r^{-1}. \tag{4.27}$$

Moreover, note that item (iii) of Lemma 4.1.2 and (4.24) ensure that for all $r \in [1, \infty)$, $K \in \mathbb{N} \cap [1, L)$, $i \in \{1, \dots, \ell_{K+1}\}$, $x \in \mathbb{R}^{\ell_0}$ it holds that

$$\begin{aligned}
&\big|\mathscr{R}_{r^{1/(K+1)}}(\mathcal{N}^{K+1,\theta}_{r,i}(x)) - \mathscr{R}_{r^{1/(K+1)}}(\mathcal{N}^{K+1,\theta}_{\infty,i}(x))\big| \leqslant \mathbf{L}|\mathcal{N}^{K+1,\theta}_{r,i}(x) - \mathcal{N}^{K+1,\theta}_{\infty,i}(x)| \\
&= \mathbf{L}\big|\textstyle\sum_{j=1}^{\ell_K} \mathfrak{w}^{K+1,\theta}_{i,j}\big(\mathscr{R}_{r^{1/K}}(\mathcal{N}^{K,\theta}_{r,j}(x)) - \mathscr{R}_\infty(\mathcal{N}^{K,\theta}_{\infty,j}(x))\big)\big| \\
&\leqslant \mathbf{L}\big[\textstyle\sum_{j=1}^{\ell_K} |\mathfrak{w}^{K+1,\theta}_{i,j}||\mathscr{R}_{r^{1/K}}(\mathcal{N}^{K,\theta}_{r,j}(x)) - \mathscr{R}_\infty(\mathcal{N}^{K,\theta}_{\infty,j}(x))|\big] \\
&\leqslant \mathbf{L}\big[\textstyle\sum_{j=1}^{\ell_K} |\mathfrak{w}^{K+1,\theta}_{i,j}|\big]\big[\max_{j\in\{1,2,\dots,\ell_K\}} |\mathscr{R}_{r^{1/K}}(\mathcal{N}^{K,\theta}_{r,j}(x)) - \mathscr{R}_\infty(\mathcal{N}^{K,\theta}_{\infty,j}(x))|\big] \\
&\leqslant \mathbf{L}\big[\textstyle\sum_{j=1}^{\mathfrak{d}} |\theta_j|\big]\big[\max_{j\in\{1,2,\dots,\ell_K\}} |\mathscr{R}_{r^{1/K}}(\mathcal{N}^{K,\theta}_{r,j}(x)) - \mathscr{R}_\infty(\mathcal{N}^{K,\theta}_{\infty,j}(x))|\big] \\
&= \mathbf{L}\Big[\sum_{j=1}^{\mathfrak{d}} |\theta_j|\Big]e_{K,r}.
\end{aligned} \tag{4.28}$$

Combining this with (4.25) ensures that for all $r \in [1, \infty)$, $K \in \mathbb{N} \cap [1, L)$, $i \in \{1, \dots, \ell_{K+1}\}$, $x \in \mathbb{R}^{\ell_0}$ it holds that

$$\begin{aligned}
&\big|\mathscr{R}_{r^{1/(K+1)}}(\mathcal{N}^{K+1,\theta}_{r,i}(x)) - \mathscr{R}_\infty(\mathcal{N}^{K+1,\theta}_{\infty,i}(x))\big| \\
&\leqslant \big|\mathscr{R}_{r^{1/(K+1)}}(\mathcal{N}^{K+1,\theta}_{r,i}(x)) - \mathscr{R}_{r^{1/(K+1)}}(\mathcal{N}^{K+1,\theta}_{\infty,i}(x))\big| \\
&\quad + \big|\mathscr{R}_{r^{1/(K+1)}}(\mathcal{N}^{K+1,\theta}_{\infty,i}(x)) - \mathscr{R}_\infty(\mathcal{N}^{K+1,\theta}_{\infty,i}(x))\big| \\
&\leqslant \mathbf{L}\big[\textstyle\sum_{j=1}^{\mathfrak{d}} |\theta_j|\big]e_{K,r} + \mathscr{B}r^{-1/(K+1)} = e_{K,r}c + \mathscr{B}r^{-1/(K+1)}.
\end{aligned} \tag{4.29}$$

Hence, we obtain for all $r \in [1, \infty)$, $K \in \mathbb{N} \cap [1, L)$ that $e_{K+1,r} \leqslant e_{K,r}c + \mathscr{B}r^{-1/(K+1)}$. This shows for all $r \in [1, \infty)$, $K \in \mathbb{N} \cap (1, L]$ that

$$\begin{aligned}
&e_{K,r} \\
&\leqslant c\big[e_{K-1,r}\big] + \mathscr{B}r^{-1/K} \leqslant c\big[e_{K-2,r}c + \mathscr{B}r^{-1/(K-1)}\big] + \mathscr{B}r^{-1/K} \\
&= c^2 e_{K-2,r} + \textstyle\sum_{j=0}^{1} \mathscr{B}c^j r^{-1/(K-j)} \leqslant \ldots \leqslant c^{K-1}e_{K-(K-1),r} + \textstyle\sum_{j=0}^{K-2} \mathscr{B}c^j r^{-1/(K-j)} \\
&= c^{K-1}e_{1,r} + \textstyle\sum_{j=0}^{K-2} \mathscr{B}c^j r^{-1/(K-j)}.
\end{aligned} \tag{4.30}$$

Combining this with (4.27) demonstrates for all $r \in [1,\infty)$, $K \in \mathbb{N} \cap [1,L]$ that

$$\begin{aligned} e_{K,r} &\leqslant c^{K-1} e_{1,r} + \textstyle\sum_{j=0}^{K-2} \mathscr{B} c^j r^{-1/(K-j)} \leqslant c^{K-1} \mathscr{B} r^{-1} + \textstyle\sum_{j=0}^{K-2} \mathscr{B} c^j r^{-1/(K-j)} \\ &= \textstyle\sum_{j=0}^{K-1} \mathscr{B} c^j r^{-1/(K-j)} \leqslant \big[\textstyle\sum_{j=0}^{K-1} \mathscr{B} c^j\big] r^{\max\{-1/K,-1/(K-1),\dots,-1\}} \\ &= \Big[\sum_{j=0}^{K-1} c^j\Big] \mathscr{B} r^{-1/K}. \end{aligned} \tag{4.31}$$

This establishes item (ii). Observe that (4.2), (4.5) and (4.23) ensure that for all $r \in [1,\infty)$, $K \in \mathbb{N} \cap (1,L]$, $i \in \{1,2,\dots,\ell_K\}$, $x \in \mathbb{R}^{\ell_0}$ it holds that

$$\begin{aligned} &|\mathcal{N}^{K,\theta}_{r,i}(x) - \mathcal{N}^{K,\theta}_{\infty,i}(x)| = \big|\textstyle\sum_{j=1}^{\ell_{K-1}} \mathfrak{w}^{K,\theta}_{i,j} \big(\mathscr{R}_{r^{1/(K-1)}}(\mathcal{N}^{K-1,\theta}_{r,j}(x)) - \mathscr{R}_\infty(\mathcal{N}^{K-1,\theta}_{\infty,j}(x))\big)\big| \\ &\leqslant \textstyle\sum_{j=1}^{\ell_{K-1}} \big(|\mathfrak{w}^{K,\theta}_{i,j}| |\mathscr{R}_{r^{1/(K-1)}}(\mathcal{N}^{K-1,\theta}_{r,j}(x)) - \mathscr{R}_\infty(\mathcal{N}^{K-1,\theta}_{\infty,j}(x))|\big) \\ &\leqslant \big(\textstyle\sum_{j=1}^{\ell_{K-1}} |\mathfrak{w}^{K,\theta}_{i,j}|\big) \big(\max_{j\in\{1,2,\dots,\ell_{K-1}\}} |\mathscr{R}_{r^{1/(K-1)}}(\mathcal{N}^{K-1,\theta}_{r,j}(x)) - \mathscr{R}_\infty(\mathcal{N}^{K-1,\theta}_{\infty,j}(x))|\big) \\ &\leqslant \big(\textstyle\sum_{j=1}^{\mathfrak{d}} |\theta_j|\big)\, e_{K-1,r}. \end{aligned} \tag{4.32}$$

Therefore, we obtain for all $K \in \mathbb{N} \cap (1,L]$ that

$$\begin{aligned} &\textstyle\sup_{r\in[1,\infty)} \sup_{x\in\mathbb{R}^{\ell_0}} \max_{i\in\{1,2,\dots,\ell_K\}} \big(r^{1/(\max\{K-1,1\})} |\mathcal{N}^{K,\theta}_{r,i}(x) - \mathcal{N}^{K,\theta}_{\infty,i}(x)|\big) \\ &\leqslant \big(\textstyle\sum_{j=1}^{\mathfrak{d}} |\theta_j|\big) \big[\sup_{r\in[1,\infty)} \big(r^{1/(\max\{K-1,1\})} e_{K-1,r}\big)\big] \\ &= \Big(\sum_{j=1}^{\mathfrak{d}} |\theta_j|\Big) \Big[\sup_{r\in[1,\infty)} \big(r^{1/(K-1)} e_{K-1,r}\big)\Big]. \end{aligned} \tag{4.33}$$

Combining this with (4.5), (4.23), and item (ii) establishes item (iii). Note that items (ii) and (iii) prove item (iv). It thus remains to prove item (v). For this observe that item (iv) assures for all $i \in \{1,\dots,\ell_k\}$, $x \in \mathbb{R}^{\ell_0}$ with $\mathcal{N}^{k,\theta}_{\infty,i}(x) < 0$ that there exists $R \in [1,\infty)$ such that for all $r \in [R,\infty)$ it holds that $\mathcal{N}^{k,\theta}_{r,i}(x) < 0$. Combining this with (4.3) demonstrates for all $i \in \{1,\dots,\ell_k\}$, $x \in \mathbb{R}^{\ell_0}$ with $\mathcal{N}^{k,\theta}_{\infty,i}(x) < 0$ that there exists $R \in [1,\infty)$ such that for all $r \in [R,\infty)$ it holds that

$$(\mathscr{R}_{r^{1/k}})'(\mathcal{N}^{k,\theta}_{r,i}(x)) = 0 = \mathbb{1}_{(0,\infty)}\big(\mathcal{N}^{k,\theta}_{\infty,i}(x)\big) = \mathbb{1}_{\mathcal{X}^{k,\theta}_i}(x). \tag{4.34}$$

In addition, note that item (iv) shows for all $i \in \{1,\dots,\ell_k\}$, $x \in \mathbb{R}^{\ell_0}$ with $\mathcal{N}^{k,\theta}_{\infty,i}(x) > 0$ that there exists $R \in [1,\infty)$ such that for all $r \in [R,\infty)$ it holds that $\mathcal{N}^{k,\theta}_{r,i}(x) > \mathscr{B} r^{-1/k}$. Combining this with (4.3) demonstrates for all $i \in \{1,\dots,\ell_k\}$, $x \in \mathbb{R}^{\ell_0}$ with $\mathcal{N}^{k,\theta}_{\infty,i}(x) > 0$ that there exists $R \in [1,\infty)$ such that for all $r \in [R,\infty)$ it holds that

$$(\mathscr{R}_{r^{1/k}})'(\mathcal{N}^{k,\theta}_{r,i}(x)) = 1 = \mathbb{1}_{(0,\infty)}\big(\mathcal{N}^{k,\theta}_{\infty,i}(x)\big) = \mathbb{1}_{\mathcal{X}^{k,\theta}_i}(x). \tag{4.35}$$

Furthermore, observe that item (iii) assures that for all $i \in \{1,2,\dots,\ell_k\}$, $x \in \mathbb{R}^{\ell_0}$ it holds that

$$\begin{aligned} &\textstyle\sup_{r\in[1,\infty)} \big(r^{(1+\mathbb{1}_{\{1\}}(k))/(k-\mathbb{1}_{(1,\infty)}(k))} |\mathcal{N}^{k,\theta}_{r,i}(x) - \mathcal{N}^{k,\theta}_{\infty,i}(x)|\big) \\ &= \textstyle\sup_{r\in[1,\infty)} \big(r^{(1+\mathbb{1}_{\{1\}}(k))/(\max\{k-1,1\})} |\mathcal{N}^{k,\theta}_{r,i}(x) - \mathcal{N}^{k,\theta}_{\infty,i}(x)|\big) < \infty. \end{aligned} \tag{4.36}$$

This and the fact that $-(1+\mathbb{1}_{\{1\}}(k))/(k-\mathbb{1}_{(1,\infty)}(k)) < -1/k$ assure for all $i \in \{1,2,\dots,\ell_k\}$, $x \in \mathbb{R}^{\ell_0}$ with $\mathcal{N}^{k,\theta}_{\infty,i}(x) = 0$ that there exist $\mathfrak{C}, R \in [1,\infty)$ such that for all $r \in [R,\infty)$ it holds that

$$|\mathcal{N}^{k,\theta}_{r,i}(x)| = |\mathcal{N}^{k,\theta}_{r,i}(x) - \mathcal{N}^{k,\theta}_{\infty,i}(x)| \leqslant \mathfrak{C}\big[r^{-(1+\mathbb{1}_{\{1\}}(k))/(k-\mathbb{1}_{(1,\infty)}(k))}\big] < \mathscr{A} r^{-1/k}. \tag{4.37}$$

Moreover, note that (4.3) assures that for all $r \in [1,\infty)$, $y \in \mathbb{R}$ with $|y| < \mathscr{A} r^{-1/k}$ it holds that $(\mathscr{R}_{r^{-1/k}})'(y) = 0$. Combining this with (4.37) proves for all $i \in \{1,2,\dots,\ell_k\}$, $x \in \mathbb{R}^{\ell_0}$ with $\mathcal{N}^{k,\theta}_{\infty,i}(x) = 0$ that there exists $R \in [1,\infty)$ such that for all $r \in [R,\infty)$ it holds that

$$(\mathscr{R}_{r^{-1/k}})'(\mathcal{N}^{k,\theta}_{r,i}(x)) = 0 = \mathbb{1}_{(0,\infty)}\big(\mathcal{N}^{k,\theta}_{\infty,i}(x)\big) = \mathbb{1}_{\mathcal{X}^{k,\theta}_i}(x). \tag{4.38}$$

Combining this, (4.34), and (4.35) establishes item (v). The proof of Proposition 4.1.5 is thus complete. □

The next step is to calculate the gradients of the approximation of the realization function. The gradients are presented in the following lemma.

Lemma 4.1.6. *Assume Setting 4.1.1, for every $k \in \mathbb{N}_0$ let $\mathbf{d}_k \in \mathbb{N}_0$ satisfy $\mathbf{d}_k = \sum_{n=1}^k \ell_n(\ell_{n-1}+1)$, and let $x = (x_1,\dots,x_{\ell_0}) \in [a,b]^{\ell_0}$, $r \in [1,\infty)$. Then*

(i) it holds for all $k \in \{1,\dots,L\}$, $i \in \{1,\dots,\ell_k\}$ that $\mathbb{R}^{\mathfrak{d}} \ni \theta \mapsto \mathcal{N}^{k,\theta}_{r,i}(x) \in \mathbb{R}^{\mathfrak{d}}$ is differentiable,

(ii) it holds for all $K \in \{1,\dots,L\}$, $k \in \{1,\dots,K\}$, $i \in \{1,\dots,\ell_k\}$, $j \in \{1,\dots,\ell_{k-1}\}$, $h \in \{1,\dots,\ell_K\}$, $\theta = (\theta_1,\dots,\theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ that

$$\begin{aligned}&\frac{\partial}{\partial\theta_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}}\big(\mathcal{N}^{K,\theta}_{r,h}(x)\big)\\ &= \sum_{\substack{v_k,v_{k+1},\dots,v_K\in\mathbb{N},\\ \forall w\in\mathbb{N}\cap[k,K]\colon v_w\le\ell_w}} \Big[\mathscr{R}_{r^{1/(\max\{k-1,1\})}}(\mathcal{N}^{\max\{k-1,1\},\theta}_{r,j}(x))\mathbb{1}_{(1,K]}(k) + x_j\mathbb{1}_{\{1\}}(k)\Big]\\ &\quad\cdot\Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathbb{1}_{\{h\}}(v_K)\Big]\Big[\textstyle\prod_{n=k+1}^K\big(\mathfrak{w}^{n,\theta}_{v_n,v_{n-1}}\big[(\mathscr{R}_{r^{1/(n-1)}})'(\mathcal{N}^{n-1,\theta}_{r,v_{n-1}}(x))\big]\big)\Big],\end{aligned}\tag{4.39}$$

and

(iii) it holds for all $K \in \{1,\dots,L\}$, $k \in \{1,\dots,K\}$, $i \in \{1,\dots,\ell_k\}$, $h \in \{1,\dots,\ell_K\}$, $\theta = (\theta_1,\dots,\theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ that

$$\begin{aligned}&\frac{\partial}{\partial\theta_{\ell_k\ell_{k-1}+i+\mathbf{d}_{k-1}}}\big(\mathcal{N}^{K,\theta}_{r,h}(x)\big)\\ &= \sum_{\substack{v_k,v_{k+1},\dots,v_K\in\mathbb{N},\\ \forall w\in\mathbb{N}\cap[k,K]\colon v_w\le\ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathbb{1}_{\{h\}}(v_K)\Big]\\ &\quad\cdot\Big[\textstyle\prod_{n=k+1}^K\big(\mathfrak{w}^{n,\theta}_{v_n,v_{n-1}}\big[(\mathscr{R}_{r^{1/(n-1)}})'(\mathcal{N}^{n-1,\theta}_{r,v_{n-1}}(x))\big]\big)\Big].\end{aligned}\tag{4.40}$$

Proof of Lemma 4.1.6. Observe that items (ii) and (iii) of Lemma 4.1.2 and the assumption that $\mathscr{R}_r \in C^1(\mathbb{R},\mathbb{R})$ establish item (i). We now prove (4.39) and (4.40) by induction on $K \in \{1,\dots,L\}$. Note that item (ii) of Lemma 4.1.2 implies that for all $i \in \{1,\dots,\ell_1\}$, $j \in \{1,\dots,\ell_0\}$, $h \in \{1,\dots,\ell_1\}$, $\theta = (\theta_1,\dots,\theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ it holds that

$$\frac{\partial}{\partial\theta_{(i-1)\ell_0+j}}\big(\mathcal{N}^{1,\theta}_{r,h}(x)\big) = x_j\mathbb{1}_{\{h\}}(i) \qquad\text{and}\qquad \frac{\partial}{\partial\theta_{\ell_0\ell_1+i}}\big(\mathcal{N}^{1,\theta}_{r,h}(x)\big) = \mathbb{1}_{\{h\}}(i). \tag{4.41}$$

This establishes (4.39) and (4.40) in the base case $K=1$. For the induction step let $K \in \mathbb{N}\cap[1,L)$ satisfy for all $k \in \{1,\dots,K\}$, $i \in \{1,\dots,\ell_k\}$, $j \in \{1,\dots,\ell_{k-1}\}$, $h \in \{1,\dots,\ell_K\}$,

$\theta = (\theta_1, \ldots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ that

$$\frac{\partial}{\partial\theta_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}}\big(\mathcal{N}^{K,\theta}_{r,h}(x)\big) = \sum_{\substack{v_k,v_{k+1},\ldots,v_K\in\mathbb{N},\\ \forall w\in\mathbb{N}\cap[k,K]\colon v_w\leqslant\ell_w}} \Big[\mathscr{R}_{r^{1/(\max\{k-1,1\})}}\big(\mathcal{N}^{\max\{k-1,1\},\theta}_{r,j}(x)\big)\mathbb{1}_{(1,K]}(k) + x_j\mathbb{1}_{\{1\}}(k)\Big] \cdot \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathbb{1}_{\{h\}}(v_K)\Big]\Big[\textstyle\prod_{n=k+1}^{K}\big(\mathfrak{w}^{n,\theta}_{v_n,v_{n-1}}\big[(\mathscr{R}_{r^{1/(n-1)}})'(\mathcal{N}^{n-1,\theta}_{r,v_{n-1}}(x))\big]\big)\Big] \tag{4.42}$$

and

$$\frac{\partial}{\partial\theta_{\ell_k\ell_{k-1}+i+\mathbf{d}_{k-1}}}\big(\mathcal{N}^{K,\theta}_{r,h}(x)\big) = \sum_{\substack{v_k,v_{k+1},\ldots,v_K\in\mathbb{N},\\ \forall w\in\mathbb{N}\cap[k,K]\colon v_w\leqslant\ell_w)}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathbb{1}_{\{h\}}(v_K)\Big] \cdot \Big[\textstyle\prod_{n=k+1}^{K}\big(\mathfrak{w}^{n,\theta}_{v_n,v_{n-1}}\big[(\mathscr{R}_{r^{1/(n-1)}})'(\mathcal{N}^{n-1,\theta}_{r,v_{n-1}}(x))\big]\big)\Big]. \tag{4.43}$$

Observe that item (iii) of Lemma 4.1.2 and (4.42) demonstrate that for all $k \in \{1, \ldots, K\}$, $i \in \{1, \ldots, \ell_k\}$, $j \in \{1, \ldots, \ell_{k-1}\}$, $h \in \{1, \ldots, \ell_{K+1}\}$, $\theta = (\theta_1, \ldots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ it holds that

$$\begin{aligned}
&\frac{\partial}{\partial\theta_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}}\big(\mathcal{N}^{K+1,\theta}_{r,h}(x)\big) \\
&= \frac{\partial}{\partial\theta_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}}\bigg(\mathfrak{b}^{K+1,\theta}_h + \sum_{i=1}^{\ell_K}\mathfrak{w}^{K+1,\theta}_{h,i}\mathscr{R}_{r^{1/K}}\big(\mathcal{N}^{K,\theta}_{r,i}(x)\big)\bigg) \\
&= \sum_{i=1}^{\ell_K}\bigg[\mathfrak{w}^{K+1,\theta}_{h,i}\big[(\mathscr{R}_{r^{1/K}})'(\mathcal{N}^{K,\theta}_{r,i}(x))\big]\bigg(\frac{\partial}{\partial\theta_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}}\big(\mathcal{N}^{K,\theta}_{r,i}(x)\big)\bigg)\bigg] \\
&= \sum_{i=1}^{\ell_K}\bigg[\mathfrak{w}^{K+1,\theta}_{h,i}\big[(\mathscr{R}_{r^{1/K}})'(\mathcal{N}^{K,\theta}_{r,i}(x))\big] \\
&\quad\cdot \sum_{\substack{v_k,v_{k+1},\ldots,v_K\in\mathbb{N},\\ \forall w\in\mathbb{N}\cap[k,K]\colon v_w\leqslant\ell_w}} \Big[\mathscr{R}_{r^{1/(\max\{k-1,1\})}}\big(\mathcal{N}^{\max\{k-1,1\},\theta}_{r,j}(x)\big)\mathbb{1}_{(1,K]}(k) + x_j\mathbb{1}_{\{1\}}(k)\Big] \\
&\quad\cdot \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathbb{1}_{\{i\}}(v_K)\Big]\Big[\textstyle\prod_{n=k+1}^{K}\big(\mathfrak{w}^{n,\theta}_{v_n,v_{n-1}}\big[(\mathscr{R}_{r^{1/(n-1)}})'(\mathcal{N}^{n-1,\theta}_{r,v_{n-1}}(x))\big]\big)\Big]\bigg] \\
&= \sum_{\substack{v_k,v_{k+1},\ldots,v_{K+1}\in\mathbb{N},\\ \forall w\in\mathbb{N}\cap[k,K+1]\colon v_w\leqslant\ell_w}} \Big[\mathscr{R}_{r^{1/(\max\{k-1,1\})}}\big(\mathcal{N}^{\max\{k-1,1\},\theta}_{r,j}(x)\big)\mathbb{1}_{(1,K]}(k) + x_j\mathbb{1}_{\{1\}}(k)\Big] \\
&\quad\cdot \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathbb{1}_{\{h\}}(v_{K+1})\Big]\Big[\textstyle\prod_{n=k+1}^{K+1}\big(\mathfrak{w}^{n,\theta}_{v_n,v_{n-1}}\big[(\mathscr{R}_{r^{1/(n-1)}})'(\mathcal{N}^{n-1,\theta}_{r,v_{n-1}}(x))\big]\big)\Big]\Big].
\end{aligned} \tag{4.44}$$

Furthermore, note that item (iii) of Lemma 4.1.2 ensures that for all $i \in \{1, \ldots, \ell_{K+1}\}$, $j \in \{1, \ldots, \ell_K\}$, $h \in \{1, \ldots, \ell_{K+1}\}$, $\theta = (\theta_1, \ldots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ it holds that

$$\begin{aligned}
\frac{\partial}{\partial\theta_{(i-1)\ell_K+j+\mathbf{d}_K}}\big(\mathcal{N}^{K+1,\theta}_{r,h}(x)\big) &= \frac{\partial}{\partial\theta_{(i-1)\ell_K+j+\mathbf{d}_K}}\bigg(\mathfrak{b}^{K+1,\theta}_h + \sum_{i=1}^{\ell_K}\mathfrak{w}^{K+1,\theta}_{h,i}\mathscr{R}_{r^{1/K}}\big(\mathcal{N}^{K,\theta}_{r,i}(x)\big)\bigg) \\
&= \mathscr{R}_{r^{1/K}}\big(\mathcal{N}^{K,\theta}_{r,j}(x)\big)\mathbb{1}_{\{h\}}(i).
\end{aligned} \tag{4.45}$$

Moreover, observe that item (iii) of Lemma 4.1.2 and (4.43) demonstrate that for all $k \in \{1, \ldots, K\}$, $i \in \{1, \ldots, \ell_k\}$, $h \in \{1, \ldots, \ell_{K+1}\}$, $\theta = (\theta_1, \ldots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ it holds that

$$\begin{aligned}
&\frac{\partial}{\partial \theta_{\ell_k \ell_{k-1}+i+\mathbf{d}_{k-1}}} \big(\mathcal{N}_{r,h}^{K+1,\theta}(x)\big) = \frac{\partial}{\partial \theta_{\ell_k \ell_{k-1}+i+\mathbf{d}_{k-1}}} \Bigg(\mathfrak{b}_h^{K+1,\theta} + \sum_{\imath=1}^{\ell_K} \mathfrak{w}_{h,\imath}^{K+1,\theta} \mathscr{R}_{r^{1/K}}\big(\mathcal{N}_{r,\imath}^{K,\theta}(x)\big)\Bigg) \\
&= \sum_{\imath=1}^{\ell_K} \mathfrak{w}_{h,\imath}^{K+1,\theta} \big[(\mathscr{R}_{r^{1/K}})'\big(\mathcal{N}_{r,\imath}^{K,\theta}(x)\big)\big] \Bigg(\frac{\partial}{\partial \theta_{\ell_k \ell_{k-1}+i+\mathbf{d}_{k-1}}} \big(\mathcal{N}_{r,\imath}^{K,\theta}(x)\big)\Bigg) \\
&= \sum_{i_K=1}^{\ell_K} \Bigg[\mathfrak{w}_{h,\imath}^{K+1,\theta} \big[(\mathscr{R}_{r^{1/K}})'\big(\mathcal{N}_{r,\imath}^{K,\theta}(x)\big)\big] \sum_{\substack{v_k, v_{k+1}, \ldots, v_K \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,K] \colon v_w \leq \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big] \Big[\mathbb{1}_{\{\imath\}}(v_K)\Big] \\
&\quad \cdot \Big[\textstyle\prod_{n=k+1}^{K} \Big(\mathfrak{w}_{v_n,v_{n-1}}^{n,\theta} \big[(\mathscr{R}_{r^{1/(n-1)}})'\big(\mathcal{N}_{r,v_{n-1}}^{n-1,\theta}(x)\big)\big]\Big)\Big] \Bigg] \\
&= \sum_{\substack{v_k, v_{k+1}, \ldots, v_{K+1} \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,K+1] \colon v_w \leq \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big] \Big[\mathbb{1}_{\{h\}}(v_{K+1})\Big] \Big[\textstyle\prod_{n=k+1}^{K+1} \mathfrak{w}_{v_n,v_{n-1}}^{n,\theta} \big[(\mathscr{R}_{r^{1/(n-1)}})'\big(\mathcal{N}_{r,v_{n-1}}^{n-1,\theta}(x)\big)\big]\Big].
\end{aligned} \tag{4.46}$$

In addition, note that item (iii) of Lemma 4.1.2 shows for all $i \in \{1, \ldots, \ell_{K+1}\}$, $h \in \{1, \ldots, \ell_{K+1}\}$, $\theta = (\theta_1, \ldots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ that

$$\begin{aligned}
&\frac{\partial}{\partial \theta_{\ell_{K+1}\ell_K+i+\mathbf{d}_K}} \big(\mathcal{N}_{r,h}^{K+1,\theta}(x)\big) \\
&= \frac{\partial}{\partial \theta_{\ell_{K+1}\ell_K+i+\mathbf{d}_K}} \Bigg(\mathfrak{b}_h^{K+1,\theta} + \sum_{\imath=1}^{\ell_K} \mathfrak{w}_{h,\imath}^{K+1,\theta} \mathscr{R}_{r^{1/K}}\big(\mathcal{N}_{r,\imath}^{K,\theta}(x)\big)\Bigg) \\
&= \mathbb{1}_{\{h\}}(i).
\end{aligned} \tag{4.47}$$

Induction thus establishes (4.39) and (4.40). The proof of Lemma 4.1.6 is thus complete. □

4.1.3 Explicit representations for the generalized gradients of the risk function

The following lemma provides an a-priori bound for the introduced approximations of the realization function and its gradients.

Lemma 4.1.7. *Assume Setting 4.1.1 and let $K \subseteq \mathbb{R}^{\mathfrak{d}}$ be compact. Then*

(i) it holds for all $x \in [0, \infty)$ that

$$\sup_{r \in [1,\infty)} \sup_{y \in [-x,x]} \big(|\mathscr{R}_r(y)| + |(\mathscr{R}_r)'(y)|\big) < \infty, \tag{4.48}$$

(ii) it holds for all $k \in \{1, \ldots, L\}$ that

$$\sup_{\theta \in K} \sup_{r \in [1,\infty)} \sup_{i \in \{1,\ldots,\ell_k\}} \sup_{x \in [a,b]^{\ell_0}} |\mathcal{N}_{r,i}^{k,\theta}(x)| < \infty, \tag{4.49}$$

(iii) it holds for all $k \in \{1, \ldots, L\}$ that

$$\begin{aligned}
\sup_{\theta \in K} \sup_{r,s,t \in [1,\infty)} \sup_{i \in \{1,\ldots,\ell_k\}} \sup_{x \in [a,b]^{\ell_0}} \big(|\mathscr{R}_s(\mathcal{N}_{r,i}^{k,\theta}(x))| \\
+ |(\mathscr{R}_t)'(\mathcal{N}_{r,i}^{k,\theta}(x))|\big) < \infty,
\end{aligned} \tag{4.50}$$

and

(iv) it holds that

$$\sup_{\theta\in K}\sup_{r\in[1,\infty)}\sup_{i\in\{1,\ldots,\ell_L\}}\sup_{j\in\{1,\ldots,\mathfrak{d}\}}\sup_{x\in[a,b]^{\ell_0}}\left|\tfrac{\partial}{\partial\theta_j}\big(\mathcal{N}^{L,\theta}_{r,i}(x)\big)\right| < \infty. \tag{4.51}$$

Proof of Lemma 4.1.7. Observe that the fundamental theorem of calculus and the assumption that for all $r \in [1,\infty)$ it holds that $\mathscr{R}_r \in C^1(\mathbb{R},\mathbb{R})$ ensure that for all $r \in [1,\infty)$, $x \in \mathbb{R}$ we have that $\mathscr{R}_r(x) = \mathscr{R}_r(0) + \int_0^x (\mathscr{R}_r)'(y)\,\mathrm{d}y = \int_0^x (\mathscr{R}_r)'(y)\,\mathrm{d}y$. Combining this with the assumption that $\sup_{r\in[1,\infty)}\sup_{y\in\mathbb{R}}|(\mathscr{R}_r)'(y)| < \infty$ shows that for all $x \in [0,\infty)$ it holds that $\sup_{r\in[1,\infty)}\sup_{y\in[-x,x]}|\mathscr{R}_r(y)| < \infty$. This establishes item (i). We now prove (4.49) by induction on $k \in \{1,\ldots,L\}$. Note that the assumption that K is compact ensures that there exists $\mathfrak{C} \in \mathbb{R}$ which satisfies that

$$\sup_{\theta\in K}\|\theta\| < \mathfrak{C}. \tag{4.52}$$

Observe that (4.52) and item (ii) of Lemma 4.1.2 demonstrate for all $\theta \in K$, $r \in [1,\infty)$, $i \in \{1,\ldots,\ell_1\}$, $x \in [a,b]^{\ell_0}$ that

$$|\mathcal{N}^{1,\theta}_{r,i}(x)| \leq |\mathfrak{b}^{1,\theta}_i| + \textstyle\sum_{j=1}^{\ell_0}|\mathfrak{w}^{1,\theta}_{i,j}||x_j| \leq \mathbf{a}(\ell_0+1)\mathfrak{C}. \tag{4.53}$$

This shows (4.49) in the base case $k = 1$. For the induction step let $k \in \mathbb{N} \cap [1,L)$ satisfy that

$$\sup_{\theta\in K}\sup_{r\in[1,\infty)}\sup_{i\in\{1,\ldots,\ell_k\}}\sup_{x\in[a,b]^{\ell_0}}|\mathcal{N}^{k,\theta}_{r,i}(x)| < \infty. \tag{4.54}$$

Note that (4.52) and item (iii) of Lemma 4.1.2 imply for all $\theta \in K$, $r \in [1,\infty)$, $i \in \{1,\ldots,\ell_{k+1}\}$, $x \in [a,b]^{\ell_0}$ that

$$|\mathcal{N}^{k+1,\theta}_{r,i}(x)| \leq |\mathfrak{b}^{k+1,\theta}_i| + \textstyle\sum_{j=1}^{\ell_k}|\mathfrak{w}^{k+1,\theta}_{i,j}||\mathscr{R}_r(\mathcal{N}^{k,\theta}_{r,j}(x))| \leq \mathfrak{C}\big(1 + \ell_k \max_{j\in\{1,\ldots,\ell_k\}}|\mathscr{R}_r(\mathcal{N}^{k,\theta}_{r,j}(x))|\big). \tag{4.55}$$

This, item (i), and (4.54) demonstrate that

$$\sup_{\theta\in K}\sup_{r\in[1,\infty)}\sup_{i\in\{1,\ldots,\ell_{k+1}\}}\sup_{x\in[a,b]^{\ell_0}}|\mathcal{N}^{k+1,\theta}_{r,i}(x)| < \infty. \tag{4.56}$$

Induction thus establishes (4.49). This completes the proof of item (ii). Observe that items (i) and (ii) prove item (iii). Note that item (iii) and Lemma 4.1.6 establish item (iv). The proof of Lemma 4.1.7 is thus complete. □

The following lemma is an elementary integrability result for the target function.

Lemma 4.1.8. *Assume Setting 4.1.1. Then*

$$\int_{[a,b]^{\ell_0}} \|f(x)\|^2\,\mu(\mathrm{d}x) + \int_{[a,b]^{\ell_0}} \|f(x)\|\,\mu(\mathrm{d}x) < \infty. \tag{4.57}$$

Proof of Lemma 4.1.8. Observe that (4.6), the fact that $\mathcal{L}_\infty(0) \in \mathbb{R}$, and the fact that for all $x \in \mathbb{R}^{\ell_0}$ it holds that $\mathcal{N}^{L,0}_\infty(x) = 0$ assure that

$$\textstyle\int_{[a,b]^{\ell_0}} \|f(x)\|^2\,\mu(\mathrm{d}x) = \mathcal{L}_\infty(0) < \infty. \tag{4.58}$$

Hölder's inequality and the fact that $\mathfrak{m} = \mu([a,b]^{\ell_0}) \in \mathbb{R}$ hence show that

$$\begin{aligned}\textstyle\int_{[a,b]^{\ell_0}} \|f(x)\|\,\mu(\mathrm{d}x) &\leq \Big[\textstyle\int_{[a,b]^{\ell_0}} 1\,\mu(\mathrm{d}x)\Big]^{1/2}\Big[\textstyle\int_{[a,b]^{\ell_0}} \|f(x)\|^2\,\mu(\mathrm{d}x)\Big]^{1/2} \\ &= \sqrt{\mathfrak{m}}\Big[\textstyle\int_{[a,b]^{\ell_0}} \|f(x)\|^2\,\mu(\mathrm{d}x)\Big]^{1/2} < \infty.\end{aligned} \tag{4.59}$$

Combining this with (4.58) establishes (4.57). The proof of Lemma 4.1.8 is thus complete. □

The following theorem establishes an explicit representation for the generalized gradient function of the risk function. Theorem 4.1.11 extends [63, Lemma 2.5] from the case of shallow ANNs to deep ANNs with an arbitrary number of hidden layers.

Theorem 4.1.9. *Assume Setting 4.1.1, for every $k \in \mathbb{N}_0$ let $\mathbf{d}_k \in \mathbb{N}_0$ satisfy $\mathbf{d}_k = \sum_{n=1}^{k} \ell_n(\ell_{n-1}+1)$, and let $\theta = (\theta_1, \dots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$. Then*

(i) it holds for all $r \in [1,\infty)$ that $\mathcal{L}_r \in C^1(\mathbb{R}^{\mathfrak{d}}, \mathbb{R})$

(ii) it holds for all $r \in [1,\infty)$, $k \in \{1,\dots,L\}$, $i \in \{1,\dots,\ell_k\}$, $j \in \{1,\dots,\ell_{k-1}\}$ that

$$\begin{aligned}
&\left(\frac{\partial \mathcal{L}_r}{\partial \theta_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}}\right)(\theta) \\
&= \sum_{\substack{v_k, v_{k+1}, \dots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N}\cap[k,L] \colon v_w \leq \ell_w}} \int_{[a,b]^{\ell_0}} 2\Big[\mathscr{R}_{r^{1/(\max\{k-1,1\})}}(\mathcal{N}_{r,j}^{\max\{k-1,1\},\theta}(x))\mathbb{1}_{(1,L]}(k) \\
&\quad + x_j \mathbb{1}_{\{1\}}(k)\Big]\Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathcal{N}_{r,v_L}^{L,\theta}(x) - f_{v_L}(x)\Big] \\
&\quad \cdot \Big[\textstyle\prod_{n=k+1}^{L}\big(\mathfrak{w}_{v_n,v_{n-1}}^{n,\theta}\big[(\mathscr{R}_{r^{1/(n-1)}})'(\mathcal{N}_{r,v_{n-1}}^{n-1,\theta}(x))\big]\big)\Big]\,\mu(\mathrm{d}x),
\end{aligned} \tag{4.60}$$

(iii) it holds for all $r \in [1,\infty)$, $k \in \{1,\dots,L\}$, $i \in \{1,\dots,\ell_k\}$ that

$$\begin{aligned}
\left(\frac{\partial \mathcal{L}_r}{\partial \theta_{\ell_k\ell_{k-1}+i+\mathbf{d}_{k-1}}}\right)(\theta) &= \sum_{\substack{v_k, v_{k+1}, \dots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N}\cap[k,L] \colon v_w \leq \ell_w}} \int_{[a,b]^{\ell_0}} 2\Big[\mathbb{1}_{\{i\}}(v_k)\Big] \\
&\cdot \Big[\mathcal{N}_{r,v_L}^{L,\theta}(x) - f_{v_L}(x)\Big]\Big[\textstyle\prod_{n=k+1}^{L}\big(\mathfrak{w}_{v_n,v_{n-1}}^{n,\theta}\big[(\mathscr{R}_{r^{1/(n-1)}})'(\mathcal{N}_{r,v_{n-1}}^{n-1,\theta}(x))\big]\big)\Big]\,\mu(\mathrm{d}x),
\end{aligned} \tag{4.61}$$

(iv) it holds that $\limsup_{r\to\infty}(|\mathcal{L}_r(\theta) - \mathcal{L}_\infty(\theta)| + \|(\nabla\mathcal{L}_r)(\theta) - \mathcal{G}(\theta)\|) = 0$,

(v) it holds for all $k \in \{1,\dots,L\}$, $i \in \{1,\dots,\ell_k\}$, $j \in \{1,\dots,\ell_{k-1}\}$ that

$$\begin{aligned}
&\mathcal{G}_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}(\theta) \\
&= \sum_{\substack{v_k, v_{k+1}, \dots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N}\cap[k,L] \colon v_w \leq \ell_w}} \int_{[a,b]^{\ell_0}} 2\Big[\mathscr{R}_\infty(\mathcal{N}_{\infty,j}^{\max\{k-1,1\},\theta}(x))\mathbb{1}_{(1,L]}(k) + x_j \mathbb{1}_{\{1\}}(k)\Big] \\
&\quad \cdot \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathcal{N}_{\infty,v_L}^{L,\theta}(x) - f_{v_L}(x)\Big]\Big[\textstyle\prod_{n=k+1}^{L}\big(\mathfrak{w}_{v_n,v_{n-1}}^{n,\theta}\mathbb{1}_{\mathcal{X}_{v_{n-1}}^{n-1,\theta}}(x)\big)\Big]\,\mu(\mathrm{d}x),
\end{aligned} \tag{4.62}$$

and

(vi) it holds for all $k \in \{1,\dots,L\}$, $i \in \{1,\dots,\ell_k\}$ that

$$\begin{aligned}
\mathcal{G}_{\ell_k\ell_{k-1}+i+\mathbf{d}_{k-1}}(\theta) &= \sum_{\substack{v_k, v_{k+1}, \dots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N}\cap[k,L] \colon v_w \leq \ell_w}} \int_{[a,b]^{\ell_0}} 2\Big[\mathbb{1}_{\{i\}}(v_k)\Big] \\
&\cdot \Big[\mathcal{N}_{\infty,v_L}^{L,\theta}(x) - f_{v_L}(x)\Big]\Big[\textstyle\prod_{n=k+1}^{L}\big(\mathfrak{w}_{v_n,v_{n-1}}^{n,\theta}\mathbb{1}_{\mathcal{X}_{v_{n-1}}^{n-1,\theta}}(x)\big)\Big]\,\mu(\mathrm{d}x).
\end{aligned} \tag{4.63}$$

Proof of Theorem 4.1.9. Note that Lemma 4.1.6 and the chain rule show that for all $r \in [1,\infty)$, $i \in \{1,\dots,\mathfrak{d}\}$, $x \in [a,b]^{\ell_0}$ it holds that $\mathbb{R}^{\mathfrak{d}} \ni \vartheta \mapsto \|\mathcal{N}_r^{L,\vartheta}(x) - f(x)\|^2 \in \mathbb{R}$ is differentiable at θ and that

$$\tfrac{\partial}{\partial\theta_i}\|\mathcal{N}_r^{L,\theta}(x) - f(x)\|^2 = 2\textstyle\sum_{j=1}^{\ell_L}\big((\mathcal{N}_{r,j}^{L,\theta}(x) - f_j(x))\tfrac{\partial}{\partial\theta_i}\big[\mathcal{N}_{r,j}^{L,\theta}(x)\big]\big). \tag{4.64}$$

Combining this with Lemma 4.1.8, Lemma 4.1.7, Lemma 4.1.6, and the dominated convergence theorem establishes items (i), (ii) and (iii). Observe that the dominated convergence theorem and the fact that for all $x \in [a,b]^{\ell_0}$ it holds that $\limsup_{r\to\infty} |(\mathcal{N}_r^{L,\theta}(x) - f(x)) - (\mathcal{N}_\infty^{L,\theta}(x) - f(x))| = 0$ ensure that

$$\limsup_{r\to\infty} |\mathcal{L}_r(\theta) - \mathcal{L}_\infty(\theta)| = 0. \tag{4.65}$$

Furthermore, note that items (iv) and (v) of Proposition 4.1.5 demonstrate that for all $k \in \{1,\dots,L\}$, $i \in \{1,\dots,\ell_k\}$, $j \in \{1,\dots,\ell_{k-1}\}$, $x \in [a,b]^{\ell_0}$ it holds that

$$\begin{aligned}
&\lim_{r\to\infty}\Bigg(\sum_{\substack{v_k,v_{k+1},\dots,v_L\in\mathbb{N},\\ \forall w\in\mathbb{N}\cap[k,L]\colon v_w\leqslant\ell_w}} \Big[\mathscr{R}_{r^{1/(\max\{k-1,1\})}}(\mathcal{N}_{r,j}^{\max\{k-1,1\},\theta}(x))\mathbb{1}_{(1,L]}(k) + x_j\mathbb{1}_{\{1\}}(k)\Big] \\
&\quad\cdot \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathcal{N}_{r,v_L}^{L,\theta}(x) - f_{v_L}(x)\Big]\Big[\textstyle\prod_{n=k+1}^{L}\big(\mathfrak{w}_{v_n,v_{n-1}}^{n,\theta}\big[(\mathscr{R}_{r^{1/(n-1)}})'(\mathcal{N}_{r,v_{n-1}}^{n-1,\theta}(x))\big]\big)\Big]\Bigg) \\
&= \sum_{\substack{v_k,v_{k+1},\dots,v_L\in\mathbb{N},\\ \forall w\in\mathbb{N}\cap[k,L]\colon v_w\leqslant\ell_w}} \Big[\mathscr{R}_\infty(\mathcal{N}_{\infty,j}^{\max\{k-1,1\},\theta}(x))\mathbb{1}_{(1,L]}(k) + x_j\mathbb{1}_{\{1\}}(k)\Big] \\
&\quad\cdot \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathcal{N}_{\infty,v_L}^{L,\theta}(x) - f_{v_L}(x)\Big]\Big[\textstyle\prod_{n=k+1}^{L}\big(\mathfrak{w}_{v_n,v_{n-1}}^{n,\theta}\mathbb{1}_{\mathcal{X}_{n-1}^{n-1,\theta}}(x)\big)\Big]
\end{aligned} \tag{4.66}$$

and

$$\begin{aligned}
&\lim_{r\to\infty}\Bigg(\sum_{\substack{v_k,v_{k+1},\dots,v_L\in\mathbb{N},\\ \forall w\in\mathbb{N}\cap[k,L]\colon v_w\leqslant\ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathcal{N}_{r,v_L}^{L,\theta}(x) - f_{v_L}(x)\Big] \\
&\quad\cdot \Big[\textstyle\prod_{n=k+1}^{L}\big(\mathfrak{w}_{v_n,v_{n-1}}^{n,\theta}\big[(\mathscr{R}_{r^{1/(n-1)}})'(\mathcal{N}_{r,v_{n-1}}^{n-1,\theta}(x))\big]\big)\Big]\Bigg) \\
&= \sum_{\substack{v_k,v_{k+1},\dots,v_L\in\mathbb{N},\\ \forall w\in\mathbb{N}\cap[k,L]\colon v_w\leqslant\ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathcal{N}_{\infty,v_L}^{L,\theta}(x) - f_{v_L}(x)\Big]\Big[\textstyle\prod_{n=k+1}^{L}\big(\mathfrak{w}_{v_n,v_{n-1}}^{n,\theta}\mathbb{1}_{\mathcal{X}_{n-1}^{n-1,\theta}}(x)\big)\Big]
\end{aligned} \tag{4.67}$$

Combining Lemma 4.1.7, Lemma 4.1.8, (4.65), (4.66), (4.67), and the dominated convergence theorem establishes items (iv), (v) and (vi). The proof of Theorem 4.1.9 is thus complete. ☐

4.1.4 Local Lipschitz continuity properties of the risk function

In the next result we prove that the risk and the realization functions are Lipschitz continuous. In this sense, Lemma 4.1.10 extends [63, Lemma 2.4].

Lemma 4.1.10. *Assume Setting 4.1.1 and let $\mathscr{K} \subseteq \mathbb{R}^{\mathfrak{d}}$ be compact. Then there exists $\mathscr{L} \in \mathbb{R}$ which satisfies for all $\theta, \vartheta \in \mathscr{K}$ that*

$$|\mathcal{L}_\infty(\theta) - \mathcal{L}_\infty(\vartheta)| + \sup_{x\in[a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\theta}(x) - \mathcal{N}_\infty^{L,\vartheta}(x)\| \leqslant \mathscr{L}\|\theta - \vartheta\|. \tag{4.68}$$

Proof of Lemma 4.1.10. Observe that, e.g., Beck et al. [11, Theorem 2.1] (applied with $d \curvearrowleft \mathfrak{d}$, $l = (l_0, l_1, \dots, l_L) \curvearrowleft (\ell_0, \ell_1, \dots, \ell_L)$ in the notation of [11, Theorem 2.1]) implies that for all $\theta, \vartheta \in \mathscr{K}$ it holds that

$$\sup_{x\in[a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\theta}(x) - \mathcal{N}_\infty^{L,\vartheta}(x)\| \leqslant L\mathbf{a}\big[\textstyle\prod_{p=0}^{L-1}(\ell_p + 1)\big]\sqrt{\ell_L}\big(\max\{1, \|\vartheta\|, \|\theta\|\}\big)\|\theta - \vartheta\|. \tag{4.69}$$

Furthermore, note that the fact that $\mathscr{K}$ is compact ensures that there exists $\kappa \in [1, \infty)$ which satisfies for all $\theta \in \mathscr{K}$ that

$$\|\theta\| \leqslant \kappa < \infty. \tag{4.70}$$

Observe that (4.69) and (4.70) demonstrate that there exists $\mathscr{L} \in \mathbb{R}$ which satisfies for all $\theta, \vartheta \in \mathscr{K}$ that

$$\sup_{x \in [a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\theta}(x) - \mathcal{N}_\infty^{L,\vartheta}(x)\| \leqslant \mathscr{L}\|\theta - \vartheta\|. \tag{4.71}$$

Combining this with the Cauchy-Schwarz inequality shows for all $\theta, \vartheta \in \mathscr{K}$ that

$$\begin{aligned}
&|\mathcal{L}_\infty(\theta) - \mathcal{L}_\infty(\vartheta)| \\
&= \left| \left[\int_{[a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\theta}(x) - f(x)\|^2 \, \mu(\mathrm{d}x) \right] - \left[\int_{[a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\vartheta}(x) - f(x)\|^2 \, \mu(\mathrm{d}x) \right] \right| \\
&\leqslant \int_{[a,b]^{\ell_0}} \big| \|\mathcal{N}_\infty^{L,\theta}(x) - f(x)\|^2 - \|\mathcal{N}_\infty^{L,\vartheta}(x) - f(x)\|^2 \big| \, \mu(\mathrm{d}x) \\
&= \int_{[a,b]^{\ell_0}} \big| \big\langle \mathcal{N}_\infty^{L,\theta}(x) - \mathcal{N}_\infty^{L,\vartheta}(x), \mathcal{N}_\infty^{L,\theta}(x) + \mathcal{N}_\infty^{L,\vartheta}(x) - 2f(x) \big\rangle \big| \, \mu(\mathrm{d}x) \\
&\leqslant \int_{[a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\theta}(x) - \mathcal{N}_\infty^{L,\vartheta}(x)\| \|\mathcal{N}_\infty^{L,\theta}(x) + \mathcal{N}_\infty^{L,\vartheta}(x) - 2f(x)\| \, \mu(\mathrm{d}x) \\
&\leqslant \mathscr{L}\|\theta - \vartheta\| \left[\int_{[a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\theta}(x) + \mathcal{N}_\infty^{L,\vartheta}(x) - 2f(x)\| \, \mu(\mathrm{d}x) \right].
\end{aligned} \tag{4.72}$$

This, (4.70), (4.71), and the fact that for all $x \in [a,b]^{\ell_0}$ it holds that $\mathcal{N}_\infty^{L,0}(x) = 0$ imply that for all $\theta, \vartheta \in \mathscr{K}$ we have that

$$\begin{aligned}
&|\mathcal{L}_\infty(\theta) - \mathcal{L}_\infty(\vartheta)| \leqslant \mathscr{L}\|\theta - \vartheta\| \left[\int_{[a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\theta}(x) + \mathcal{N}_\infty^{L,\vartheta}(x)\| + \|2f(x)\| \, \mu(\mathrm{d}x) \right] \\
&\leqslant \mathscr{L}\|\theta - \vartheta\| \left[\mathfrak{m}\big(\sup_{y \in [a,b]^{\ell_0}} [\|\mathcal{N}_\infty^{L,\theta}(y)\| + \|\mathcal{N}_\infty^{L,\vartheta}(y)\|] \big) + 2 \int_{[a,b]^{\ell_0}} \|f(x)\| \, \mu(\mathrm{d}x) \right] \\
&= \mathscr{L}\|\theta - \vartheta\| \Bigg[\mathfrak{m}\big(\sup_{y \in [a,b]^{\ell_0}} [\|\mathcal{N}_\infty^{L,\theta}(y) - \mathcal{N}^{L,0}(y)\| + \|\mathcal{N}_\infty^{L,\vartheta}(y) - \mathcal{N}^{L,0}(y)\|] \big) \\
&\quad + 2 \int_{[a,b]^{\ell_0}} \|f(x)\| \, \mu(\mathrm{d}x) \Bigg] \\
&\leqslant \mathscr{L}\|\theta - \vartheta\| \left[\mathfrak{m}\big(\mathscr{L}\|\theta\| + \mathscr{L}\|\vartheta\| \big) + 2 \int_{[a,b]^{\ell_0}} \|f(x)\| \, \mu(\mathrm{d}x) \right] \\
&\leqslant 2\mathscr{L} \left[\kappa \mathscr{L} \mathfrak{m} + \int_{[a,b]^{\ell_0}} \|f(x)\| \, \mu(\mathrm{d}x) \right] \|\theta - \vartheta\|.
\end{aligned} \tag{4.73}$$

Combining this, Lemma 4.1.8, and (4.71) establishes (4.68). The proof of Lemma 4.1.10 is thus complete. □

4.1.5 Upper estimates for the norm of the generalized gradients of the risk function

The following theorem provides an explicit polynomial growth estimate for the generalized gradient functions of the risk function. Theorem 4.1.11 generalizes [63, Lemma 2.5] to the setting of deep ANNs.

Theorem 4.1.11. *Assume Setting 4.1.1, let $\theta \in \mathbb{R}^{\mathfrak{d}}$, for every $k \in \mathbb{N}$, $i \in \{1, \ldots, \ell_k\}$ let $Q_{k,i} \in \mathbb{R}$ satisfy $Q_{k,i} = |\mathfrak{b}_i^{k,\theta}|^2 + \sum_{j=1}^{\ell_{k-1}} |\mathfrak{w}_{i,j}^{k,\theta}|^2$, and let $\mathcal{Q}_k \in \mathbb{R}$, $k \in \mathbb{N}_0$, satisfy for all $k \in \mathbb{N}$ that $\mathcal{Q}_0 = 1$ and $\mathcal{Q}_k = 1 + \sum_{i=1}^{\ell_k} Q_{k,i}$. Then*

(i) *it holds for all* $k \in \{1, \ldots, L\}$, $i \in \{1, \ldots, \ell_k\}$, $x \in [a,b]^{\ell_0}$ *that*

$$|\mathscr{R}_\infty(\mathcal{N}^{k,\theta}_{\infty,i}(x))|^2 \leqslant |\mathcal{N}^{k,\theta}_{\infty,i}(x)|^2 \leqslant \mathbf{a}^2 Q_{k,i}\big[\textstyle\prod_{p=0}^{k-1}\big((\ell_p+1)\mathcal{Q}_p\big)\big], \tag{4.74}$$

(ii) *it holds for all* $K \in \{1, \ldots, L\}$ *that* $\prod_{p=0}^{K} \mathcal{Q}_p \leqslant (\|\theta\|^2 + 1)^K$,

(iii) *it holds for all* $K \in \{1, \ldots, L\}$, $k \in \{1, \ldots, K\}$, $i \in \{1, \ldots, \ell_k\}$ *that*

$$\sum_{\substack{v_k, v_{k+1}, \ldots, v_K \in \mathbb{N},\\ \forall w \in \mathbb{N} \cap [k,K]\colon v_w \leqslant \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{K} |\mathfrak{w}^{n,\theta}_{v_n, v_{n-1}}|^2\Big] \leqslant \|\theta\|^{2(K-k)}, \tag{4.75}$$

and

(iv) *it holds that*

$$\|\mathcal{G}(\theta)\|^2 \leqslant 4L\mathfrak{m}\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](\|\theta\|^2+1)^{L-1}\mathcal{L}_\infty(\theta). \tag{4.76}$$

Proof of Theorem 4.1.11. We first prove item (i) by induction on $k \in \{1, \ldots, L\}$. Note that (4.5), the fact that for all $x \in \mathbb{R}$ it holds that $|\mathscr{R}_\infty(x)| = |\max\{x, 0\}| \leqslant |x|$, and the Cauchy-Schwarz inequality ensure that for all $i \in \{1, \ldots, \ell_1\}$, $x = (x_1, \ldots, x_{\ell_0}) \in [a,b]^{\ell_0}$ it holds that

$$\begin{aligned}
&|\mathscr{R}_\infty(\mathcal{N}^{1,\theta}_{\infty,i}(x))|^2 \leqslant |\mathcal{N}^{1,\theta}_{\infty,i}(x)|^2 = \big|\mathfrak{b}^{1,\theta}_i + \textstyle\sum_{j=1}^{\ell_0} \mathfrak{w}^{1,\theta}_{i,j} x_i\big|^2 \\
&\leqslant (\ell_0+1)\big(|\mathfrak{b}^{1,\theta}_i|^2 + \textstyle\sum_{j=1}^{\ell_0} |\mathfrak{w}^{1,\theta}_{i,j}|^2 |x_j|^2\big) \\
&\leqslant \mathbf{a}^2(\ell_0+1)\big(|\mathfrak{b}^{1,\theta}_i|^2 + \textstyle\sum_{j=1}^{\ell_0} |\mathfrak{w}^{1,\theta}_{i,j}|^2\big) = \mathbf{a}^2(\ell_0+1)Q_{1,i}.
\end{aligned} \tag{4.77}$$

This establishes item (i) in the base case $k = 1$. For the induction step let $k \in \mathbb{N} \cap [1, L)$ satisfy for all $i \in \{1, \ldots, \ell_k\}$, $x \in [a,b]^{\ell_0}$ that

$$|\mathscr{R}_\infty(\mathcal{N}^{k,\theta}_{\infty,i}(x))|^2 \leqslant |\mathcal{N}^{k,\theta}_{\infty,i}(x)|^2 \leqslant \mathbf{a}^2 Q_{k,i}\big[\textstyle\prod_{p=0}^{k-1}\big((\ell_p+1)\mathcal{Q}_p\big)\big]. \tag{4.78}$$

Observe that (4.78), the fact that for all $x \in \mathbb{R}$ it holds that $|\mathscr{R}_\infty(x)| = |\max\{x, 0\}| \leqslant |x|$, the fact that for all $p \in \mathbb{N}_0$ it holds that $\mathcal{Q}_p \geqslant 1$, and the Cauchy-Schwarz inequality demonstrate that for all $i \in \{1, \ldots, \ell_{k+1}\}$, $x = (x_1, \ldots, x_{\ell_0}) \in [a,b]^{\ell_0}$ it holds that

$$\begin{aligned}
&|\mathscr{R}_\infty(\mathcal{N}^{k+1,\theta}_{\infty,i}(x))|^2 \\
&\leqslant |\mathcal{N}^{k+1,\theta}_{\infty,i}(x)|^2 = \big|\mathfrak{b}^{k+1,\theta}_i + \textstyle\sum_{j=1}^{\ell_k} \mathfrak{w}^{k+1,\theta}_{i,j} \mathscr{R}_\infty(\mathcal{N}^{k,\theta}_{\infty,j}(x))\big|^2 \\
&\leqslant (\ell_k+1)\Big(|\mathfrak{b}^{k+1,\theta}_i|^2 + \textstyle\sum_{j=1}^{\ell_k} |\mathfrak{w}^{k+1,\theta}_{i,j}|^2 |\mathscr{R}_\infty(\mathcal{N}^{k,\theta}_{\infty,j}(x))|^2\Big) \\
&\leqslant (\ell_k+1)\Big(|\mathfrak{b}^{k+1,\theta}_i|^2 + \textstyle\sum_{j=1}^{\ell_k}\big[|\mathfrak{w}^{k+1,\theta}_{i,j}|^2 \mathbf{a}^2 Q_{k,j}\big[\prod_{p=0}^{k-1}\big((\ell_p+1)\mathcal{Q}_p\big)\big]\big]\Big) \\
&\leqslant (\ell_k+1)\mathbf{a}^2\big[\textstyle\prod_{p=0}^{k-1}\big((\ell_p+1)\mathcal{Q}_p\big)\big]\big(|\mathfrak{b}^{k+1,\theta}_i|^2 + \sum_{j=1}^{\ell_k} |\mathfrak{w}^{k+1,\theta}_{i,j}|^2\big)\big(1 + \sum_{j=1}^{\ell_k} Q_{k,j}\big) \\
&= \mathbf{a}^2\big[\textstyle\prod_{p=0}^{k}\big((\ell_p+1)\mathcal{Q}_p\big)\big]\big(|\mathfrak{b}^{k+1,\theta}_i|^2 + \sum_{j=1}^{\ell_k} |\mathfrak{w}^{k+1,\theta}_{i,j}|^2\big) \\
&= \mathbf{a}^2 Q_{k+1,i}\big[\textstyle\prod_{p=0}^{k}\big((\ell_p+1)\mathcal{Q}_p\big)\big].
\end{aligned} \tag{4.79}$$

Induction thus establishes item (i). Note that the fact that for all $k \in \{1, \ldots, L\}$ it holds that $\mathcal{Q}_0 = 1$ and

$$\mathcal{Q}_k = 1 + \textstyle\sum_{i=1}^{\ell_k} Q_{k,i} = 1 + \sum_{i=1}^{\ell_k}\big(|\mathfrak{b}^{k,\theta}_i|^2 + \sum_{j=1}^{\ell_{k-1}} |\mathfrak{w}^{k,\theta}_{i,j}|^2\big) \leqslant 1 + \|\theta\|^2 \tag{4.80}$$

establishes item (ii). We now prove item (iii) by induction on $K \in \{1, \ldots, L\}$. Observe that the fact that for all $i \in \{1, \ldots, \ell_K\}$ it holds that $\sum_{v=1}^{\ell_K} \mathbb{1}_{\{i\}}(v) = 1$ establishes (4.75) in the base case $K = 1$. For the induction step let $K \in \mathbb{N} \cap [1, L)$ satisfy for all $k \in \{1, \ldots, K\}$, $i \in \{1, \ldots, \ell_k\}$ that

$$\sum_{\substack{v_k, v_{k+1}, \ldots, v_K \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,K] \colon v_w \leqslant \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{K} |\mathfrak{w}^{n,\theta}_{v_n, v_{n-1}}|^2\Big] \leqslant \|\theta\|^{2(K-k)}. \tag{4.81}$$

Note that (4.81) implies that for all $k \in \{1, \ldots, K+1\}$, $i \in \{1, \ldots, \ell_k\}$ it holds that

$$\begin{split}
&\sum_{\substack{v_k, v_{k+1}, \ldots, v_{K+1} \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,K+1] \colon v_w \leqslant \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{K+1} |\mathfrak{w}^{n,\theta}_{v_n, v_{n-1}}|^2\Big] \\
&= \left[\sum_{\substack{v_k, v_{k+1}, \ldots, v_K \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,K] \colon v_w \leqslant \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{K} |\mathfrak{w}^{n,\theta}_{v_n, v_{n-1}}|^2\Big]\right]\left[\sum_{v_{K+1}=1}^{\ell_{K+1}} |\mathfrak{w}^{K+1,\theta}_{v_{K+1}, v_K}|^2\right] \\
&\leqslant \left[\sum_{\substack{v_k, v_{k+1}, \ldots, v_K \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,K] \colon v_w \leqslant \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{K} |\mathfrak{w}^{n,\theta}_{v_n, v_{n-1}}|^2\Big]\right]\|\theta\|^2 \leqslant \|\theta\|^{2(K+1-k)}.
\end{split} \tag{4.82}$$

Induction thus establishes item (iii). It thus remains to prove item (iv). For this assume without loss of generality that $\mathfrak{m} > 0$, for every $k \in \mathbb{N}_0$ let $\mathbf{d}_k \in \mathbb{N}_0$ satisfy $\mathbf{d}_k = \sum_{n=1}^{k} \ell_n(\ell_{n-1} + 1)$, and for every $k \in \{1, \ldots, L\}$, $i \in \{1, \ldots, \ell_k\}$ let $W_{k,i} \in \mathbb{R}$ satisfy

$$W_{k,i} = \sum_{\substack{v_k, v_{k+1}, \ldots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{L} |\mathfrak{w}^{n,\theta}_{v_n, v_{n-1}}|\Big]. \tag{4.83}$$

Observe that item (vi) of Theorem 4.1.9 proves that for all $k \in \{1, \ldots, L\}$, $i \in \{1, \ldots, \ell_k\}$ it holds that

$$\begin{split}
&|\mathcal{G}_{\ell_k \ell_{k-1} + i + \mathbf{d}_{k-1}}(\theta)|^2 \\
&= \left[\sum_{\substack{v_k, v_{k+1}, \ldots, v_L \in \mathbb{N}, \\ \forall\, w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \int_{[a,b]^{\ell_0}} 2 \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathcal{N}^{L,\theta}_{\infty, v_L}(x) - f_{v_L}(x)\Big]\right. \\
&\qquad \left. \cdot \Big[\textstyle\prod_{n=k+1}^{L} \big(\mathfrak{w}^{n,\theta}_{v_n, v_{n-1}} \mathbb{1}_{\mathcal{X}^{n-1,\theta}_{v_{n-1}}}(x)\big)\Big] \, \mu(\mathrm{d}x)\right]^2 \\
&\leqslant 4\left[\int_{[a,b]^{\ell_0}} \|\mathcal{N}^{L,\theta}_{\infty}(x) - f(x)\|\right. \\
&\qquad \left. \cdot \left[\sum_{\substack{v_k, v_{k+1}, \ldots, v_L \in \mathbb{N}, \\ \forall\, w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{L} |\mathfrak{w}^{n,\theta}_{v_n, v_{n-1}}|\Big]\right] \mu(\mathrm{d}x)\right]^2 \\
&= 4\left[W_{k,i} \int_{[a,b]^{\ell_0}} \|\mathcal{N}^{L,\theta}_{\infty}(x) - f(x)\| \, \mu(\mathrm{d}x)\right]^2.
\end{split} \tag{4.84}$$

Jensen's inequality hence shows that for all $k \in \{1, \dots, L\}$, $i \in \{1, \dots, \ell_k\}$ it holds that

$$\begin{aligned} |\mathcal{G}_{\ell_k \ell_{k-1}+i+\mathbf{d}_{k-1}}(\theta)|^2 &\leqslant 4\mathfrak{m}^2 (W_{k,i})^2 \left[\frac{1}{\mathfrak{m}} \int_{[a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\theta}(x) - f(x)\| \, \mu(\mathrm{d}x) \right]^2 \\ &\leqslant 4\mathfrak{m}^2 (W_{k,i})^2 \left[\frac{1}{\mathfrak{m}} \int_{[a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\theta}(x) - f(x)\|^2 \, \mu(\mathrm{d}x) \right] = 4\mathfrak{m} (W_{k,i})^2 \mathcal{L}_\infty(\theta). \end{aligned} \tag{4.85}$$

Furthermore, note that item (v) of Theorem 4.1.9 and Jensen's inequality imply for all $k \in \{1, \dots, L\}$, $i \in \{1, \dots, \ell_k\}$, $j \in \{1, \dots, \ell_{k-1}\}$ that

$$\begin{aligned} &|\mathcal{G}_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}(\theta)|^2 \\ &= \Bigg[\sum_{\substack{v_k, v_{k+1}, \dots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \int_{[a,b]^{\ell_0}} 2 \left[\mathscr{R}_\infty(\mathcal{N}_{\infty,j}^{\max\{k-1,1\},\theta}(x)) \mathbb{1}_{(1,L]}(k) + x_j \mathbb{1}_{\{1\}}(k) \right] \\ &\quad \cdot \left[\mathbb{1}_{\{i\}}(v_k) \right] \left[\mathcal{N}_{\infty,v_L}^{L,\theta}(x) - f_{v_L}(x) \right] \left[\textstyle\prod_{n=k+1}^L \big(\mathfrak{w}_{v_n, v_{n-1}}^{n,\theta} \mathbb{1}_{\mathcal{X}_{v_{n-1}}^{n-1,\theta}}(x) \big) \right] \mu(\mathrm{d}x) \Bigg]^2 \\ &\leqslant 4 \Bigg[\int_{[a,b]^{\ell_0}} \left[\mathscr{R}_\infty(\mathcal{N}_{\infty,j}^{\max\{k-1,1\},\theta}(x)) \mathbb{1}_{(1,L]}(k) + x_j \mathbb{1}_{\{1\}}(k) \right] \\ &\quad \cdot \|\mathcal{N}_\infty^{L,\theta}(x) - f(x)\| W_{k,i} \, \mu(\mathrm{d}x) \Bigg]^2 \\ &= 4\mathfrak{m}^2 \Bigg[W_{k,i} \Bigg[\frac{1}{\mathfrak{m}} \int_{[a,b]^{\ell_0}} \left[\mathscr{R}_\infty(\mathcal{N}_{\infty,j}^{\max\{k-1,1\},\theta}(x)) \mathbb{1}_{(1,L]}(k) + x_j \mathbb{1}_{\{1\}}(k) \right] \\ &\quad \cdot \|\mathcal{N}_\infty^{L,\theta}(x) - f(x)\| \, \mu(\mathrm{d}x) \Bigg] \Bigg]^2 \\ &\leqslant 4\mathfrak{m}^2 (W_{k,i})^2 \Bigg[\frac{1}{\mathfrak{m}} \int_{[a,b]^{\ell_0}} \left[\mathscr{R}_\infty(\mathcal{N}_{\infty,j}^{\max\{k-1,1\},\theta}(x)) \mathbb{1}_{(1,L]}(k) + x_j \mathbb{1}_{\{1\}}(k) \right]^2 \\ &\quad \cdot \|\mathcal{N}_\infty^{L,\theta}(x) - f(x)\|^2 \, \mu(\mathrm{d}x) \Bigg]. \end{aligned} \tag{4.86}$$

This ensures for all $i \in \{1, \dots, \ell_1\}$, $j \in \{1, \dots, \ell_0\}$ that

$$|\mathcal{G}_{(i-1)\ell_0+j}(\theta)|^2 \leqslant 4\mathfrak{m}\mathbf{a}^2 (W_{1,i})^2 \mathcal{L}_\infty(\theta). \tag{4.87}$$

Moreover, observe that (4.86) and item (i) demonstrate for all $k \in \mathbb{N} \cap (1, L]$, $i \in \{1, \dots, \ell_k\}$, $j \in \{1, \dots, \ell_{k-1}\}$ that

$$\begin{aligned} &|\mathcal{G}_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}(\theta)|^2 \\ &\leqslant 4\mathfrak{m}(W_{k,i})^2 \Big[\textstyle\int_{[a,b]^{\ell_0}} \mathbf{a}^2 Q_{k-1,j} \big[\prod_{p=0}^{k-2} (\ell_p + 1) \mathcal{Q}_p \big] \|\mathcal{N}_\infty^{L,\theta}(x) - f(x)\|^2 \mu(\mathrm{d}x) \Big] \\ &\leqslant 4\mathfrak{m}(W_{k,i})^2 \mathbf{a}^2 Q_{k-1,j} \big[\textstyle\prod_{p=0}^{k-2} (\ell_p + 1) \mathcal{Q}_p \big] \mathcal{L}_\infty(\theta). \end{aligned} \tag{4.88}$$

In addition, note that the Cauchy-Schwarz inequality and item (iii) imply for all $k \in \{1, \dots, L\}$, $i \in \{1, \dots, \ell_k\}$ that

$$\begin{aligned} (W_{k,i})^2 &\leqslant \big[\textstyle\prod_{p=k+1}^L \ell_p \big] \Big[\textstyle\sum_{\substack{v_k, v_{k+1}, \dots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \big[\mathbb{1}_{\{i\}}(v_k) \big] \big[\prod_{n=k+1}^L |\mathfrak{w}_{v_n, v_{n-1}}^{n,\theta}|^2 \big] \Big] \\ &\leqslant \big[\textstyle\prod_{p=k+1}^L \ell_p \big] \|\theta\|^{2(L-k)}. \end{aligned} \tag{4.89}$$

Combining this with (4.85) and (4.87) assures that

$$\begin{aligned}
&\sum_{i=1}^{\ell_1}\left[|\mathcal{G}_{\ell_1\ell_0+i}(\theta)|^2+\sum_{j=1}^{\ell_0}|\mathcal{G}_{(i-1)\ell_0+j}(\theta)|^2\right]\\
&\leqslant \sum_{i=1}^{\ell_1}\left[\left[4\mathfrak{m}(W_{1,i})^2+\sum_{j=1}^{\ell_0}4\mathfrak{m}\mathbf{a}^2(W_{1,i})^2\right]\mathcal{L}_\infty(\theta)\right]\\
&\leqslant 4\mathfrak{m}\big[\textstyle\prod_{p=2}^{L}\ell_p\big]\|\theta\|^{2(L-1)}(\ell_1+\mathbf{a}^2\ell_1\ell_0)\mathcal{L}_\infty(\theta)\\
&\leqslant 4\mathfrak{m}\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big]\|\theta\|^{2(L-1)}\mathcal{L}_\infty(\theta).
\end{aligned} \tag{4.90}$$

Furthermore, observe that (4.85), (4.88), and (4.89) prove that

$$\begin{aligned}
&\sum_{k=2}^{L}\sum_{i=1}^{\ell_k}\left[|\mathcal{G}_{\ell_k\ell_{k-1}+i+\mathbf{d}_{k-1}}(\theta)|^2+\sum_{j=1}^{\ell_{k-1}}|\mathcal{G}_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}(\theta)|^2\right]\\
&\leqslant 4\mathfrak{m}\left[\sum_{k=2}^{L}\left[\sum_{i=1}^{\ell_k}\left((W_{k,i})^2+\mathbf{a}^2\sum_{j=1}^{\ell_{k-1}}(W_{k,i})^2\mathcal{Q}_{k-1,j}\big[\textstyle\prod_{p=0}^{k-2}(\ell_p+1)\mathcal{Q}_p\big]\right)\right]\right]\mathcal{L}_\infty(\theta)\\
&= 4\mathfrak{m}\left[\sum_{k=2}^{L}\left[\sum_{i=1}^{\ell_k}\left((W_{k,i})^2\left(1+\mathbf{a}^2(\mathcal{Q}_{k-1}-1)\big[\textstyle\prod_{p=0}^{k-2}(\ell_p+1)\mathcal{Q}_p\big]\right)\right)\right]\right]\mathcal{L}_\infty(\theta)\\
&\leqslant 4\mathfrak{m}\left[\sum_{k=2}^{L}\left[\big[\textstyle\prod_{p=k+1}^{L}\ell_p\big]\|\theta\|^{2(L-k)}\ell_k\right.\right.\\
&\qquad\left.\left.\cdot\left(1+\mathbf{a}^2(\mathcal{Q}_{k-1}-1)\Big[\prod_{p=0}^{k-2}\big((\ell_p+1)\mathcal{Q}_p\big)\Big]\right)\right]\mathcal{L}_\infty(\theta)\right]\\
&\leqslant 4\mathbf{a}^2\mathfrak{m}\left[\sum_{k=2}^{L}\left[\big[\textstyle\prod_{p=k}^{L}\ell_p\big]\|\theta\|^{2(L-k)}\right.\right.\\
&\qquad\left.\left.\cdot\left(1+(\mathcal{Q}_{k-1}-1)\Big[\prod_{p=0}^{k-2}(\ell_p+1)\Big]\Big[\prod_{p=0}^{k-2}\mathcal{Q}_p\Big]\right)\right]\right]\mathcal{L}_\infty(\theta).
\end{aligned} \tag{4.91}$$

This, item (ii), and the fact that for all $k \in \mathbb{N}$ it holds that $1 \leqslant \prod_{p=0}^{k-2}\mathcal{Q}_p$ show that

$$\begin{aligned}
&\sum_{k=2}^{L}\sum_{i=1}^{\ell_k}\left[|\mathcal{G}_{\ell_k\ell_{k-1}+i+\mathbf{d}_{k-1}}(\theta)|^2+\sum_{j=1}^{\ell_{k-1}}|\mathcal{G}_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}(\theta)|^2\right]\\
&\leqslant 4\mathbf{a}^2\mathfrak{m}\left[\sum_{k=2}^{L}\left[\big[\textstyle\prod_{p=k}^{L}\ell_p\big]\|\theta\|^{2(L-k)}\big[\textstyle\prod_{p=0}^{k-2}(\ell_p+1)\big]\right.\right.\\
&\qquad\left.\left.\cdot\left(1+(\mathcal{Q}_{k-1}-1)\Big[\prod_{p=0}^{k-2}\mathcal{Q}_p\Big]\right)\right]\right]\mathcal{L}_\infty(\theta)\\
&\leqslant 4\mathbf{a}^2\mathfrak{m}\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big]\left[\displaystyle\sum_{k=2}^{L}\left(\|\theta\|^{2(L-k)}\big[\textstyle\prod_{p=0}^{k-1}\mathcal{Q}_p\big]\right)\right]\mathcal{L}_\infty(\theta)\\
&\leqslant 4\mathbf{a}^2\mathfrak{m}\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big]\left[\displaystyle\sum_{k=2}^{L}\left(\|\theta\|^{2(L-k)}(\|\theta\|^2+1)^{k-1}\right)\right]\mathcal{L}_\infty(\theta)\\
&\leqslant 4\mathbf{a}^2\mathfrak{m}(L-1)\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](\|\theta\|^2+1)^{L-1}\mathcal{L}_\infty(\theta).
\end{aligned} \tag{4.92}$$

Combining this with (4.90) ensures that

$$\begin{aligned}\|\mathcal{G}(\theta)\|^2 &= \sum_{k=1}^{L}\sum_{i=1}^{\ell_k}\Bigg[|\mathcal{G}_{\ell_k\ell_{k-1}+i+\mathbf{d}_{k-1}}(\theta)|^2 + \sum_{j=1}^{\ell_{k-1}}|\mathcal{G}_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}(\theta)|^2\Bigg]\\ &\leqslant 4\mathfrak{m}\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big]\|\theta\|^{2(L-1)}\mathcal{L}_\infty(\theta)\\ &\quad + 4\mathbf{a}^2\mathfrak{m}(L-1)\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](\|\theta\|^2+1)^{L-1}\mathcal{L}_\infty(\theta)\\ &\leqslant 4\mathbf{a}^2\mathfrak{m}L\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](\|\theta\|^2+1)^{L-1}\mathcal{L}_\infty(\theta).\end{aligned} \tag{4.93}$$

This establishes item (iv). The proof of Theorem 4.1.11 is thus complete. □

Combining Lemma 4.1.10 and Theorem 4.1.11 proves that the generalized gradient function is bounded. This is stated in the following corollary, Corollary 4.1.12, which generalizes [63, Corollary 2.6].

Corollary 4.1.12. Assume Setting 4.1.1 and let $K \subseteq \mathbb{R}^{\mathfrak{d}}$ be compact. Then

$$\sup\nolimits_{\theta\in K}\|\mathcal{G}(\theta)\| < \infty. \tag{4.94}$$

Proof of Corollary 4.1.12. Note that Lemma 4.1.10 and the assumption that K is compact demonstrate that $\sup_{\theta\in K}\mathcal{L}_\infty(\theta) < \infty$. Item (iv) of Theorem 4.1.11 therefore establishes (4.94). The proof of Corollary 4.1.12 is thus complete. □

4.1.6 Convexity properties of the risk function

In the next proposition we recall the fact that convex functions from $\mathbb{R}^{\mathfrak{d}}$ to $\mathbb{R}$ do not own any non-global local minimum points.

Proposition 4.1.13. *Let $\mathfrak{d} \in \mathbb{N}$ and let $\mathcal{L}\colon \mathbb{R}^{\mathfrak{d}} \to \mathbb{R}$ satisfy for all $\theta, \vartheta \in \mathbb{R}^{\mathfrak{d}}$, $\lambda \in [0,1]$ that $\mathcal{L}(\lambda\theta + (1-\lambda)\vartheta) \leqslant \lambda\mathcal{L}(\theta) + (1-\lambda)\mathcal{L}(\vartheta)$. Then*

$$\big\{\theta\in\mathbb{R}^{\mathfrak{d}}\colon \big(\exists\,\varepsilon\in(0,\infty)\colon \inf\nolimits_{\vartheta\in\{\varphi\in\mathbb{R}^{\mathfrak{d}}\colon\|\varphi-\theta\|\leqslant\varepsilon\}}\mathcal{L}(\vartheta) = \mathcal{L}(\theta) > \inf\nolimits_{\vartheta\in\mathbb{R}^{\mathfrak{d}}}\mathcal{L}(\vartheta)\big)\big\} = \varnothing. \tag{4.95}$$

Proof of Proposition 4.1.13. We prove (4.95) by contradiction. For this let $\theta, \vartheta \in \mathbb{R}^{\mathfrak{d}}$, $\varepsilon \in (0,\infty)$ satisfy

$$\inf\nolimits_{\varphi\in\{\phi\in\mathbb{R}^{\mathfrak{d}}\colon\|\phi-\theta\|\leqslant\varepsilon\}}\mathcal{L}(\varphi) = \mathcal{L}(\theta) > \mathcal{L}(\vartheta). \tag{4.96}$$

Observe that (4.96) and the fact that for all $\lambda \in [0,1]$ it holds that $\mathcal{L}(\lambda\vartheta + (1-\lambda)\theta) \leqslant \lambda\mathcal{L}(\vartheta) + (1-\lambda)\mathcal{L}(\theta)$ imply that for all $\lambda \in (0,1)$ it holds that

$$\mathcal{L}(\lambda\vartheta + (1-\lambda)\theta) \leqslant \lambda\mathcal{L}(\vartheta) + (1-\lambda)\mathcal{L}(\theta) < \lambda\mathcal{L}(\theta) + (1-\lambda)\mathcal{L}(\theta) = \mathcal{L}(\theta). \tag{4.97}$$

Combing this and (4.96) with the fact that $\inf_{\varphi\in\{\phi\in\mathbb{R}^{\mathfrak{d}}\colon\|\phi-\theta\|\leqslant\varepsilon\}}\mathcal{L}(\varphi) = \inf_{\varphi\in\{\phi\in\mathbb{R}^{\mathfrak{d}}\colon\|\phi\|\leqslant\varepsilon\}}\mathcal{L}(\theta+\varphi)$ shows that for all $\lambda \in (0,1)$ it holds that

$$\mathcal{L}(\theta + \lambda(\vartheta-\theta)) < \inf\nolimits_{\varphi\in\{\phi\in\mathbb{R}^{\mathfrak{d}}\colon\|\phi\|\leqslant\varepsilon\}}\mathcal{L}(\theta+\varphi). \tag{4.98}$$

This contradiction establishes (4.95). The proof of Proposition 4.1.13 is thus complete. □

The following elementary lemma presents a characterization for affine linear functions.

Lemma 4.1.14. *Let V and W be $\mathbb{R}$-vector spaces and let $\varphi\colon V \to W$ be a function. Then the following three statements are equivalent:*

(i) It holds for all $\lambda \in [0,1]$, $v, w \in V$ that $\varphi(\lambda v + (1-\lambda)w) = \lambda\varphi(v) + (1-\lambda)\varphi(w)$.

(ii) *It holds for all* $\lambda \in (0,1)$, $v, w \in V$ *that* $\varphi(\lambda v + (1-\lambda)w) = \lambda\varphi(v) + (1-\lambda)\varphi(w)$.

(iii) *It holds for all* $\lambda \in \mathbb{R}$, $v, w \in V$ *that* $\varphi(\lambda v + w) - \varphi(0) = \lambda(\varphi(v) - \varphi(0)) + (\varphi(w) - \varphi(0))$.

Proof of Lemma 4.1.14. Note that the fact that for all $\lambda \in \{0,1\}$, $v, w \in V$ it holds that $\varphi(\lambda v + (1-\lambda)w) = \lambda\varphi(v) + (1-\lambda)\varphi(w)$ establishes that (item (i) $\leftrightarrow$ item (ii)). We now prove that (item (iii) $\rightarrow$ item (i)). Observe that item (iii) ensures that for all $\lambda \in [0,1]$, $v, w \in V$ it holds that

$$\begin{aligned} &\varphi(\lambda v + (1-\lambda)w) = \lambda(\varphi(v) - \varphi(0)) + \varphi((1-\lambda)w) \\ &= \lambda(\varphi(v) - \varphi(0)) + (1-\lambda)(\varphi(w) - \varphi(0)) + \varphi(0) = \lambda\varphi(v) + (1-\lambda)\varphi(w). \end{aligned} \tag{4.99}$$

This establishes that (item (iii) $\rightarrow$ item (i)). We now prove that (item (i) $\rightarrow$ item (iii)). Note that item (i) implies that for all $\lambda \in [0,1]$, $v \in V$ it holds that

$$\varphi(\lambda v) = \varphi(\lambda v + (1-\lambda)0) = \lambda(\varphi(v)) + (1-\lambda)\varphi(0) = \lambda(\varphi(v) - \varphi(0)) + \varphi(0). \tag{4.100}$$

Furthermore, observe that item (i) shows that for all $\lambda \in (1,\infty)$, $v \in V$ it holds that

$$\begin{aligned} \lambda\varphi(v) &= \lambda\varphi(\tfrac{1}{\lambda}\lambda v + (1 - \tfrac{1}{\lambda})0) \\ &= \lambda\tfrac{1}{\lambda}\varphi(\lambda v) + \lambda(1 - \tfrac{1}{\lambda})\varphi(0) = \varphi(\lambda v) - (1-\lambda)\varphi(0). \end{aligned} \tag{4.101}$$

This and (4.100) prove that for all $\lambda \in [0,\infty)$, $v \in V$ it holds that

$$\varphi(\lambda v) = \lambda(\varphi(v) - \varphi(0)) + \varphi(0). \tag{4.102}$$

Combining this with item (i) ensures that for all $v \in V$ it holds that

$$\begin{aligned} 0 &= \varphi(0) - \varphi(0) = \varphi(\tfrac{1}{2}2v - \tfrac{1}{2}2v) - \varphi(0) = \tfrac{1}{2}\varphi(2v) + \tfrac{1}{2}\varphi(-2v) - \varphi(0) \\ &= \tfrac{1}{2}2(\varphi(v) - \varphi(0)) + \tfrac{1}{2}\varphi(0) + \tfrac{1}{2}2(\varphi(-v) - \varphi(0)) + \tfrac{1}{2}\varphi(0) - \varphi(0) \\ &= \varphi(v) + \varphi(-v) - 2\varphi(0). \end{aligned} \tag{4.103}$$

This and (4.102) show that for all $\lambda \in (-\infty,0)$, $v \in V$ it holds that

$$\begin{aligned} \varphi(\lambda v) - \varphi(0) &= -\varphi(-\lambda v) + \varphi(0) = -(\varphi(-\lambda v) - \varphi(0)) \\ &= -(-\lambda(\varphi(v) - \varphi(0)) + \varphi(0) - \varphi(0)) = \lambda(\varphi(v) - \varphi(0)). \end{aligned} \tag{4.104}$$

Combining this with (4.102) proves that for all $\lambda \in \mathbb{R}$, $v \in V$ it holds that

$$\varphi(\lambda v) = \lambda(\varphi(v) - \varphi(0)) + \varphi(0). \tag{4.105}$$

Hence, we obtain for all $v, w \in V$ that

$$\begin{aligned} &\varphi(v + w) = \varphi(\tfrac{1}{2}2v + \tfrac{1}{2}2w) = \tfrac{1}{2}\varphi(2v) + \tfrac{1}{2}\varphi(2w) \\ &= \tfrac{1}{2}2(\varphi(v) - \varphi(0)) + \tfrac{1}{2}\varphi(0) + \tfrac{1}{2}2(\varphi(w) - \varphi(0)) + \tfrac{1}{2}\varphi(0) = \varphi(v) + \varphi(w) - \varphi(0). \end{aligned} \tag{4.106}$$

This and (4.105) demonstrate that for all $\lambda \in \mathbb{R}$, $v, w \in V$ it holds that

$$\begin{aligned} \varphi(\lambda v + w) - \varphi(0) &= \varphi(\lambda v) + \varphi(w) - 2\varphi(0) \\ &= \lambda(\varphi(v) - \varphi(0)) + \varphi(0) + \varphi(w) - 2\varphi(0) \\ &= \lambda(\varphi(v) - \varphi(0)) + (\varphi(w) - \varphi(0)). \end{aligned} \tag{4.107}$$

This establishes that (item (i) $\rightarrow$ item (iii)). The proof of Lemma 4.1.14 is thus complete. □

We will use the characterization stated in Lemma 4.1.14 to prove in the following proposition that the risk function with respect to the last affine linear transformation is convex.

Proposition 4.1.15. Assume Setting 4.1.1, let $\delta \in \mathbb{N}_0$ satisfy $\delta = \sum_{k=1}^{L-1} \ell_k(\ell_{k-1}+1)$, let $\theta = (\theta_j)_{j \in \mathbb{N} \cap [0,\delta]} \colon \mathbb{N} \cap [0,\delta] \to \mathbb{R}$ be a function, and let $\mathbb{L} \colon \mathbb{R}^{\ell_L(\ell_{L-1}+1)} \to \mathbb{R}$ satisfy for all $v = (v_1, \dots, v_{\ell_L(\ell_{L-1}+1)}) \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$ that

$$\mathbb{L}(v) = \mathcal{L}_\infty\big(\theta_1, \theta_2, \dots, \theta_\delta, v_1, v_2, \dots, v_{\ell_L(\ell_{L-1}+1)}\big). \tag{4.108}$$

Then it holds for all $v, w \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$, $\lambda \in [0,1]$ that

$$\mathbb{L}(\lambda v + (1-\lambda) w) \leqslant \lambda \mathbb{L}(v) + (1-\lambda)\mathbb{L}(w). \tag{4.109}$$

Proof of Proposition 4.1.15. Throughout this proof let $\psi = (\psi_1, \dots, \psi_{\mathfrak{d}}) \colon \mathbb{R}^{\ell_L(\ell_L+1)} \to \mathbb{R}^{\mathfrak{d}}$ satisfy for all $v = (v_1, \dots, v_{\ell_L(\ell_{L-1}+1)}) \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$ that

$$\psi(v) = (\theta_1, \theta_2, \dots, \theta_\delta, v_1, v_2, \dots, v_{\ell_L(\ell_{L-1}+1)}) \tag{4.110}$$

and for every $v \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$ let $\mathscr{N}^v \colon \mathbb{R}^{\ell_0} \to \mathbb{R}^{\ell_k}$ satisfy for all $x \in \mathbb{R}^{\ell_0}$ that

$$\mathscr{N}^v(x) = \mathcal{N}_\infty^{L,\psi(v)}(x). \tag{4.111}$$

Note that Lemma 4.1.14, (4.2), and (4.110) ensure that for all $v, w \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$, $\lambda \in [0,1]$, $x \in \mathbb{R}^{\ell_{L-1}}$ it holds that

$$\begin{aligned}
&\mathfrak{b}^{L,\psi(\lambda v+(1-\lambda)w)} + \mathfrak{w}^{L,\psi(\lambda v+(1-\lambda)w)} x = \mathfrak{b}^{L,\lambda\psi(v)+(1-\lambda)\psi(w)} + \mathfrak{w}^{L,\lambda\psi(v)+(1-\lambda)\psi(w)} x \\
&= \lambda \mathfrak{b}^{L,\psi(v)} + (1-\lambda)\mathfrak{b}^{L,\psi(w)} + \lambda \mathfrak{w}^{L,\psi(v)} x + (1-\lambda)\mathfrak{w}^{L,\psi(w)} x \\
&= \lambda\big[\mathfrak{b}^{L,\psi(v)} + \mathfrak{w}^{L,\psi(v)} x\big] + (1-\lambda)\big[\mathfrak{b}^{L,\psi(w)} + \mathfrak{w}^{L,\psi(w)} x\big].
\end{aligned} \tag{4.112}$$

Next observe that (4.5) shows that for all $v \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$, $x \in \mathbb{R}^{\ell_0}$ it holds that

$$\mathcal{N}_\infty^{L,\psi(v)}(x) = \begin{cases} \mathfrak{b}^{L,\psi(v)} + \mathfrak{w}^{L,\psi(v)} x & \colon L = 1 \\ \mathfrak{b}^{L,\psi(v)} + \mathfrak{w}^{L,\psi(v)} \big(\mathfrak{M}_\infty(\mathcal{N}_\infty^{L-1,\psi(v)}(x))\big) & \colon L > 1. \end{cases} \tag{4.113}$$

Therefore, we obtain that for all $v, w \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$, $\lambda \in [0,1]$, $x \in \mathbb{R}^{\ell_0}$ it holds that

$$\begin{aligned}
&\mathcal{N}_\infty^{L,\psi(\lambda v+(1-\lambda)w)}(x) \\
&= \begin{cases} \mathfrak{b}^{L,\psi(\lambda v+(1-\lambda)w)} + \mathfrak{w}^{L,\psi(\lambda v+(1-\lambda)w)} x & \colon L = 1 \\ \mathfrak{b}^{L,\psi(\lambda v+(1-\lambda)w)} + \mathfrak{w}^{L,\psi(\lambda v+(1-\lambda)w)} \big(\mathfrak{M}_\infty(\mathcal{N}_\infty^{L-1,\psi(\lambda v+(1-\lambda)w)}(x))\big) & \colon L > 1. \end{cases}
\end{aligned} \tag{4.114}$$

Combining this with (4.112) implies that for all $v, w \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$, $\lambda \in [0,1]$, $x \in \mathbb{R}^{\ell_0}$ it holds that

$$\begin{aligned}
&\mathcal{N}_\infty^{L,\psi(\lambda v+(1-\lambda)w)}(x) \\
&= \begin{cases} \lambda\big[\mathfrak{b}^{L,\psi(v)} + \mathfrak{w}^{L,\psi(v)} x\big] + (1-\lambda)\big[\mathfrak{b}^{L,\psi(w)} + \mathfrak{w}^{L,\psi(w)} x\big] & \colon L = 1 \\ \lambda\big[\mathfrak{b}^{L,\psi(v)} + \mathfrak{w}^{L,\psi(v)} \big(\mathfrak{M}_\infty(\mathcal{N}_\infty^{L-1,\psi(\lambda v+(1-\lambda)w)}(x))\big)\big] & \\ +(1-\lambda)\big[\mathfrak{b}^{L,\psi(w)} + \mathfrak{w}^{L,\psi(w)} \big(\mathfrak{M}_\infty(\mathcal{N}_\infty^{L-1,\psi(\lambda v+(1-\lambda)w)}(x))\big)\big] & \colon L > 1. \end{cases}
\end{aligned} \tag{4.115}$$

Item (i) of Proposition 4.1.5, (4.2), and the fact that for all $i \in \mathbb{N} \cap [1, \sum_{k=1}^{L-1} \ell_k(\ell_{k-1}+1)]$, $v, w \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$ it holds that $\psi_i(v) = \psi_i(w)$ hence imply that for all $v, w \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$, $\lambda \in [0,1]$, $x \in \mathbb{R}^{\ell_0}$ it holds that

$$\begin{aligned} &\mathcal{N}_\infty^{L,\psi(\lambda v+(1-\lambda)w)}(x) \\ &= \begin{cases} \lambda\big[\mathfrak{b}^{L,\psi(v)} + \mathfrak{w}^{L,\psi(v)}x\big] + (1-\lambda)\big[\mathfrak{b}^{L,\psi(w)} + \mathfrak{w}^{L,\psi(w)}x\big] & : L = 1 \\ \lambda\big[\mathfrak{b}^{L,\psi(v)} + \mathfrak{w}^{L,\psi(v)}\big(\mathfrak{M}_\infty(\mathcal{N}_\infty^{L-1,\psi(v)}(x))\big)\big] & \\ +(1-\lambda)\big[\mathfrak{b}^{L,\psi(w)} + \mathfrak{w}^{L,\psi(w)}\big(\mathfrak{M}_\infty(\mathcal{N}_\infty^{L-1,\psi(w)}(x))\big)\big] & : L > 1. \end{cases} \end{aligned} \tag{4.116}$$

Moreover, note that (4.113) ensures that for all $v, w \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$, $\lambda \in [0,1]$, $x \in \mathbb{R}^{\ell_0}$ it holds that

$$\begin{aligned} &\lambda\mathcal{N}_\infty^{L,\psi(v)}(x) + (1-\lambda)\mathcal{N}_\infty^{L,\psi(w)}(x) \\ &= \begin{cases} \lambda\big[\mathfrak{b}^{L,\psi(v)} + \mathfrak{w}^{L,\psi(v)}x\big] + (1-\lambda)\big[\mathfrak{b}^{L,\psi(w)} + \mathfrak{w}^{L,\psi(w)}x\big] & : L = 1 \\ \lambda\big[\mathfrak{b}^{L,\psi(v)} + \mathfrak{w}^{L,\psi(v)}\big(\mathfrak{M}_\infty(\mathcal{N}_\infty^{L-1,\psi(v)}(x))\big)\big] & \\ +(1-\lambda)\big[\mathfrak{b}^{L,\psi(w)} + \mathfrak{w}^{L,\psi(w)}\big(\mathfrak{M}_\infty(\mathcal{N}_\infty^{L-1,\psi(w)}(x))\big)\big] & : L > 1. \end{cases} \end{aligned} \tag{4.117}$$

Combining this with (4.116) demonstrates that for all $v, w \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$, $\lambda \in [0,1]$, $x \in \mathbb{R}^{\ell_0}$ it holds that

$$\mathcal{N}_\infty^{L,\psi(\lambda v+(1-\lambda)w)}(x) = \lambda\mathcal{N}_\infty^{L,\psi(v)}(x) + (1-\lambda)\mathcal{N}_\infty^{L,\psi(w)}(x). \tag{4.118}$$

This and (4.111) show that for all $v, w \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$, $\lambda \in [0,1]$, $x \in \mathbb{R}^{\ell_0}$ it holds that

$$\begin{aligned} \mathscr{N}^{\lambda v+(1-\lambda)w}(x) &= \mathcal{N}_\infty^{L,\psi(\lambda v+(1-\lambda)w)}(x) = \lambda\mathcal{N}_\infty^{L,\psi(v)}(x) + (1-\lambda)\mathcal{N}_\infty^{L,\psi(w)}(x) \\ &= \lambda\mathscr{N}^v(x) + (1-\lambda)\mathscr{N}^w(x). \end{aligned} \tag{4.119}$$

Next observe that (4.6), (4.108), (4.110), and (4.111) ensure that for all $v \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$ it holds that

$$\begin{aligned} \mathbb{L}(v) = \mathcal{L}_\infty(\psi(v)) &= \int_{[a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\psi(v)}(x) - f(x)\|^2 \,\mu(\mathrm{d}x) \\ &= \int_{[a,b]^{\ell_0}} \|\mathscr{N}^v(x) - f(x)\|^2 \,\mu(\mathrm{d}x). \end{aligned} \tag{4.120}$$

Combining this, (4.119), and the fact that for all $\lambda \in [0,1]$, $x, y \in \mathbb{R}$ it holds that $(\lambda x + (1-\lambda)y)^2 \leqslant \lambda x^2 + (1-\lambda)y^2$ shows that for all $v, w \in \mathbb{R}^{\ell_L(\ell_{L-1}+1)}$, $\lambda \in [0,1]$ it holds that

$$\begin{aligned} &\mathbb{L}(\lambda v + (1-\lambda)w) = \int_{[a,b]^{\ell_0}} \|\mathscr{N}^{\lambda v+(1-\lambda)w}(x) - f(x)\|^2 \,\mu(\mathrm{d}x) \\ &= \int_{[a,b]^{\ell_0}} \|\lambda(\mathscr{N}^v(x) - f(x)) + (1-\lambda)(\mathscr{N}^w(x) - f(x))\|^2 \,\mu(\mathrm{d}x) \\ &\leqslant \int_{[a,b]^{\ell_0}} \Big[\lambda\|\mathscr{N}^v(x) - f(x)\| + (1-\lambda)\|\mathscr{N}^w(x) - f(x)\|\Big]^2 \,\mu(\mathrm{d}x) \\ &\leqslant \int_{[a,b]^{\ell_0}} \Big[\lambda\|\mathscr{N}^v(x) - f(x)\|^2 + (1-\lambda)\|\mathscr{N}^w(x) - f(x)\|^2\Big] \,\mu(\mathrm{d}x) \\ &= \lambda\bigg[\int_{[a,b]^{\ell_0}} \|\mathscr{N}^v(x) - f(x)\|^2 \,\mu(\mathrm{d}x)\bigg] + (1-\lambda)\bigg[\int_{[a,b]^{\ell_0}} \|\mathscr{N}^w(x) - f(x)\|^2 \,\mu(\mathrm{d}x)\bigg] \\ &= \lambda\mathbb{L}(v) + (1-\lambda)\mathbb{L}(w). \end{aligned} \tag{4.121}$$

This establishes (4.109). The proof of Proposition 4.1.15 is thus complete. □

This leads to the following corollary which proves the convexity of the risk function.

Corollary 4.1.16. Assume Setting 4.1.1 and assume $(L-1)\mathfrak{m} = 0$. Then it holds for all $\theta, \vartheta \in \mathbb{R}^{\mathfrak{d}}$, $\lambda \in [0,1]$ that

$$\mathcal{L}_\infty(\lambda\theta + (1-\lambda)\vartheta) \leqslant \lambda\mathcal{L}_\infty(\theta) + (1-\lambda)\mathcal{L}_\infty(\vartheta). \tag{4.122}$$

Proof of Corollary 4.1.16. Throughout this proof we distinguish between the case $\mathfrak{m} = 0$ and the case $\mathfrak{m} \neq 0$. We first prove (4.122) in the case

$$\mathfrak{m} = 0. \tag{4.123}$$

Note that (4.6) and (4.123) ensure that for all $\theta \in \mathbb{R}^{\mathfrak{d}}$ it holds that

$$\mathcal{L}_\infty(\theta) = \int_{[a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\theta}(x) - \xi\|^2 \,\mu(\mathrm{d}x) = 0. \tag{4.124}$$

Therefore, we obtain that for all $\theta, \vartheta \in \mathbb{R}^{\mathfrak{d}}$, $\lambda \in [0,1]$ it holds that

$$\mathcal{L}_\infty(\lambda\theta + (1-\lambda)\vartheta) = 0 = \lambda\mathcal{L}_\infty(\theta) + (1-\lambda)\mathcal{L}_\infty(\vartheta). \tag{4.125}$$

This establishes (4.122) in the case $\mathfrak{m} = 0$. Next we prove (4.122) in the case

$$\mathfrak{m} \neq 0. \tag{4.126}$$

Observe that (4.126) and the assumption that $(L-1)\mathfrak{m} = 0$ imply that $L = 1$. Proposition 4.1.15 hence demonstrates that for all $\theta, \vartheta \in \mathbb{R}^{\mathfrak{d}}$, $\lambda \in [0,1]$ it holds that

$$\mathcal{L}_\infty(\lambda\theta + (1-\lambda)\vartheta) \leqslant \lambda\mathcal{L}_\infty(\theta) + (1-\lambda)\mathcal{L}_\infty(\vartheta). \tag{4.127}$$

This establishes (4.122) in the case $\mathfrak{m} \neq 0$. The proof of Corollary 4.1.16 is thus complete. □

4.1.7 Non-convexity properties of the risk function

The following proposition, Proposition 4.1.17, computes arithmetical averages in the argument of the risk function.

Proposition 4.1.17. *Assume Setting 4.1.1, assume $L > 1$, let $\xi = (\xi_1, \ldots, \xi_{\ell_L}) \in \mathbb{R}^{\ell_L}$, let $(\alpha_{i,j,k})_{(i,j,k)\in(\mathbb{N}_0)^3} \subseteq \mathbb{R}$ satisfy for all $i, j \in \mathbb{N}_0$, $k \in \mathbb{N}$ that $\alpha_{i,j,k} = i\ell_{k-1} + j + \sum_{h=1}^{k-1} \ell_h(\ell_{h-1} + 1)$, and let $\theta = (\theta_1, \ldots, \theta_{\mathfrak{d}})$, $\vartheta = (\vartheta_1, \ldots, \vartheta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ satisfy for all $i \in \{1, \ldots, \ell_L\}$, $j \in \{1, \ldots, \mathfrak{d}\} \backslash (\cup_{k=1}^{\ell_L} \{\alpha_{\ell_{L-1},1,L-1}, \alpha_{0,1,L}, \alpha_{\ell_L,k,L}\})$ that*

$$\begin{aligned} \theta_{\alpha_{\ell_{L-1},1,L-1}} &= \vartheta_{\alpha_{0,1,L}} = 1, \qquad \theta_{\alpha_{\ell_L,i,L}} = \vartheta_{\alpha_{\ell_L,i,L}} = \xi_i, \\ \text{and} \qquad \theta_{\alpha_{0,1,L}} &= \vartheta_{\alpha_{\ell_{L-1},1,L-1}} = \theta_j = \vartheta_j = 0. \end{aligned} \tag{4.128}$$

Then $\mathcal{L}_\infty(\theta) = \mathcal{L}_\infty(\vartheta)$ and

$$\mathcal{L}_\infty\!\left(\frac{\theta+\vartheta}{2}\right) = \left[\frac{\mathcal{L}_\infty(\theta) + \mathcal{L}_\infty(\vartheta)}{2}\right] + \frac{\mathfrak{m}}{16} + \frac{1}{2}\left[\xi_1\mathfrak{m} - \int_{[a,b]^{\ell_0}} f_1(x)\,\mu(\mathrm{d}x)\right]. \tag{4.129}$$

Proof of Proposition 4.1.17. Note that items (ii) and (iii) of Lemma 4.1.2, and (4.128) ensure that for all $x \in [a,b]^{\ell_0}$ it holds that $\mathcal{N}_\infty^{L,\theta}(x) = \mathcal{N}_\infty^{L,\vartheta}(x) = \xi$. This and (4.6) imply that

$$\mathcal{L}_\infty(\theta) = \mathcal{L}_\infty(\vartheta) = \int_{[a,b]^{\ell_0}} \|\xi - f(x)\|^2 \,\mu(\mathrm{d}x). \tag{4.130}$$

Furthermore, observe that items (ii) and (iii) of Lemma 4.1.2 and (4.128) demonstrate that for all $x \in [a,b]^{\ell_0}$ it holds that

$$\mathcal{N}_{\infty,1}^{L,(\theta+\vartheta)/2}(x) = \tfrac{1}{2}\max\{\tfrac{1}{2},0\} + \xi_1 = \xi_1 + \tfrac{1}{4}. \tag{4.131}$$

Moreover, note that items (ii) and (iii) of Lemma 4.1.2 and (4.128) ensure that for all $i \in \mathbb{N} \cap (1,\ell_L]$, $x \in [a,b]^{\ell_0}$ it holds that $\mathcal{N}_{\infty,i}^{L,(\theta+\vartheta)/2}(x) = \xi_i$. Combining this with (4.130) and (4.131) shows that

$$\begin{aligned}
\mathcal{L}_\infty\Big(\frac{\theta+\vartheta}{2}\Big) &= \int_{[a,b]^{\ell_0}} \Bigg[\big(\xi_1 + \tfrac{1}{4} - f_1(x)\big)^2 + \sum_{i=2}^{\ell_L}\big(\xi_i - f_i(x)\big)^2\Bigg]\,\mu(\mathrm{d}x) \\
&= \Bigg[\int_{[a,b]^{\ell_0}} \|\xi - f(x)\|^2\,\mu(\mathrm{d}x)\Bigg] + \Bigg[\int_{[a,b]^{\ell_0}} \tfrac{1}{16}\,\mu(\mathrm{d}x)\Bigg] + \Bigg[\int_{[a,b]^{\ell_0}} \tfrac{(\xi_1 - f_1(x))}{2}\,\mu(\mathrm{d}x)\Bigg] \\
&= \Bigg[\frac{\mathcal{L}_\infty(\theta) + \mathcal{L}_\infty(\vartheta)}{2}\Bigg] + \frac{\mathfrak{m}}{16} + \frac{1}{2}\Bigg[\xi_1\mathfrak{m} - \int_{[a,b]^{\ell_0}} f_1(x)\,\mu(\mathrm{d}x)\Bigg].
\end{aligned} \tag{4.132}$$

This establishes (4.129). The proof of Proposition 4.1.17 is thus complete. □

Proposition 4.1.17 is used in the following corollary to establish that the risk function is not convex.

Corollary 4.1.18. Assume Setting 4.1.1, assume $(L-1)\mathfrak{m} \neq 0$, let $\xi = (\xi_1,\dots,\xi_{\ell_L}) \in \mathbb{R}^{\ell_L}$ satisfy $\xi_1 = \mathfrak{m}^{-1}\int_{[a,b]^{\ell_0}} f_1(x)\,\mu(\mathrm{d}x)$, let $(\alpha_{i,j,k})_{(i,j,k)\in(\mathbb{N}_0)^3} \subseteq \mathbb{R}$ satisfy for all $i,j \in \mathbb{N}_0$, $k \in \mathbb{N}$ that $\alpha_{i,j,k} = i\ell_{k-1} + j + \sum_{h=1}^{k-1} \ell_h(\ell_{h-1}+1)$, and let $\theta = (\theta_1,\dots,\theta_{\mathfrak{d}})$, $\vartheta = (\vartheta_1,\dots,\vartheta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ satisfy for all $i \in \{1,\dots,\ell_L\}$, $j \in \{1,\dots,\mathfrak{d}\}\backslash(\cup_{k=1}^{\ell_L}\{\alpha_{\ell_{L-1},1,L-1}, \alpha_{0,1,L}, \alpha_{\ell_L,k,L}\})$ that

$$\begin{aligned}
\theta_{\alpha_{\ell_{L-1},1,L-1}} = \vartheta_{\alpha_{0,1,L}} = 1, \qquad \theta_{\alpha_{\ell_L,i,L}} = \vartheta_{\alpha_{\ell_L,i,L}} = \xi_i, \\
\text{and} \qquad \theta_{\alpha_{0,1,L}} = \vartheta_{\alpha_{\ell_{L-1},1,L-1}} = \theta_j = \vartheta_j = 0.
\end{aligned} \tag{4.133}$$

Then

$$\mathcal{L}_\infty\Big(\frac{\theta+\vartheta}{2}\Big) = \Bigg[\frac{\mathcal{L}_\infty(\theta) + \mathcal{L}_\infty(\vartheta)}{2}\Bigg] + \frac{\mathfrak{m}}{16} > \frac{\mathcal{L}_\infty(\theta) + \mathcal{L}_\infty(\vartheta)}{2}. \tag{4.134}$$

Proof of Corollary 4.1.18. Observe that Proposition 4.1.17 and the assumption that $\xi_1 = \mathfrak{m}^{-1}\int_{[a,b]^{\ell_0}} f_1(x)\,\mu(\mathrm{d}x)$ demonstrate that

$$\begin{aligned}
\mathcal{L}_\infty\Big(\frac{\theta+\vartheta}{2}\Big) &= \Bigg[\frac{\mathcal{L}_\infty(\theta) + \mathcal{L}_\infty(\vartheta)}{2}\Bigg] + \frac{\mathfrak{m}}{16} + \frac{1}{2}\Bigg[\xi_1\mathfrak{m} - \int_{[a,b]^{\ell_0}} f_1(x)\mu(\mathrm{d}x)\Bigg] \\
&= \Bigg[\frac{\mathcal{L}_\infty(\theta) + \mathcal{L}_\infty(\vartheta)}{2}\Bigg] + \frac{\mathfrak{m}}{16}.
\end{aligned} \tag{4.135}$$

The fact that $\mathfrak{m} \neq 0$ therefore implies that

$$\mathcal{L}_\infty\Big(\frac{\theta+\vartheta}{2}\Big) > \frac{\mathcal{L}_\infty(\theta) + \mathcal{L}_\infty(\vartheta)}{2}. \tag{4.136}$$

The proof of Corollary 4.1.18 is thus complete. □

This leads to the following characterization of the convexity of the risk function. In particular, Corollary 4.1.19 shows that - provided that the underlying measure is not the zero measure - for every measurable target function it holds that the risk function is not convex.

Corollary 4.1.19. Assume Setting 4.1.1. Then it holds that $(L-1)\mathfrak{m} = 0$ if and only if it holds for all $\theta, \vartheta \in \mathbb{R}^{\mathfrak{d}}$, $\lambda \in [0,1]$ that

$$\mathcal{L}_\infty(\lambda\theta + (1-\lambda)\vartheta) \leqslant \lambda\mathcal{L}_\infty(\theta) + (1-\lambda)\mathcal{L}_\infty(\vartheta). \tag{4.137}$$

Proof of Corollary 4.1.19. Note that Corollary 4.1.16 and Corollary 4.1.18 establish (4.137). The proof of Corollary 4.1.19 is thus complete. □

4.2 Gradient flow (GF) processes in the training of deep ANNs

This section's purpose is to study GF processes in the training of deep ANNs with an arbitrary number of hidden layers under the assumptions that the target function is constant. In Subsection 4.2.1 we establish properties of the considered Lyapunov function in the setting of GF processes. Subsection 4.2.2 and Subsection 4.2.3 present a weak chain rule for the composition of Lyapunov functions and GF solutions and compositions of the risk of GF processes, respectively. Subsection 4.2.4 contains this sections main result, Theorem 4.2.8, which proves under the assumption of a constant target function that the risk of every solution of the associated GF differential equation converges with rate 1 to zero.

4.2.1 Lyapunov type estimates for the dynamics of GF processes

The following proposition shows a presentation of the Lyapunov function as well as regularity properties. Proposition 4.2.1 generalizes [63, Proposition 2.8] from shallow ANNs to deep ANNs.

Proposition 4.2.1. Assume Setting 4.1.1 and let $\theta \in \mathbb{R}^{\mathfrak{d}}$. Then

(i) it holds that

$$V(\theta) = \textstyle\sum_{k=1}^{L}\big(k\big[\sum_{i=1}^{\ell_k}|\mathfrak{b}_i^{k,\theta}|^2\big] + \big[\sum_{i=1}^{\ell_k}\sum_{j=1}^{\ell_{k-1}}|\mathfrak{w}_{i,j}^{k,\theta}|^2\big] - 2L\big[\sum_{i=1}^{\ell_L} f_i(0)\mathfrak{b}_i^{L,\theta}\big]\big), \tag{4.138}$$

(ii) it holds that $\frac{1}{2}\|\theta\|^2 - 2L^2\|f(0)\|^2 \leqslant V(\theta) \leqslant 2L\|\theta\|^2 + L\|f(0)\|^2$, and

(iii) it holds that

$$\begin{aligned}(\nabla V)(\theta) \\ = 2\big(\mathfrak{w}^{1,\theta}, \mathfrak{b}^{1,\theta}, \mathfrak{w}^{2,\theta}, 2\mathfrak{b}^{2,\theta}, \ldots, \mathfrak{w}^{L-1,\theta}, (L-1)\mathfrak{b}^{L-1,\theta}, \mathfrak{w}^{L,\theta}, L(\mathfrak{b}^{L,\theta} - f(0))\big)\,.\end{aligned} \tag{4.139}$$

Proof of Proposition 4.2.1. Observe that (4.7) establishes item (i). Next note that (4.7), the Cauchy-Schwarz inequality, and the Young inequality demonstrate that

$$\begin{aligned} V(\theta) &= \left[\sum_{k=1}^{L} \left(k \|\mathfrak{b}^{k,\theta}\|^2 + \sum_{i=1}^{\ell_k} \sum_{j=1}^{\ell_{k-1}} |\mathfrak{w}^{k,\theta}_{i,j}|^2 \right) \right] - 2L\langle f(0), \mathfrak{b}^{L,\theta} \rangle \\ &= \|\theta\|^2 + \sum_{k=1}^{L} (k-1) \|\mathfrak{b}^{k,\theta}\|^2 - 2L\langle f(0), \mathfrak{b}^{L,\theta} \rangle \geqslant \|\theta\|^2 - 2L\langle f(0), \mathfrak{b}^{L,\theta} \rangle \\ &\geqslant \|\theta\|^2 - 2L\|f(0)\|\|\mathfrak{b}^{L,\theta}\| \geqslant \|\theta\|^2 - 2L^2\|f(0)\|^2 - \frac{1}{2}\|\mathfrak{b}^{L,\theta}\|^2 \\ &\geqslant \frac{1}{2}\|\theta\|^2 - 2L^2\|f(0)\|^2. \end{aligned} \tag{4.140}$$

Furthermore, observe that (4.7) shows that

$$\begin{aligned} V(\theta) &= \left[\sum_{k=1}^{L} \left(k \|\mathfrak{b}^{k,\theta}\|^2 + \sum_{i=1}^{\ell_k} \sum_{j=1}^{\ell_{k-1}} |\mathfrak{w}^{k,\theta}_{i,j}|^2 \right) \right] - 2L\langle f(0), \mathfrak{b}^{L,\theta} \rangle \\ &= \|\theta\|^2 + \sum_{k=1}^{L} (k-1) \|\mathfrak{b}^{k,\theta}\|^2 - 2L\langle f(0), \mathfrak{b}^{L,\theta} \rangle \\ &\leqslant \|\theta\|^2 + \sum_{k=1}^{L} (k-1) \|\mathfrak{b}^{k,\theta}\|^2 + |2L\langle f(0), \mathfrak{b}^{L,\theta} \rangle|. \end{aligned} \tag{4.141}$$

The Cauchy-Schwarz inequality and the Young inequality hence imply that

$$\begin{aligned} V(\theta) &\leqslant \|\theta\|^2 + \sum_{k=1}^{L} (k-1) \|\mathfrak{b}^{k,\theta}\|^2 + 2L\|f(0)\|\|\mathfrak{b}^{L,\theta}\| \\ &\leqslant \|\theta\|^2 + \sum_{k=1}^{L} (k-1) \|\mathfrak{b}^{k,\theta}\|^2 + L\|f(0)\|^2 + L\|\mathfrak{b}^{L,\theta}\|^2 \leqslant 2L\|\theta\|^2 + L\|f(0)\|^2. \end{aligned} \tag{4.142}$$

Combining this and (4.140) proves item (ii). Note that item (i) establishes item (iii). The proof of Proposition 4.2.1 is thus complete. □

In the following proposition we determine the scalar product of the gradient of the Lyapunov function and the generalized gradient function associated to the risk function. In this sense, Proposition 4.2.2 generalizes [63, Proposition 2.9].

Proposition 4.2.2. Assume Setting 4.1.1 and let $\theta \in \mathbb{R}^{\mathfrak{d}}$. Then

$$\langle (\nabla V)(\theta), \mathcal{G}(\theta) \rangle = 4L \left[\int_{[a,b]^{\ell_0}} \left\langle \mathcal{N}^{L,\theta}_{\infty}(x) - f(x), \mathcal{N}^{L,\theta}_{\infty}(x) - f(0) \right\rangle \mu(\mathrm{d}x) \right]. \tag{4.143}$$

Proof of Proposition 4.2.2. Throughout this proof let $(\mathbf{d}_k)_{k \in \mathbb{N}_0} \subseteq \mathbb{N}_0$ satisfy for all $k \in \mathbb{N}_0$ that $\mathbf{d}_k = \sum_{n=1}^{k} \ell_n(\ell_{n-1} + 1)$. Observe that item (iii) of Proposition 4.2.1 demonstrates that

$$\begin{aligned} \langle (\nabla V(\theta), \mathcal{G}(\theta) \rangle &= \sum_{k=1}^{L} \sum_{i_k=1}^{\ell_k} 2k \mathfrak{b}^{k,\theta}_{i_k} \mathcal{G}_{\ell_k \ell_{k-1} + i_k + \mathbf{d}_{k-1}}(\theta) \\ &+ \sum_{k=1}^{L} \sum_{i_k=1}^{\ell_k} \sum_{i_{k-1}=1}^{\ell_{k-1}} 2\mathfrak{w}^{k,\theta}_{i_k, i_{k-1}} \mathcal{G}_{(i_k - 1)\ell_{k-1} + i_{k-1} + \mathbf{d}_{k-1}}(\theta) \\ &- \sum_{i_L=1}^{\ell_L} 2L f_{i_L}(0) \mathcal{G}_{\ell_L \ell_{L-1} + i_L + \mathbf{d}_{L-1}}(\theta). \end{aligned} \tag{4.144}$$

Next we claim that for all $k \in \{1, \ldots, L\}$ it holds that

$$\begin{aligned} &\sum_{m=k}^{L} \sum_{i_m=1}^{\ell_m} \mathfrak{b}_{i_m}^{m,\theta} \mathcal{G}_{\ell_m \ell_{m-1}+i_m+\mathbf{d}_{m-1}}(\theta) + \sum_{i_k=1}^{\ell_k} \sum_{i_{k-1}=1}^{\ell_{k-1}} \mathfrak{w}_{i_k,i_{k-1}}^{k,\theta} \mathcal{G}_{(i_k-1)\ell_{k-1}+i_{k-1}+\mathbf{d}_{k-1}}(\theta) \\ &= 2 \int_{[a,b]^{\ell_0}} \left\langle \mathcal{N}_{\infty}^{L,\theta}(x) - f(x), \mathcal{N}_{\infty}^{L,\theta}(x) \right\rangle \mu(\mathrm{d}x). \end{aligned} \tag{4.145}$$

We prove (4.145) by induction on $k \in \{1, \ldots, L\}$. Note that items (v) and (vi) of Theorem 4.1.9 and (4.5) ensure that

$$\begin{aligned} &\sum_{i=1}^{\ell_L} \mathfrak{b}_i^{L,\theta} \mathcal{G}_{\ell_L \ell_{L-1}+i+\mathbf{d}_{L-1}}(\theta) + \sum_{i=1}^{\ell_L} \sum_{j=1}^{\ell_{L-1}} \mathfrak{w}_{i,j}^{L,\theta} \mathcal{G}_{(i-1)\ell_{L-1}+j+\mathbf{d}_{L-1}}(\theta) \\ &= \sum_{i=1}^{\ell_L} 2\mathfrak{b}_i^{L,\theta} \int_{[a,b]^{\ell_0}} (\mathcal{N}_{\infty,i}^{L,\theta}(x) - f_i(x)) \mu(\mathrm{d}x) \\ &\quad + \sum_{i=1}^{\ell_L} \sum_{j=1}^{\ell_{L-1}} 2\mathfrak{w}_{i,j}^{L,\theta} \int_{[a,b]^{\ell_0}} \left[\mathscr{R}_{\infty}(\mathcal{N}_{\infty,j}^{\max\{L-1,1\},\theta}(x)) \mathbb{1}_{(1,\infty)}(L) + x_j \mathbb{1}_{\{1\}}(L) \right] \\ &\qquad \cdot (\mathcal{N}_{\infty,i}^{L,\theta}(x) - f_i(x)) \mu(\mathrm{d}x) \\ &= 2 \int_{[a,b]^{\ell_0}} \left\langle \mathcal{N}_{\infty}^{L,\theta}(x) - f(x), \mathcal{N}_{\infty}^{L,\theta}(x) \right\rangle \mu(\mathrm{d}x). \end{aligned} \tag{4.146}$$

This establishes (4.145) in the base case $k = L$. For the induction step let $k \in \mathbb{N} \cap [2, L]$ satisfy

$$\begin{aligned} &\sum_{m=k}^{L} \sum_{i_m=1}^{\ell_m} \mathfrak{b}_{i_m}^{m,\theta} \mathcal{G}_{\ell_m \ell_{m-1}+i_m+\mathbf{d}_{m-1}}(\theta) + \sum_{i_k=1}^{\ell_k} \sum_{i_{k-1}=1}^{\ell_{k-1}} \mathfrak{w}_{i_k,i_{k-1}}^{k,\theta} \mathcal{G}_{(i_k-1)\ell_{k-1}+i_{k-1}+\mathbf{d}_{k-1}}(\theta) \\ &= 2 \int_{[a,b]^{\ell_0}} \left\langle \mathcal{N}_{\infty}^{L,\theta}(x) - f(x), \mathcal{N}_{\infty}^{L,\theta}(x) \right\rangle \mu(\mathrm{d}x). \end{aligned} \tag{4.147}$$

Observe that items (v) and (vi) of Theorem 4.1.9 show that

$$\begin{aligned}
&\sum_{m=k-1}^{L}\sum_{i_m=1}^{\ell_m} \mathfrak{b}^{m,\theta}_{i_m}\mathcal{G}_{\ell_m\ell_{m-1}+i_m+\mathbf{d}_{m-1}}(\theta) \\
&\quad + \sum_{i_{k-1}=1}^{\ell_{k-1}}\sum_{i_{k-2}=1}^{\ell_{k-2}} \mathfrak{w}^{k-1,\theta}_{i_{k-1},i_{k-2}}\mathcal{G}_{(i_{k-1}-1)\ell_{k-2}+i_{k-2}+\mathbf{d}_{k-2}}(\theta) \\
&= \sum_{m=k}^{L}\sum_{i_m=1}^{\ell_m} \mathfrak{b}^{m,\theta}_{i_m}\mathcal{G}_{\ell_m\ell_{m-1}+i_m+\mathbf{d}_{m-1}}(\theta) + \sum_{i_{k-1}=1}^{\ell_{k-1}} \mathfrak{b}^{k-1,\theta}_{i_{k-1}}\mathcal{G}_{\ell_{k-1}\ell_{k-2}+i_{k-1}+\mathbf{d}_{k-2}}(\theta) \\
&\quad + \sum_{i_{k-1}=1}^{\ell_{k-1}}\sum_{i_{k-2}=1}^{\ell_{k-2}} \mathfrak{w}^{k-1,\theta}_{i_{k-1},i_{k-2}}\mathcal{G}_{(i_{k-1}-1)\ell_{k-2}+i_{k-2}+\mathbf{d}_{k-2}}(\theta) \\
&= \sum_{m=k}^{L}\sum_{i_m=1}^{\ell_m} \mathfrak{b}^{m,\theta}_{i_m}\mathcal{G}_{\ell_m\ell_{m-1}+i_m+\mathbf{d}_{m-1}}(\theta) \\
&\quad + 2\sum_{i_{k-1}=1}^{\ell_{k-1}}\Bigg[\int_{[a,b]^{\ell_0}}\Bigg(\mathfrak{b}^{k-1,\theta}_{i_{k-1}} + \sum_{i_{k-2}=1}^{\ell_{k-2}}\mathfrak{w}^{k-1,\theta}_{i_{k-1},i_{k-2}} \\
&\quad \cdot\Big[\mathscr{R}_\infty(\mathcal{N}^{\max\{k-2,1\},\theta}_{\infty,i_{k-2}}(x))\mathbb{1}_{(1,L]}(k-1) + x_{i_{k-2}}\mathbb{1}_{\{1\}}(k-1)\Big]\Bigg) \\
&\quad \cdot\textstyle\sum_{\substack{v_{k-1},v_k,\ldots,v_L\in\mathbb{N},\\(\forall m\in\mathbb{N}\cap[k-1,L]\colon v_m\leq\ell_m)}}\Big[\big(\mathcal{N}^{L,\theta}_{\infty,v_L}(x)-f_{v_L}(x)\big) \\
&\quad \cdot\mathbb{1}_{\{i_{k-1}\}}(v_{k-1})\Big[\textstyle\prod_{q=k}^{L}\big(\mathfrak{w}^{q,\theta}_{v_q,v_{q-1}}\big[\mathbb{1}_{\mathcal{X}^{q-1,\theta}_{v_{q-1}}}(x)\big]\big)\Big]\Big]\,\mu(\mathrm{d}x)\Bigg].
\end{aligned} \tag{4.148}$$

Therefore, we obtain that

$$\begin{aligned}
&\sum_{m=k-1}^{L}\sum_{i_m=1}^{\ell_m} \mathfrak{b}^{m,\theta}_{i_m}\mathcal{G}_{\ell_m\ell_{m-1}+i_m+\mathbf{d}_{m-1}}(\theta) \\
&\quad + \sum_{i_{k-1}=1}^{\ell_{k-1}}\sum_{i_{k-2}=1}^{\ell_{k-2}} \mathfrak{w}^{k-1,\theta}_{i_{k-1},i_{k-2}}\mathcal{G}_{(i_{k-1}-1)\ell_{k-2}+i_{k-2}+\mathbf{d}_{k-2}}(\theta) \\
&= \sum_{m=k}^{L}\sum_{i_m=1}^{\ell_m} \mathfrak{b}^{m,\theta}_{i_m}\mathcal{G}_{\ell_m\ell_{m-1}+i_m+\mathbf{d}_{m-1}}(\theta) \\
&\quad + 2\sum_{i_{k-1}=1}^{\ell_{k-1}}\Bigg[\int_{[a,b]^{\ell_0}}\sum_{i_k=1}^{\ell_k}\mathfrak{w}^{k,\theta}_{i_k,i_{k-1}}\big[\mathbb{1}_{\mathcal{X}^{k-1,\theta}_{i_{k-1}}}(x)\big]\Bigg(\mathfrak{b}^{k-1,\theta}_{i_{k-1}} + \sum_{i_{k-2}=1}^{\ell_{k-2}}\mathfrak{w}^{k-1,\theta}_{i_{k-1},i_{k-2}} \\
&\quad \Big[\mathscr{R}_\infty(\mathcal{N}^{\max\{k-2,1\},\theta}_{\infty,i_{k-2}}(x))\mathbb{1}_{(1,L]}(k-1) + x_{i_{k-2}}\mathbb{1}_{\{1\}}(k-1)\Big]\Bigg) \\
&\quad \cdot\sum_{\substack{v_k,v_{k+1},\ldots,v_L\in\mathbb{N},\\ \forall w\in\mathbb{N}\cap[k,L]\colon v_w\leq\ell_w}}\Big[\mathcal{N}^{L,\theta}_{\infty,v_L}(x)-f_{v_L}(x)\Big]\Big[\mathbb{1}_{\{i_{k-1}\}}(v_k)\Big] \\
&\quad \cdot\Big[\textstyle\prod_{n=k+1}^{L}\big(\mathfrak{w}^{n,\theta}_{v_n,v_{n-1}}\big[\mathbb{1}_{\mathcal{X}^{n-1,\theta}_{v_{n-1}}}(x)\big]\big)\Big]\,\mu(\mathrm{d}x)\Bigg].
\end{aligned} \tag{4.149}$$

Item (v) of Theorem 4.1.9, (4.5), and (4.147) hence imply that

$$\begin{aligned}
&\sum_{m=k-1}^{L} \sum_{i_m=1}^{\ell_m} \mathfrak{b}_{i_m}^{m,\theta} \mathcal{G}_{\ell_m \ell_{m-1}+i_m+\mathbf{d}_{m-1}}(\theta) \\
&\quad + \sum_{i_{k-1}=1}^{\ell_{k-1}} \sum_{i_{k-2}=1}^{\ell_{k-2}} \mathfrak{w}_{i_{k-1},i_{k-2}}^{k-1,\theta} \mathcal{G}_{(i_{k-1}-1)\ell_{k-2}+i_{k-2}+\mathbf{d}_{k-2}}(\theta) \\
&= \sum_{m=k}^{L} \sum_{i_m=1}^{\ell_m} \mathfrak{b}_{i_m}^{m,\theta} \mathcal{G}_{\ell_m \ell_{m-1}+i_m+\mathbf{d}_{m-1}}(\theta) \\
&\quad + 2 \sum_{i_{k-1}=1}^{\ell_{k-1}} \Bigg[\int_{[a,b]^{\ell_0}} \sum_{i_k=1}^{\ell_k} \mathfrak{w}_{i_k,i_{k-1}}^{k,\theta} \big[\mathscr{R}_\infty(\mathcal{N}_{\infty,i_{k-1}}^{k-1,\theta}(x))\big] \\
&\quad \cdot \sum_{\substack{v_k, v_{k+1}, \ldots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \Big[\mathcal{N}_{\infty,v_L}^{L,\theta}(x) - f_{v_L}(x)\Big] \Big[\mathbb{1}_{\{i_{k-1}\}}(v_k)\Big] \\
&\quad \cdot \Big[\textstyle\prod_{n=k+1}^{L} \big(\mathfrak{w}_{v_n,v_{n-1}}^{n,\theta} \mathbb{1}_{\mathcal{X}_{v_{n-1}}^{n-1,\theta}}(x)\big)\Big] \, \mu(\mathrm{d}x) \Bigg] \\
&= \sum_{m=k}^{L} \sum_{i_m=1}^{\ell_m} \mathfrak{b}_{i_m}^{m,\theta} \mathcal{G}_{\ell_m \ell_{m-1}+i_m+\mathbf{d}_{m-1}}(\theta) + \sum_{i_k=1}^{\ell_k} \sum_{i_{k-1}=1}^{\ell_{k-1}} \mathfrak{w}_{i_k,i_{k-1}}^{k,\theta} \mathcal{G}_{(i_k-1)\ell_{k-1}+i_{k-1}+\mathbf{d}_{k-1}}(\theta) \\
&= 2 \int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) \big\rangle \, \mu(\mathrm{d}x).
\end{aligned} \tag{4.150}$$

Induction thus establishes (4.145). Next note that (4.145) ensures that

$$\begin{aligned}
&\sum_{k=1}^{L} k \sum_{i_k=1}^{\ell_k} \mathfrak{b}_{i_k}^{k,\theta} \mathcal{G}_{\ell_k \ell_{k-1}+i_k+\mathbf{d}_{k-1}}(\theta) + \sum_{k=1}^{L} \sum_{i_k=1}^{\ell_k} \sum_{i_{k-1}=1}^{\ell_{k-1}} \mathfrak{w}_{i_k,i_{k-1}}^{k,\theta} \mathcal{G}_{(i_k-1)\ell_{k-1}+i_{k-1}+\mathbf{d}_{k-1}}(\theta) \\
&= 2L \int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) \big\rangle \, \mu(\mathrm{d}x).
\end{aligned} \tag{4.151}$$

Combining this and (4.144) demonstrates that

$$\begin{aligned}
\langle (\nabla V(\theta), \mathcal{G}(\theta) \rangle &= \sum_{k=1}^{L} \sum_{i_k=1}^{\ell_k} 2k \mathfrak{b}_{i_k}^{k,\theta} \mathcal{G}_{\ell_k \ell_{k-1}+i_k+\mathbf{d}_{k-1}}(\theta) \\
&\quad + \sum_{k=1}^{L} \sum_{i_k=1}^{\ell_k} \sum_{i_{k-1}=1}^{\ell_{k-1}} 2\mathfrak{w}_{i_k,i_{k-1}}^{k,\theta} \mathcal{G}_{(i_k-1)\ell_{k-1}+i_{k-1}+\mathbf{d}_{k-1}}(\theta) \\
&\quad - \sum_{i_L=1}^{\ell_L} 2L f_{i_L}(0) \mathcal{G}_{\ell_L \ell_{L-1}+i_L+\mathbf{d}_{L-1}}(\theta) \\
&= 4L \int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) \big\rangle \, \mu(\mathrm{d}x) \\
&\quad - 4L \sum_{i_L=1}^{\ell_L} \left[\int_{[a,b]^{\ell_0}} f_{i_L}(0) (\mathcal{N}_{\infty,i_L}^{L,\theta}(x) - f_{i_L}(x)) \, \mu(\mathrm{d}x) \right] \\
&= 4L \int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) - f(0) \big\rangle \, \mu(\mathrm{d}x).
\end{aligned} \tag{4.152}$$

The proof of Proposition 4.2.2 is thus complete. □

Proposition 4.2.2 leads to the following corollary ensuring that the scalar product of the gradient of the Lyapunov function and the generalized gradient function of the risks is non-negative. This implies that V is indeed a Lyapunov function for the considered GF process. Corollary 4.2.3 in Subsection 4.2.1 extends [63, Corollary 2.10].

Corollary 4.2.3. Assume Setting 4.1.1, assume for all $x \in [a,b]^{\ell_0}$ that $f(x) = f(0)$, and let $\theta \in \mathbb{R}^{\mathfrak{d}}$. Then

$$\langle (\nabla V)(\theta), \mathcal{G}(\theta) \rangle = 4L\mathcal{L}_\infty(\theta)\,. \tag{4.153}$$

Proof of Corollary 4.2.3. Observe that Proposition 4.2.2 and the fact that for all $r \in [1,\infty]$ it holds that $\mathcal{L}_r(\theta) = \int_{[a,b]^{\ell_0}} \|\mathcal{N}_r^{L,\theta}(x) - f(0)\|^2 \mu(\mathrm{d}x)$ establish (4.153). The proof of Corollary 4.2.3 is thus complete. □

4.2.2 Weak chain rule for compositions of Lyapunov functions and GF processes

The following proposition obtains a weak chain rule for compositions of the Lyapunov function V and GF solutions.

Proposition 4.2.4. *Assume Setting 4.1.1 and let $T \in (0,\infty)$, $\Theta \in C([0,T],\mathbb{R}^{\mathfrak{d}})$ satisfy for all $t \in [0,T]$ that $\Theta_t = \Theta_0 - \int_0^t \mathcal{G}(\Theta_s)\,\mathrm{d}s$. Then it holds for all $t \in [0,T]$ that*

$$V(\Theta_t) = V(\Theta_0) - 4L \int_0^t \int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) - f(0) \big\rangle \, \mu(\mathrm{d}x)\,\mathrm{d}s. \tag{4.154}$$

Proof of Proposition 4.2.4. Note that Corollary 4.1.12 and the assumption that $\Theta \in C([0,T],\mathbb{R}^{\mathfrak{d}})$ ensure that $[0,T] \ni t \mapsto \mathcal{G}(\Theta_t) \in \mathbb{R}^{\mathfrak{d}}$ is bounded. Proposition 4.2.2 and, e.g., Cheridito et al. [24, Lemma 3.1] (applied with $T \curvearrowleft T$, $n \curvearrowleft \mathfrak{d}$, $F \curvearrowleft (\mathbb{R}^{\mathfrak{d}} \ni \theta \mapsto V(\theta) \in \mathbb{R})$, $\vartheta \curvearrowleft ([0,T] \ni t \mapsto \mathcal{G}(\Theta_t) \in \mathbb{R}^{\mathfrak{d}})$ in the notation of [24, Lemma 3.1]) therefore prove that for all $t \in [0,T]$ it holds that

$$\begin{aligned} V(\Theta_t) - V(\Theta_0) &= -\int_0^t \langle (\nabla V)(\Theta_s), \mathcal{G}(\Theta_s) \rangle \,\mathrm{d}s \\ &= -4L \int_0^t \int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) - f(0) \big\rangle \, \mu(\mathrm{d}x)\,\mathrm{d}s. \end{aligned} \tag{4.155}$$

The proof of Proposition 4.2.4 is thus complete. □

Under the assumption of a constant target function this leads to the following chain rule for compositions of the Lyapunov function V and GF solutions. Together with Proposition 4.2.4, Corollary 4.2.5 generalizes [24, Lemma 3.2].

Corollary 4.2.5. Assume Setting 4.1.1, assume for all $x \in [a,b]^{\ell_0}$ that $f(x) = f(0)$, and let $T \in (0,\infty)$, $\Theta \in C([0,T],\mathbb{R}^{\mathfrak{d}})$ satisfy for all $t \in [0,T]$ that $\Theta_t = \Theta_0 - \int_0^t \mathcal{G}(\Theta_s)\,\mathrm{d}s$. Then it holds for all $t \in [0,T]$ that

$$V(\Theta_t) = V(\Theta_0) - 4L \int_0^t \mathcal{L}_\infty(\Theta_s)\,\mathrm{d}s. \tag{4.156}$$

Proof of Corollary 4.2.5. Observe Proposition 4.2.4 and the fact that $\mathcal{L}_\infty(\theta) = \int_{[a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\theta}(x) - f(x)\|^2 \mu(\mathrm{d}x) = \int_{[a,b]^{\ell_0}} \langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) - f(0) \rangle \mu(\mathrm{d}x)$ establish (4.156). The proof of Corollary 4.2.5 is thus complete. □

4.2.3 Weak chain rule for the risk of GF processes

The following lemma provides uniform local boundedness of the gradients of the approximated risk function. In particular, Lemma 4.2.6 is an expansion of [24, Lemma 3.4].

Lemma 4.2.6. *Assume Setting 4.1.1 and let $K \subseteq \mathbb{R}^{\mathfrak{d}}$ be compact. Then*

$$\sup\nolimits_{\theta \in K} \sup\nolimits_{r \in [1,\infty)} \|(\nabla \mathcal{L}_r)(\theta)\| < \infty. \tag{4.157}$$

Proof of Lemma 4.2.6. Throughout this proof assume without loss of generality that $\mathfrak{m} > 0$ and for every $k \in \mathbb{N}_0$ let $\mathbf{d}_k \in \mathbb{N}_0$ satisfy $\mathbf{d}_k = \sum_{n=1}^{k} \ell_n(\ell_{n-1}+1)$. Note that Lemma 4.1.7 ensures that there exists $\mathfrak{D} \in [1,\infty)$ which satisfies for all $k \in \{1,\dots,L\}$ that

$$\begin{split}\mathfrak{D} \geqslant \mathbf{a} + \sup_{\theta \in K} \sup_{r,s,t \in [1,\infty)} \sup_{i \in \{1,\dots,\ell_k\}} \sup_{x \in [a,b]^{\ell_0}} \Big(|\mathcal{N}^{k,\theta}_{r,i}(x)| \\ + |\mathscr{R}_s(\mathcal{N}^{k,\theta}_{r,i}(x))| + |(\mathscr{R}_t)'(\mathcal{N}^{k,\theta}_{r,i}(x))| \Big).\end{split} \tag{4.158}$$

Furthermore, observe that item (iii) of Theorem 4.1.9 proves that for all $\theta \in K$, $r \in [1,\infty)$, $k \in \{1,\dots,L\}$, $i \in \{1,\dots,\ell_k\}$ it holds that

$$\begin{split}&\left|\Big(\frac{\partial \mathcal{L}_r}{\partial \theta_{\ell_k \ell_{k-1}+i+\mathbf{d}_{k-1}}}\Big)(\theta)\right|^2 \\
&= \Bigg(\sum_{\substack{v_k, v_{k+1}, \dots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \int_{[a,b]^{\ell_0}} 2 \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathcal{N}^{L,\theta}_{r,v_L}(x) - f_{v_L}(x)\Big] \\
&\quad \cdot \Big[\textstyle\prod_{n=k+1}^{L} \big(\mathfrak{w}^{n,\theta}_{v_n,v_{n-1}}\big[(\mathscr{R}_{r^{1/(n-1)}})'(\mathcal{N}^{n-1,\theta}_{r,v_{n-1}}(x))\big]\big)\Big]\,\mu(\mathrm{d}x)\Bigg)^2 \\
&\leqslant 4\mathfrak{D}^{2L} \Bigg(\int_{[a,b]^{\ell_0}} \|\mathcal{N}^{L,\theta}_r(x) - f(x)\| \\
&\quad \cdot \Bigg[\sum_{\substack{v_k, v_{k+1}, \dots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{L} |\mathfrak{w}^{n,\theta}_{v_n,v_{n-1}}|\Big]\Bigg]\,\mu(\mathrm{d}x)\Bigg)^2 \\
&= 4\mathfrak{m}^2\mathfrak{D}^{2L} \Bigg(\Bigg[\sum_{\substack{v_k, v_{k+1}, \dots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{L} |\mathfrak{w}^{n,\theta}_{v_n,v_{n-1}}|\Big]\Bigg] \\
&\quad \cdot \frac{1}{\mathfrak{m}} \int_{[a,b]^{\ell_0}} \|\mathcal{N}^{L,\theta}_r(x) - f(x)\|\,\mu(\mathrm{d}x)\Bigg)^2.\end{split} \tag{4.159}$$

Jensen's inequality hence shows for all $\theta \in K$, $r \in [1,\infty)$, $k \in \{1,\dots,L\}$, $i \in \{1,\dots,\ell_k\}$

that

$$
\begin{aligned}
&\left|\left(\frac{\partial \mathcal{L}_r}{\partial \theta_{\ell_k \ell_{k-1}+i+\mathbf{d}_{k-1}}}\right)(\theta)\right|^2 \\
&\leqslant 4\mathfrak{m}^2 \mathfrak{D}^{2L} \Bigg[\sum_{\substack{v_k, v_{k+1}, \ldots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{L} |\mathfrak{w}^{n,\theta}_{v_n, v_{n-1}}|\Big]\Bigg]^2 \\
&\qquad \cdot \frac{1}{\mathfrak{m}} \int_{[a,b]^{\ell_0}} \|\mathcal{N}^{L,\theta}_r(x) - f(x)\|^2 \, \mu(\mathrm{d}x) \\
&= 4\mathfrak{m}\mathfrak{D}^{2L} \Bigg[\sum_{\substack{v_k, v_{k+1}, \ldots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{L} |\mathfrak{w}^{n,\theta}_{v_n, v_{n-1}}|\Big]\Bigg]^2 \mathcal{L}_r(\theta).
\end{aligned}
\tag{4.160}
$$

Moreover, note that item (ii) of Theorem 4.1.9 demonstrates for all $\theta \in K$, $r \in [1, \infty)$, $k \in \{1, \ldots, L\}$, $i \in \{1, \ldots, \ell_k\}$, $j \in \{1, \ldots, \ell_{k-1}\}$ that

$$
\begin{aligned}
&\left|\left(\frac{\partial \mathcal{L}_r}{\partial \theta_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}}\right)(\theta)\right|^2 \\
&= \Bigg(\sum_{\substack{v_k, v_{k+1}, \ldots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \int_{[a,b]^{\ell_0}} 2 \Big[\mathscr{R}_{r^{1/(\max\{k-1,1\})}}(\mathcal{N}^{\max\{k-1,1\},\theta}_{r,j}(x)) \mathbb{1}_{(1,L]}(k) \\
&\qquad + x_j \mathbb{1}_{\{1\}}(k)\Big]\Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathcal{N}^{L,\theta}_{r,v_L}(x) - f_{v_L}(x)\Big] \\
&\qquad \cdot \Big[\textstyle\prod_{n=k+1}^{L} \big(\mathfrak{w}^{n,\theta}_{v_n, v_{n-1}} \big[(\mathscr{R}_{r^{1/(n-1)}})'(\mathcal{N}^{n-1,\theta}_{r,v_{n-1}}(x))\big]\big)\Big] \, \mu(\mathrm{d}x) \Bigg)^2 \\
&\leqslant 4\mathfrak{D}^{2L} \Bigg(\int_{[a,b]^{\ell_0}} \|\mathcal{N}^{L,\theta}_r(x) - f(x)\| \\
&\qquad \cdot \Bigg[\sum_{\substack{v_k, v_{k+1}, \ldots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{L} |\mathfrak{w}^{n,\theta}_{v_n, v_{n-1}}|\Big]\Bigg] \, \mu(\mathrm{d}x) \Bigg)^2 \\
&= 4\mathfrak{m}^2 \mathfrak{D}^{2L} \Bigg(\Bigg[\sum_{\substack{v_k, v_{k+1}, \ldots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k,L] \colon v_w \leqslant \ell_w}} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{L} |\mathfrak{w}^{n,\theta}_{v_n, v_{n-1}}|\Big]\Bigg] \\
&\qquad \cdot \frac{1}{\mathfrak{m}} \int_{[a,b]^{\ell_0}} \|\mathcal{N}^{L,\theta}_\infty(x) - f(x)\| \, \mu(\mathrm{d}x) \Bigg)^2.
\end{aligned}
\tag{4.161}
$$

Jensen's inequality therefore proves that for all $\theta \in K$, $r \in [1, \infty)$, $k \in \{1, \ldots, L\}$, $i \in$

$\{1,\dots,\ell_k\}$, $j\in\{1,\dots,\ell_{k-1}\}$ we have that

$$\begin{aligned}
&\left|\Big(\frac{\partial \mathcal{L}_r}{\partial\theta_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}}\Big)(\theta)\right|^2\\
&\leqslant 4\mathfrak{m}^2\mathfrak{D}^{2L}\left[\sum_{\substack{v_k,v_{k+1},\dots,v_L\in\mathbb{N},\\ \forall w\in\mathbb{N}\cap[k,L]\colon v_w\leqslant\ell_w}}\Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{L}|\mathfrak{w}^{n,\theta}_{v_n,v_{n-1}}|\Big]\right]^2\\
&\quad\cdot\frac{1}{\mathfrak{m}}\int_{[a,b]^{\ell_0}}\|\mathcal{N}^{L,\theta}_{\infty}(x)-f(x)\|^2\,\mu(\mathrm{d}x)\\
&=4\mathfrak{m}\mathfrak{D}^{2L}\left[\sum_{\substack{v_k,v_{k+1},\dots,v_L\in\mathbb{N},\\ (\forall w\in\mathbb{N}\cap[k,L]\colon v_w\leqslant\ell_w)}}\Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{L}|\mathfrak{w}^{n,\theta}_{v_n,v_{n-1}}|\Big]\right]^2\mathcal{L}_r(\theta).
\end{aligned}\tag{4.162}$$

This and (4.160) assure for all $\theta\in K$, $r\in[1,\infty)$ that

$$\begin{aligned}
&\|(\nabla\mathcal{L}_r)(\theta)\|^2\\
&=\sum_{k=1}^{L}\sum_{i=1}^{\ell_k}\left[\left|\Big(\frac{\partial\mathcal{L}_r}{\partial\theta_{\ell_k\ell_{k-1}+i+\mathbf{d}_{k-1}}}\Big)(\theta)\right|^2+\sum_{j=1}^{\ell_{k-1}}\left|\Big(\frac{\partial\mathcal{L}_r}{\partial\theta_{(i-1)\ell_{k-1}+j+\mathbf{d}_{k-1}}}\Big)(\theta)\right|^2\right]\\
&\leqslant 4\mathfrak{m}\mathfrak{D}^{2L}\left[\sum_{k=1}^{L}\sum_{i=1}^{\ell_k}(\ell_{k-1}+1)\right.\\
&\quad\cdot\left.\sum_{\substack{v_k,v_{k+1},\dots,v_L\in\mathbb{N},\\ \forall w\in\mathbb{N}\cap[k,L]\colon v_w\leqslant\ell_w}}\Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\textstyle\prod_{n=k+1}^{L}|\mathfrak{w}^{n,\theta}_{v_n,v_{n-1}}|\Big]\right]^2\mathcal{L}_r(\theta).
\end{aligned}\tag{4.163}$$

Furthermore, observe that (4.6) implies for all $\theta\in K$, $r\in[1,\infty)$ that

$$\begin{aligned}
\mathcal{L}_r(\theta)&=\int_{[a,b]^{\ell_0}}\|\mathcal{N}^{L,\theta}_r(x)-f(x)\|^2\mu(\mathrm{d}x)\\
&\leqslant 2\int_{[a,b]^{\ell_0}}\Big[\|\mathcal{N}^{L,\theta}_r(x)\|^2+\|f(x)\|^2\Big]\mu(\mathrm{d}x)\\
&\leqslant 2\mathfrak{m}\Big[\sup\nolimits_{y\in[a,b]^{\ell_0}}\|\mathcal{N}^{L,\theta}_r(y)\|^2\Big]+2\int_{[a,b]^{\ell_0}}\|f(x)\|^2\mu(\mathrm{d}x).
\end{aligned}\tag{4.164}$$

This, Lemma 4.1.8, and (4.158) prove that $\sup_{\theta\in K}\sup_{r\in[1,\infty)}\mathcal{L}_r(\theta)<\infty$. Combining this with (4.163) and item (iii) of Theorem 4.1.11 establishes (4.157). The proof of Lemma 4.2.6 is thus complete. □

The following proposition delivers a chain rule for the composition of the risk function and GF solutions in the setting of a general measurable target function. This means Proposition 4.2.7 extends [24, Lemma 3.5] to DNNs.

Proposition 4.2.7. *Assume Setting 4.1.1 and let $T\in(0,\infty)$, $\Theta\in C([0,T],\mathbb{R}^{\mathfrak{d}})$ satisfy for all $t\in[0,T]$ that $\Theta_t=\Theta_0-\int_0^t\mathcal{G}(\Theta_s)\,\mathrm{d}s$. Then it holds for all $t\in[0,T]$ that*

$$\mathcal{L}_\infty(\Theta_t)=\mathcal{L}_\infty(\Theta_0)-\int_0^t\|\mathcal{G}(\Theta_s)\|^2\,\mathrm{d}s.\tag{4.165}$$

Proof of Proposition 4.2.7. Note that, e.g., Cheridito et al. [24, Lemma 3.1] (applied with $T \curvearrowleft T$, $n \curvearrowleft \mathfrak{d}$, $F \curvearrowleft (\mathbb{R}^{\mathfrak{d}} \ni \theta \mapsto \mathcal{L}_r(\theta) \in \mathbb{R})$, $\vartheta \curvearrowleft ([0,T] \ni t \mapsto \mathcal{G}(\Theta_t) \in \mathbb{R}^{\mathfrak{d}})$ in the notation of [24, Lemma 3.1]) implies that for all $r \in [1,\infty)$, $t \in [0,T]$ we have that

$$\mathcal{L}_r(\Theta_t) - \mathcal{L}_r(\Theta_0) = -\int_0^t \langle (\nabla \mathcal{L}_r(\Theta_s)), \mathcal{G}(\Theta_s) \rangle \, ds. \tag{4.166}$$

Furthermore, observe that item (iv) of Theorem 4.1.9 ensures that for all $t \in [0,T]$ it holds that $\lim_{r\to\infty} (\mathcal{L}_r(\Theta_t) - \mathcal{L}_r(\Theta_0)) = \mathcal{L}_\infty(\Theta_t) - \mathcal{L}_\infty(\Theta_0)$ and $\lim_{r\to\infty} \langle (\nabla \mathcal{L}_r)(\Theta_t), \mathcal{G}(\Theta_t) \rangle = \langle (\mathcal{G}(\Theta_t), \mathcal{G}(\Theta_t) \rangle = \|\mathcal{G}(\Theta_t)\|^2$. Moreover, note that the assumption that $\Theta \in C([0,T], \mathbb{R}^{\mathfrak{d}})$ assures that there exists a compact set $K \subseteq \mathbb{R}^{\mathfrak{d}}$ such that for all $t \in [0,T]$ it holds that $\Theta_t \in K$. This, the Cauchy-Schwarz inequality, Corollary 4.1.12, item (i) of Theorem 4.1.9, and Lemma 4.2.6 hence demonstrate that

$$\begin{split} &\sup\nolimits_{r\in[1,\infty)} \sup\nolimits_{t\in[0,T]} \big|\langle (\nabla \mathcal{L}_r)(\Theta_t), \mathcal{G}(\Theta_t) \rangle\big| \leqslant \sup\nolimits_{r\in[1,\infty)} \sup\nolimits_{\theta\in K} \big|\langle (\nabla \mathcal{L}_r)(\theta), \mathcal{G}(\theta) \rangle\big| \\ &\leqslant \sup\nolimits_{r\in[1,\infty)} \sup\nolimits_{\theta\in K} \big(\|(\nabla \mathcal{L}_r)(\theta)\| \|\mathcal{G}(\theta)\|\big) < \infty. \end{split} \tag{4.167}$$

The dominated convergence theorem and item (iv) of Theorem 4.1.9 therefore show that for all $t \in [0,T]$ it holds that

$$\begin{split} &\lim_{r\to\infty} \left[\int_0^t \langle (\nabla \mathcal{L}_\infty)(\Theta_s), \mathcal{G}(\Theta_s) \rangle \, ds \right] \\ &= \int_0^t \left[\lim_{r\to\infty} \langle (\nabla \mathcal{L}_\infty)(\Theta_s), \mathcal{G}(\Theta_s) \rangle \right] ds = \int_0^t \|\mathcal{G}(\Theta_s)\|^2 \, ds. \end{split} \tag{4.168}$$

Combining this with (4.166) establishes (4.165). The proof of Proposition 4.2.7 is thus complete. □

4.2.4 Convergence analysis for GF processes

In the following theorem we prove - under the assumption of a constant target function - that the risk of every solution of the associated GF differential equation converges to zero with rate 1. Theorem 4.2.8 is a generalization of [24, Theorem 3.7].

Theorem 4.2.8. *Assume Setting 4.1.1, assume for all $x \in [a,b]^{\ell_0}$ that $f(x) = f(0)$, and let $\Theta \in C([0,\infty), \mathbb{R}^{\mathfrak{d}})$ satisfy for all $t \in [0,\infty)$ that $\Theta_t = \Theta_0 - \int_0^t \mathcal{G}(\Theta_s) \, ds$. Then*

(i) it holds that $\sup_{t\in[0,\infty)} \|\Theta_t\| \leqslant [2V(\Theta_0) + 4L^2 \|f(0)\|^2]^{1/2} < \infty$,

(ii) it holds for all $t \in (0,\infty)$ that $\mathcal{L}_\infty(\Theta_t) \leqslant \frac{1}{2t}[\|\Theta_0\|^2 + 2L\|f(0)\|^2] < \infty$, and

(iii) it holds that $\limsup_{t\to\infty} \mathcal{L}_\infty(\Theta_t) = 0$.

Proof of Theorem 4.2.8. Observe that item (ii) of Proposition 4.2.1 implies that for all $t \in [0,\infty)$ it holds that $\|\Theta_t\| \leqslant \big[2V(\Theta_t) + 4L^2\|f(0)\|^2\big]^{1/2}$. Moreover, note that Corollary 4.2.5 and the fact that for all $\theta \in \mathbb{R}^{\mathfrak{d}}$ it holds that $\mathcal{L}_\infty(\theta) \geqslant 0$ prove that for all $t \in [0,\infty)$ we have that $V(\Theta_t) \leqslant V(\Theta_0)$. This establishes item (i). Next observe that Proposition 4.2.7 implies that $[0,\infty) \ni t \mapsto \mathcal{L}_\infty(\Theta_t) \in [0,\infty)$ is non-increasing. Combining this with Corollary 4.2.5 and item (ii) of Proposition 4.2.1 demonstrates that for all $t \in [0,\infty)$ it holds that

$$\begin{split} t\mathcal{L}_\infty(\Theta_t) &= \int_0^t \mathcal{L}_\infty(\Theta_t) \, ds \leqslant \int_0^t \mathcal{L}_\infty(\Theta_s) \, ds = \frac{1}{4L}[V(\Theta_0) - V(\Theta_t)] \\ &\leqslant \frac{1}{4L}\Big[2L\|\Theta_0\|^2 + L\|f(0)\|^2 - \frac{1}{2}\|\Theta_t\|^2 + 2L^2\|f(0)\|^2\Big] \\ &\leqslant \frac{1}{4L}\Big[2L\|\Theta_0\|^2 + (L + 2L^2)\|f(0)\|^2\Big] \leqslant \frac{1}{2}\|\Theta_0\|^2 + L\|f(0)\|^2 < \infty. \end{split} \tag{4.169}$$

Hence, we obtain for all $t \in (0, \infty)$ that

$$\mathcal{L}_\infty(\Theta_t) \leqslant \frac{1}{2t}\Big[\|\Theta_0\|^2 + 2L\|f(0)\|^2\Big]. \tag{4.170}$$

This establishes items (ii) and (iii). The proof of Theorem 4.2.8 is thus complete. □

4.3 Gradient descent (GD) processes in the training of deep ANNs

In this section we study GD processes in the training of DNNs with an arbitrary number of hidden layers. We start by establishing one-step Lyapunov estimates in Subsection 4.3.1. Subsection 4.3.2 uses these results to obtain certain time-discrete upper estimates for composition of the considered Lyapunov function and GD processes. This sections main result is Theorem 4.3.7 in Subsection 4.3.3 below. Theorem 4.3.7 proves that in the training of DNNs it holds that the sequence of risks of any time-discrete GD process converges to zero provided that the target function is constant and the step sizes are sufficiently small but not L^1-summable.

4.3.1 Lyapunov type estimates for the dynamics of GD processes

The following lemma provides a one-step Lyapunov estimate for the GD method.

Lemma 4.3.1. *Assume Setting 4.1.1 and let $\gamma \in \mathbb{R}$, $\theta \in \mathbb{R}^{\mathfrak{d}}$. Then*

$$\begin{aligned} &V(\theta - \gamma\mathcal{G}(\theta)) - V(\theta) \\ &= \gamma^2\|\mathcal{G}(\theta)\|^2 + \gamma^2\big[\textstyle\sum_{k=1}^{L}\sum_{i=1}^{\ell_k}(k-1)|\mathcal{G}_{\ell_k\ell_{k-1}+i+\sum_{h=1}^{k-1}\ell_h(\ell_{h-1}+1)}(\theta)|^2\big] \\ &\quad - 4\gamma L\big[\textstyle\int_{[a,b]^{\ell_0}}\Big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) - f(0)\Big\rangle\,\mu(\mathrm{d}x)\big] \\ &\leqslant \gamma^2 L\|\mathcal{G}(\theta)\|^2 - 4\gamma L\big[\textstyle\int_{[a,b]^{\ell_0}}\Big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) - f(0)\Big\rangle\,\mu(\mathrm{d}x)\big]. \end{aligned} \tag{4.171}$$

Proof of Lemma 4.3.1. Throughout this proof let $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{\mathfrak{d}} \in \mathbb{R}^{\mathfrak{d}}$ satisfy $\mathbf{e}_1 = (1, 0, \dots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \dots, 0)$, ..., $\mathbf{e}_{\mathfrak{d}} = (0, \dots, 0, 1)$ and let $g\colon \mathbb{R} \to \mathbb{R}$ satisfy for all $t \in \mathbb{R}$ that

$$g(t) = V(\theta - t\mathcal{G}(\theta)). \tag{4.172}$$

Note that (4.172) and the fundamental theorem of calculus demonstrate that

$$\begin{aligned} V(\theta - \gamma\mathcal{G}(\theta)) = g(\gamma) &= g(0) + \int_0^\gamma g'(t)\,\mathrm{d}t \\ &= g(0) + \int_0^\gamma \langle(\nabla V)(\theta - t\mathcal{G}(\theta)), (-\mathcal{G}(\theta))\rangle\,\mathrm{d}t \\ &= V(\theta) - \int_0^\gamma \langle(\nabla V)(\theta - t\mathcal{G}(\theta)), \mathcal{G}(\theta)\rangle\,\mathrm{d}t. \end{aligned} \tag{4.173}$$

Proposition 4.2.2 therefore proves that

$$\begin{aligned} &V(\theta - \gamma \mathcal{G}(\theta)) \\ &= V(\theta) - \int_0^\gamma \langle (\nabla V)(\theta), \mathcal{G}(\theta) \rangle \, \mathrm{d}t + \int_0^\gamma \langle (\nabla V)(\theta) - (\nabla V)(\theta - t\mathcal{G}(\theta)), \mathcal{G}(\theta) \rangle \, \mathrm{d}t \\ &= V(\theta) - 4\gamma L \bigg[\int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) - f(0) \big\rangle \, \mu(\mathrm{d}x) \bigg] \\ &\quad + \int_0^\gamma \langle (\nabla V)(\theta) - (\nabla V)(\theta - t\mathcal{G}(\theta)), \mathcal{G}(\theta) \rangle \, \mathrm{d}t. \end{aligned} \tag{4.174}$$

Moreover, observe that item (iii) of Proposition 4.2.1 implies that for all $t \in \mathbb{R}$ it holds that

$$\begin{aligned} &(\nabla V)(\theta) - (\nabla V)(\theta - t\mathcal{G}(\theta)) \\ &= 2t\mathcal{G}(\theta) + 2\Bigg[\sum_{k=1}^{L} \sum_{i=1}^{\ell_k} (k-1) \mathfrak{b}_i^{k,t\mathcal{G}(\theta)} \mathbf{e}_{\ell_k \ell_{k-1} + i + \sum_{h=1}^{k-1} \ell_h(\ell_{h-1}+1)} \Bigg]. \end{aligned} \tag{4.175}$$

Combining this with (4.174) shows that

$$\begin{aligned} &V(\theta - \gamma \mathcal{G}(\theta)) \\ &= V(\theta) - 4\gamma L \bigg[\int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) - f(0) \big\rangle \, \mu(\mathrm{d}x) \bigg] \\ &\quad + \int_0^\gamma \Bigg\langle 2t\mathcal{G}(\theta) + 2\Bigg[\sum_{k=1}^{L} \sum_{i=1}^{\ell_k} (k-1) \mathfrak{b}_i^{k,t\mathcal{G}(\theta)} \mathbf{e}_{\ell_k \ell_{k-1} + i + \sum_{h=1}^{k-1} \ell_h(\ell_{h-1}+1)} \Bigg], \mathcal{G}(\theta) \Bigg\rangle \, \mathrm{d}t \\ &= V(\theta) - 4\gamma L \bigg[\int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) - f(0) \big\rangle \, \mu(\mathrm{d}x) \bigg] \\ &\qquad + 2\|\mathcal{G}(\theta)\|^2 \bigg[\int_0^\gamma t \, \mathrm{d}t \bigg] \\ &\quad + 2\Bigg[\int_0^\gamma \Bigg\langle \sum_{k=1}^{L} \sum_{i=1}^{\ell_k} (k-1) \mathfrak{b}_i^{k,t\mathcal{G}(\theta)} \mathbf{e}_{\ell_k \ell_{k-1} + i + \sum_{h=1}^{k-1} \ell_h(\ell_{h-1}+1)}, \mathcal{G}(\theta) \Bigg\rangle \, \mathrm{d}t \Bigg] \\ &= V(\theta) - 4\gamma L \bigg[\int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) - f(0) \big\rangle \, \mu(\mathrm{d}x) \bigg] + \gamma^2 \|\mathcal{G}(\theta)\|^2 \\ &\quad + 2\Bigg[\sum_{k=1}^{L} \sum_{i=1}^{\ell_k} (k-1) \big| \big\langle \mathbf{e}_{\ell_k \ell_{k-1} + i + \sum_{h=1}^{k-1} \ell_h(\ell_{h-1}+1)}, \mathcal{G}(\theta) \big\rangle \big|^2 \Bigg] \bigg[\int_0^\gamma t \, \mathrm{d}t \bigg]. \end{aligned} \tag{4.176}$$

Hence, we obtain that

$$\begin{aligned} &V(\theta - \gamma \mathcal{G}(\theta)) \\ &= V(\theta) - 4\gamma L \bigg[\int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) - f(0) \big\rangle \, \mu(\mathrm{d}x) \bigg] + \gamma^2 \|\mathcal{G}(\theta)\|^2 \\ &\quad + \gamma^2 \Bigg[\sum_{k=1}^{L} \sum_{i=1}^{\ell_k} (k-1) \big| \big\langle \mathbf{e}_{\ell_k \ell_{k-1} + i + \sum_{h=1}^{k-1} \ell_h(\ell_{h-1}+1)}, \mathcal{G}(\theta) \big\rangle \big|^2 \Bigg] \\ &= V(\theta) - 4\gamma L \bigg[\int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) - f(0) \big\rangle \, \mu(\mathrm{d}x) \bigg] + \gamma^2 \|\mathcal{G}(\theta)\|^2 \\ &\quad + \gamma^2 \Bigg[\sum_{k=1}^{L} \sum_{i=1}^{\ell_k} (k-1) |\mathcal{G}_{\ell_k \ell_{k-1} + i + \sum_{h=1}^{k-1} \ell_h(\ell_{h-1}+1)}(\theta)|^2 \Bigg]. \end{aligned} \tag{4.177}$$

The proof of Lemma 4.3.1 is thus complete. □

Lemma 4.3.1 leads to the following one-step Lyapunov estimate in the case of a constant target function. Together with Lemma 4.3.1, Corollary 4.3.2 extends [63, Lemma 2.12] to the setting of deep ANNs.

Corollary 4.3.2. Assume Setting 4.1.1, assume for all $x \in [a,b]^{\ell_0}$ that $f(x) = f(0)$, and let $\gamma \in \mathbb{R}$, $\theta \in \mathbb{R}^{\mathfrak{d}}$. Then

$$\begin{aligned} &V(\theta - \gamma \mathcal{G}(\theta)) - V(\theta) \\ &= \gamma^2 \|\mathcal{G}(\theta)\|^2 + \gamma^2 \big[\textstyle\sum_{k=1}^{L} \sum_{i=1}^{\ell_k} (k-1) |\mathcal{G}_{\ell_k \ell_{k-1} + i + \sum_{h=1}^{k-1} \ell_h(\ell_{h-1}+1)}(\theta)|^2 \big] - 4\gamma L \mathcal{L}_\infty(\theta) \\ &\leqslant \gamma^2 L \|\mathcal{G}(\theta)\|^2 - 4\gamma L \mathcal{L}_\infty(\theta). \end{aligned} \tag{4.178}$$

Proof of Corollary 4.3.2. Note that Lemma 4.3.1, (4.6), and the assumption that for all $x \in [a,b]^{\ell_0}$ it holds that $f(x) = f(0)$ establish (4.178). The proof of Corollary 4.3.2 is thus complete. □

Additionally, combining Theorem 4.1.11 with Lemma 4.3.1 proves the following corollary. Corollary 4.3.3 is a generalization of [63, Corollary 2.13].

Corollary 4.3.3. Assume Setting 4.1.1 and let $\gamma \in [0,\infty)$, $\theta \in \mathbb{R}^{\mathfrak{d}}$. Then

$$\begin{aligned} &V(\theta - \gamma \mathcal{G}(\theta)) - V(\theta) \\ &\leqslant 4\gamma^2 \mathfrak{m} L^2 \mathbf{a}^2 \big[\textstyle\prod_{p=0}^{L} (\ell_p + 1) \big] (2V(\theta) + 4L^2 \|f(0)\|^2 + 1)^{(L-1)} \mathcal{L}_\infty(\theta) \\ &\quad - 4\gamma L \big[\textstyle\int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\theta}(x) - f(x), \mathcal{N}_\infty^{L,\theta}(x) - f(0) \big\rangle \, \mu(\mathrm{d}x) \big]. \end{aligned} \tag{4.179}$$

Proof of Corollary 4.3.3. Observe that item (iv) of Theorem 4.1.11 and item (ii) of Proposition 4.2.1 demonstrate that

$$\begin{aligned} \|\mathcal{G}(\theta)\|^2 &\leqslant 4\mathfrak{m} L \mathbf{a}^2 \big[\textstyle\prod_{p=0}^{L} (\ell_p + 1) \big] (\|\theta\|^2 + 1)^{(L-1)} \mathcal{L}_\infty(\theta) \\ &\leqslant 4\mathfrak{m} L \mathbf{a}^2 \big[\textstyle\prod_{p=0}^{L} (\ell_p + 1) \big] (2V(\theta) + 4L^2 \|f(0)\|^2 + 1)^{(L-1)} \mathcal{L}_\infty(\theta). \end{aligned} \tag{4.180}$$

Combining this with Lemma 4.3.1 establishes (4.179). The proof of Corollary 4.3.3 is thus complete. □

4.3.2 Upper estimates for compositions of Lyapunov functions and GD processes

The following corollary establishes a time-discrete upper estimate for compositions of the Lyapunov function and GD processes.

Corollary 4.3.4. Assume Setting 4.1.1, let $(\gamma_n)_{n \in \mathbb{N}_0} \subseteq [0,\infty)$, let $(\Theta_n)_{n \in \mathbb{N}_0} \colon \mathbb{N}_0 \to \mathbb{R}^{\mathfrak{d}}$ satisfy for all $n \in \mathbb{N}_0$ that $\Theta_{n+1} = \Theta_n - \gamma_n \mathcal{G}(\Theta_n)$, and let $n \in \mathbb{N}_0$. Then

$$\begin{aligned} &V(\Theta_{n+1}) - V(\Theta_n) \\ &\leqslant 4(\gamma_n)^2 \mathfrak{m} L^2 \mathbf{a}^2 \big[\textstyle\prod_{p=0}^{L} (\ell_p + 1) \big] (2V(\Theta_n) + 4L^2 \|f(0)\|^2 + 1)^{(L-1)} \mathcal{L}_\infty(\Theta_n) \\ &\quad - 4\gamma_n L \big[\textstyle\int_{[a,b]^{\ell_0}} \big\langle \mathcal{N}_\infty^{L,\Theta_n}(x) - f(x), \mathcal{N}_\infty^{L,\Theta_n}(x) - f(0) \big\rangle \, \mu(\mathrm{d}x) \big]. \end{aligned} \tag{4.181}$$

Proof of Corollary 4.3.4. Note that Corollary 4.3.3 establishes (4.181). The proof of Corollary 4.3.4 is thus complete. □

Corollary 4.3.4 immediately leads to the following time-discrete upper estimate for the composition of Lyapunov functions and GD processes under the assumption of a constant target function.

Corollary 4.3.5. Assume Setting 4.1.1, assume for all $x \in [a,b]^d$ that $f(x) = f(0)$, let $(\gamma_n)_{n\in\mathbb{N}_0} \subseteq [0,\infty)$, let $(\Theta_n)_{n\in\mathbb{N}_0} \colon \mathbb{N}_0 \to \mathbb{R}^{\mathfrak{d}}$ satisfy for all $n \in \mathbb{N}_0$ that $\Theta_{n+1} = \Theta_n - \gamma_n \mathcal{G}(\Theta_n)$, and let $n \in \mathbb{N}_0$. Then

$$\begin{aligned} &V(\Theta_{n+1}) - V(\Theta_n) \\ &\leqslant 4L\big((\gamma_n)^2 \mathfrak{m} L \mathbf{a}^2 \big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big] \\ &\quad \cdot (2V(\Theta_n) + 4L^2\|f(0)\|^2 + 1)^{(L-1)} - \gamma_n\big)\mathcal{L}_\infty(\Theta_n). \end{aligned} \tag{4.182}$$

Proof of Corollary 4.3.5. Observe that Corollary 4.3.4 and the assumption that for all $x \in [a,b]^{\ell_0}$ it holds that $f(x) = f(0)$ establish (4.182). The proof of Corollary 4.3.5 is thus complete. □

From this follows the following final upper estimate for the composition of the Lyapunov functions and the time-discrete GD process in the setting of a constant target function. Corollary 4.3.4, Corollary 4.3.5, and Lemma 4.3.6 together extend [63, Corollary 2.14 and Lemma 2.15] from shallow ANNs to deep ANNs.

Lemma 4.3.6. *Assume Setting 4.1.1, assume for all $x \in [a,b]^d$ that $f(x) = f(0)$, let $(\gamma_n)_{n\in\mathbb{N}_0} \subseteq [0,\infty)$, let $(\Theta_n)_{n\in\mathbb{N}_0} \colon \mathbb{N}_0 \to \mathbb{R}^{\mathfrak{d}}$ satisfy for all $n \in \mathbb{N}_0$ that $\Theta_{n+1} = \Theta_n - \gamma_n \mathcal{G}(\Theta_n)$, and assume*

$$\sup\nolimits_{n\in\mathbb{N}_0}(\gamma_n \mathfrak{m}) \leqslant \big(L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|f(0)\|^2 + 1)^{(L-1)}\big)^{-1}. \tag{4.183}$$

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

$$\begin{aligned} &V(\Theta_{n+1}) - V(\Theta_n) \\ &\leqslant -4L\gamma_n\big(1 - \big[\sup\nolimits_{m\in\mathbb{N}_0}\gamma_m\big]\mathfrak{m} L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big] \\ &\quad \cdot (2V(\Theta_0) + 4L^2\|f(0)\|^2 + 1)^{(L-1)}\big)\mathcal{L}_\infty(\Theta_n) \leqslant 0. \end{aligned} \tag{4.184}$$

Proof of Lemma 4.3.6. Throughout this proof let $\mathfrak{g} \in \mathbb{R}$ satisfy $\mathfrak{g} = \sup_{n\in\mathbb{N}_0}(\gamma_n \mathfrak{m})$. We prove (4.184) by induction on $n \in \mathbb{N}_0$. Note that Corollary 4.3.5, (4.183), and the fact that $\gamma_0 \mathfrak{m} \leqslant \mathfrak{g}$ imply that

$$\begin{aligned} &V(\Theta_1) - V(\Theta_0) \\ &\leqslant 4L\big(-\gamma_0 + \gamma_0^2 \mathfrak{m} L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|f(0)\|^2 + 1)^{(L-1)}\big)\mathcal{L}_\infty(\Theta_0) \\ &\leqslant 4L\big(-\gamma_0 + \gamma_0 \mathfrak{g} L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|f(0)\|^2 + 1)^{(L-1)}\big)\mathcal{L}_\infty(\Theta_0) \\ &= -4L\gamma_0\big(1 - \mathfrak{g} L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|f(0)\|^2 + 1)^{(L-1)}\big)\mathcal{L}_\infty(\Theta_0) \\ &\leqslant 0. \end{aligned} \tag{4.185}$$

This establishes (4.184) in the base case $n = 0$. For the induction step let $n \in \mathbb{N}$ satisfy for all $m \in \{0, 1, \dots, n-1\}$ that

$$\begin{aligned} &V(\Theta_{m+1}) - V(\Theta_m) \\ &\leqslant -4L\gamma_m\big(1 - \mathfrak{g} L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|f(0)\|^2 + 1)^{(L-1)}\big)\mathcal{L}_\infty(\Theta_m). \end{aligned} \tag{4.186}$$

Observe that (4.186) and the fact that for all $\theta \in \mathbb{R}^{\mathfrak{d}}$ it holds that $\mathcal{L}_\infty(\theta) \geqslant 0$ ensure that $V(\Theta_n) \leqslant V(\Theta_{n-1}) \leqslant \dots \leqslant V(\Theta_0)$. Combining this with Corollary 4.3.5, (4.183), and the fact that $\gamma_n \mathfrak{m} \leqslant \mathfrak{g}$ demonstrates that

$$
\begin{aligned}
&V(\Theta_{n+1}) - V(\Theta_n) \\
&\leqslant 4L\big(-\gamma_n + (\gamma_n)^2 \mathfrak{m} L \mathbf{a}^2 \big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big] \\
&\qquad \cdot (2V(\Theta_0) + 4L^2\|f(0)\|^2 + 1)^{(L-1)}\big)\mathcal{L}_\infty(\Theta_n) \\
&\leqslant 4L\big(-\gamma_n + \gamma_n \mathfrak{g} L \mathbf{a}^2 \big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|f(0)\|^2 + 1)^{(L-1)}\big)\mathcal{L}_\infty(\Theta_n) \\
&= -4L\gamma_n\big(1 - \mathfrak{g} L \mathbf{a}^2 \big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|f(0)\|^2 + 1)^{(L-1)}\big)\mathcal{L}_\infty(\Theta_n) \\
&\leqslant 0.
\end{aligned}
\tag{4.187}
$$

Induction thus establishes (4.184). The proof of Lemma 4.3.6 is thus complete. □

4.3.3 Convergence analysis for GD processes

The following theorem proves that the sequence of risks of any time-discrete GD process converges to zero in the training of DNNs provided that the target function is constant and the step sizes are sufficiently small but not L^1-summable. For this reason, Theorem 4.3.7 generalizes [63, Theorem 2.16].

Theorem 4.3.7. *Assume Setting 4.1.1, assume for all $x \in [a,b]^d$ that $f(x) = f(0)$, let $(\gamma_n)_{n\in\mathbb{N}_0} \subseteq [0,\infty)$, let $(\Theta_n)_{n\in\mathbb{N}_0}\colon \mathbb{N}_0 \to \mathbb{R}^{\mathfrak{d}}$ satisfy for all $n \in \mathbb{N}_0$ that $\Theta_{n+1} = \Theta_n - \gamma_n \mathcal{G}(\Theta_n)$, and assume*

$$
\sup\nolimits_{n\in\mathbb{N}_0}(\gamma_n \mathfrak{m}) < \big(L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|f(0)\|^2 + 1)^{(L-1)}\big)^{-1} \tag{4.188}
$$

and $\sum_{n=0}^{\infty} \gamma_n = \infty$. Then

(i) it holds that $\sup_{n\in\mathbb{N}_0}\|\Theta_n\| \leqslant [2V(\Theta_0) + 4L^2\|f(0)\|^2]^{1/2} < \infty$ and

(ii) it holds that $\limsup_{n\to\infty} \mathcal{L}_\infty(\Theta_n) = 0$.

Proof of Theorem 4.3.7. Throughout this proof let $\eta \in (0,\infty)$ satisfy

$$
\eta = 4L\big(1 - \big[\sup\nolimits_{m\in\mathbb{N}_0}\gamma_m\big]\mathfrak{m} L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|f(0)\|^2 + 1)^{(L-1)}\big) \tag{4.189}
$$

and let $\varepsilon \in \mathbb{R}$ satisfy $\varepsilon = (1/3)[\min\{1, \limsup_{n\to\infty}\mathcal{L}_\infty(\Theta_n)\}]$. Note that item (ii) of Proposition 4.2.1 ensures that for all $n \in \mathbb{N}_0$ it holds that

$$
\|\Theta_n\| \leqslant [2V(\Theta_n) + 4L^2\|f(0)\|^2]^{1/2}. \tag{4.190}
$$

Furthermore, observe that Lemma 4.3.6 implies that for all $n \in \mathbb{N}_0$ it holds that $V(\Theta_n) \leqslant V(\Theta_{n-1}) \leqslant \cdots \leqslant V(\Theta_0)$. This and (4.190) establish item (i). Next note that item (ii) of Proposition 4.2.1 ensures that for all $n \in \mathbb{N}$ we have that $V(\Theta_n) \geqslant \frac{1}{2}\|\Theta_n\|^2 - 2L^2\|f(0)\|^2 \geqslant -2L^2\|f(0)\|^2$. Combining this with Lemma 4.3.6 and (4.189) implies that for all $N \in \mathbb{N}$ it holds that

$$
\begin{aligned}
\eta\left[\sum_{n=0}^{N-1}\gamma_n\mathcal{L}_\infty(\Theta_n)\right] &\leqslant \sum_{n=0}^{N-1}(V(\Theta_n) - V(\Theta_{n+1})) \\
&= V(\Theta_0) - V(\Theta_N) \leqslant V(\Theta_0) + 2L^2\|f(0)\|^2.
\end{aligned}
\tag{4.191}
$$

This demonstrates that

$$
\sum_{n=0}^{\infty}[\gamma_n\mathcal{L}_\infty(\Theta_n)] \leqslant \eta^{-1}(V(\Theta_0) + 2L^2\|f(0)\|^2) < \infty. \tag{4.192}
$$

Combining this with the assumption that $\sum_{n=0}^{\infty} \gamma_n = \infty$ ensures that $\liminf_{n\to\infty} \mathcal{L}_\infty(\Theta_n) = 0$. In the following we prove item (ii) by contradiction. For this assume that

$$\limsup_{n\to\infty} \mathcal{L}_\infty(\Theta_n) > 0. \tag{4.193}$$

Observe that (4.193) implies that

$$0 = \liminf_{n\to\infty} \mathcal{L}_\infty(\Theta_n) < \varepsilon < 2\varepsilon < \limsup_{n\to\infty} \mathcal{L}_\infty(\Theta_n). \tag{4.194}$$

This ensures that there exist $(m_k, n_k) \in \mathbb{N}^2$, $k \in \mathbb{N}$, which satisfy for all $k \in \mathbb{N}$ that $m_k < n_k < m_{k+1}$, $\mathcal{L}_\infty(\Theta_{m_k}) > 2\varepsilon$, and $\mathcal{L}_\infty(\Theta_{n_k}) < \varepsilon < \min_{j\in\mathbb{N}\cap[m_k,n_k)} \mathcal{L}_\infty(\Theta_j)$. Note that (4.192) and the fact that for all $k \in \mathbb{N}$, $j \in \mathbb{N} \cap [m_k, n_k)$ it holds that $1 \leqslant \frac{1}{\varepsilon}\mathcal{L}_\infty(\Theta_j)$ demonstrate that

$$\sum_{k=1}^{\infty} \sum_{j=m_k}^{n_k-1} \gamma_j \leqslant \frac{1}{\varepsilon}\left[\sum_{k=1}^{\infty} \sum_{j=m_k}^{n_k-1} (\gamma_j \mathcal{L}_\infty(\Theta_j))\right] \leqslant \frac{1}{\varepsilon}\left[\sum_{j=0}^{\infty} (\gamma_j \mathcal{L}_\infty(\Theta_j))\right] < \infty. \tag{4.195}$$

Moreover, observe that Corollary 4.1.12 and item (i) imply that there exists $\mathfrak{C} \in \mathbb{R}$ which satisfies that

$$\sup_{n\in\mathbb{N}_0} \|\mathcal{G}(\Theta_n)\| \leqslant \mathfrak{C}. \tag{4.196}$$

Note that (4.196), the triangle inequality, and (4.195) prove that

$$\begin{aligned}\sum_{k=1}^{\infty} \|\Theta_{n_k} - \Theta_{m_k}\| &\leqslant \sum_{k=1}^{\infty} \sum_{j=m_k}^{n_k-1} \|\Theta_{j+1} - \Theta_j\| = \sum_{k=1}^{\infty} \sum_{j=m_k}^{n_k-1} (\gamma_j \|\mathcal{G}(\Theta_j)\|) \\ &\leqslant \mathfrak{C}\left[\sum_{k=1}^{\infty} \sum_{j=m_k}^{n_k-1} \gamma_j\right] < \infty.\end{aligned} \tag{4.197}$$

Next observe that Lemma 4.1.10 and item (i) ensure that there exists $\mathscr{L} \in \mathbb{R}$ which satisfies for all $m, n \in \mathbb{N}_0$ that $|\mathcal{L}(\Theta_m) - \mathcal{L}(\Theta_n)| \leqslant \mathscr{L}\|\Theta_m - \Theta_n\|$. Combining this with (4.197) shows that

$$\limsup_{k\to\infty} |\mathcal{L}_\infty(\Theta_{n_k}) - \mathcal{L}_\infty(\Theta_{m_k})| \leqslant \limsup_{k\to\infty} (\mathscr{L}\|\Theta_{n_k} - \Theta_{m_k}\|) = 0. \tag{4.198}$$

The fact that for all $k \in \mathbb{N}_0$ it holds that $\mathcal{L}_\infty(\Theta_{n_k}) < \varepsilon < 2\varepsilon < \mathcal{L}_\infty(\Theta_{m_k})$ therefore implies that

$$0 < \varepsilon \leqslant \inf_{k\in\mathbb{N}} |\mathcal{L}_\infty(\Theta_{n_k}) - \mathcal{L}_\infty(\Theta_{m_k})| \leqslant \limsup_{k\to\infty} |\mathcal{L}_\infty(\Theta_{n_k}) - \mathcal{L}_\infty(\Theta_{m_k})| = 0. \tag{4.199}$$

This contradiction establishes item (ii). The proof of Theorem 4.3.7 is thus complete. □

4.4 Stochastic gradient descent (SGD) processes in the training of deep ANNs

In this section we study SGD processes in the training of deep ReLU ANNS. Subsection 4.4.1 introduces a mathematical description of the considered SGD processes in the training of deep ReLU ANNs. In Subsection 4.4.2 we derive an explicit representation for the generalized gradients of the empirical risk function. Subsection 4.4.3 analyses measurability and independence properties and calculates expectations of the empirical risk functions. Subsection 4.4.4 establishes upper bounds for the norm of the generalized gradients of the empirical risk function and proves its uniform boundedness. In Subsection 4.3.1 we present a one-step Lyapunov estimate which is used in Subsection 4.1.5 to establish upper estimates for the composition of the considered Lyapunov function and the SGD processes. In Subsection 4.4.7 we combine this section's findings to obtain our main result, Theorem 4.4.11. Theorem 4.4.11 proves that in the training of DNNs it holds that the sequence of the risks of any time-discrete SGD process converges to zero provided that the target function is constant and the step sizes are sufficiently small but not L^1-summable.

4.4.1 Mathematical framework for SGD processes and deep ANNs with ReLU activation

During this section we are going to use the following setting.

Setting 4.4.1. *Let* $L, \mathfrak{d} \in \mathbb{N}$, $(\ell_k)_{k\in\mathbb{N}_0} \subseteq \mathbb{N}$, $\xi \in \mathbb{R}^{\ell_L}$, $a, \mathbf{a} \in \mathbb{R}$, $b \in (a,\infty)$, $\mathscr{A} \in (0,\infty)$, $\mathscr{B} \in (\mathscr{A},\infty)$ *satisfy* $\mathfrak{d} = \sum_{k=1}^{L} \ell_k(\ell_{k-1}+1)$ *and* $\mathbf{a} = \max\{|a|,|b|,1\}$, *for every* $\theta = (\theta_1,\dots,\theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ *let* $\mathfrak{w}^{k,\theta} = (\mathfrak{w}^{k,\theta}_{i,j})_{(i,j)\in\{1,\dots,\ell_k\}\times\{1,\dots,\ell_{k-1}\}} \in \mathbb{R}^{\ell_k\times\ell_{k-1}}$, $k \in \mathbb{N}$, *and* $\mathfrak{b}^{k,\theta} = (\mathfrak{b}^{k,\theta}_1,\dots,\mathfrak{b}^{k,\theta}_{\ell_k}) \in \mathbb{R}^{\ell_k}$, $k \in \mathbb{N}$, *satisfy for all* $k \in \{1,\dots,L\}$, $i \in \{1,\dots,\ell_k\}$, $j \in \{1,\dots,\ell_{k-1}\}$ *that*

$$\mathfrak{w}^{k,\theta}_{i,j} = \theta_{(i-1)\ell_{k-1}+j+\sum_{h=1}^{k-1}\ell_h(\ell_{h-1}+1)} \qquad \text{and} \qquad \mathfrak{b}^{k,\theta}_i = \theta_{\ell_k\ell_{k-1}+i+\sum_{h=1}^{k-1}\ell_h(\ell_{h-1}+1)}, \tag{4.200}$$

for every $k \in \mathbb{N}$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ *let* $\mathcal{A}^\theta_k = (\mathcal{A}^\theta_{k,1},\dots,\mathcal{A}^\theta_{k,\ell_k})\colon \mathbb{R}^{\ell_{k-1}} \to \mathbb{R}^{\ell_k}$ *satisfy for all* $x \in \mathbb{R}^{\ell_{k-1}}$ *that* $\mathcal{A}^\theta_k(x) = \mathfrak{b}^{k,\theta} + \mathfrak{w}^{k,\theta}x$, *let* $\mathscr{R}_r\colon \mathbb{R} \to \mathbb{R}$, $r \in [1,\infty]$, *satisfy for all* $r \in [1,\infty)$, $x \in (-\infty, \mathscr{A}r^{-1}]$, $y \in \mathbb{R}$, $z \in [\mathscr{B}r^{-1},\infty)$ *that*

$$\mathscr{R}_r \in C^1(\mathbb{R},\mathbb{R}),\ \mathscr{R}_r(x) = 0,\ 0 \leqslant \mathscr{R}_r(y) \leqslant \mathscr{R}_\infty(y) = \max\{y,0\},\ \text{and}\ \mathscr{R}_r(z) = z, \tag{4.201}$$

assume $\sup_{r\in[1,\infty)}\sup_{x\in\mathbb{R}} |(\mathscr{R}_r)'(x)| < \infty$, *let* $\|\cdot\|\colon (\cup_{n\in\mathbb{N}}\mathbb{R}^n) \to \mathbb{R}$, $\langle\cdot,\cdot\rangle : (\cup_{n\in\mathbb{N}}(\mathbb{R}^n \times \mathbb{R}^n)) \to \mathbb{R}$, *and* $\mathfrak{M}_r\colon (\cup_{n\in\mathbb{N}}\mathbb{R}^n) \to (\cup_{n\in\mathbb{N}}\mathbb{R}^n)$, $r \in [1,\infty]$, *satisfy for all* $r \in [1,\infty]$, $n \in \mathbb{N}$, $x = (x_1,\dots,x_n)$, $y = (y_1,\dots,y_n) \in \mathbb{R}^n$ *that*

$$\|x\| = (\textstyle\sum_{i=1}^n |x_i|^2)^{1/2},\ \langle x,y\rangle = \sum_{i=1}^n x_iy_i,\ \text{and}\ \mathfrak{M}_r(x) = (\mathscr{R}_r(x_1),\dots,\mathscr{R}_r(x_n)), \tag{4.202}$$

for every $\theta \in \mathbb{R}^{\mathfrak{d}}$ *let* $\mathcal{N}^{k,\theta}_r = (\mathcal{N}^{k,\theta}_{r,1},\dots,\mathcal{N}^{k,\theta}_{r,\ell_k})\colon \mathbb{R}^{\ell_0} \to \mathbb{R}^{\ell_k}$, $r \in [1,\infty]$, $k \in \mathbb{N}$, *and* $\mathcal{X}^{k,\theta}_i \subseteq \mathbb{R}^{\ell_0}$, $k, i \in \mathbb{N}$, *satisfy for all* $r \in [1,\infty]$, $k \in \mathbb{N}$, $i \in \{1,\dots,\ell_k\}$, $x \in \mathbb{R}^{\ell_0}$ *that*

$$\mathcal{N}^{1,\theta}_r(x) = \mathcal{A}^\theta_1(x), \qquad \mathcal{N}^{k+1,\theta}_r(x) = \mathcal{A}^\theta_{k+1}(\mathfrak{M}_{r^{1/k}}(\mathcal{N}^{k,\theta}_r(x))), \tag{4.203}$$

and $\mathcal{X}^{k,\theta}_i = \{y \in [a,b]^{\ell_0}\colon \mathcal{N}^{k,\theta}_{\infty,i}(y) > 0\}$, *let* $(\Omega,\mathcal{F},\mathbb{P})$ *be a probability space, let* $X^{n,m} = (X^{n,m}_1,\dots,X^{n,m}_{\ell_0})\colon \Omega \to [a,b]^{\ell_0}$, $n, m \in \mathbb{N}_0$ *be i.i.d. random variables, let* $\mathcal{L}\colon \mathbb{R}^{\mathfrak{d}} \to \mathbb{R}$ *and* $V\colon \mathbb{R}^{\mathfrak{d}} \to \mathbb{R}$ *satisfy for all* $\theta \in \mathbb{R}^{\mathfrak{d}}$ *that* $\mathcal{L}(\theta) = \mathbb{E}[\|\mathcal{N}^{L,\theta}_\infty(X^{0,0}) - \xi\|^2]$ *and*

$$V(\theta) = \big[\textstyle\sum_{k=1}^L \big(k\|\mathfrak{b}^{k,\theta}\|^2 + \sum_{i=1}^{\ell_k}\sum_{j=1}^{\ell_{k-1}} |\mathfrak{w}^{k,\theta}_{i,j}|^2\big)\big] - 2L\langle\xi,\mathfrak{b}^{L,\theta}\rangle, \tag{4.204}$$

let $(M_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{N}$, *for every* $n \in \mathbb{N}_0$, $r \in [1,\infty]$ *let* $\mathfrak{L}^n_r\colon \mathbb{R}^{\mathfrak{d}} \times \Omega \to \mathbb{R}$ *satisfy for all* $\theta \in \mathbb{R}^{\mathfrak{d}}$, $\omega \in \Omega$ *that* $\mathfrak{L}^n_r(\theta,\omega) = \frac{1}{M_n}\sum_{m=1}^{M_n}\|\mathcal{N}^{L,\theta}_r(X^{n,m}(\omega)) - \xi\|^2$, *for every* $n \in \mathbb{N}_0$ *let* $\mathfrak{G}^n = (\mathfrak{G}^n_1,\dots,\mathfrak{G}^n_{\mathfrak{d}})\colon \mathbb{R}^{\mathfrak{d}} \times \Omega \to \mathbb{R}^{\mathfrak{d}}$ *satisfy for all* $\theta \in \mathbb{R}^{\mathfrak{d}}$, $\omega \in \{w \in \Omega\colon ((\nabla_\theta\mathfrak{L}^n_r)(\theta,w))_{r\in[1,\infty)}$ *is convergent*$\}$ *that*

$$\mathfrak{G}^n(\theta,\omega) = \lim_{r\to\infty}(\nabla_\theta\mathfrak{L}^n_r)(\theta,\omega), \tag{4.205}$$

let $\Theta = (\Theta_n)_{n\in\mathbb{N}_0}\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^{\mathfrak{d}}$ *be a stochastic process, let* $(\gamma_n)_{n\in\mathbb{N}_0} \subseteq [0,\infty)$, *assume that* Θ_0 *and* $(X^{n,m})_{(n,m)\in(\mathbb{N}_0)^2}$ *are independent, and assume for all* $n \in \mathbb{N}_0$, $\omega \in \Omega$ *that* $\Theta_{n+1}(\omega) = \Theta_n(\omega) - \gamma_n\mathfrak{G}^n(\Theta_n(\omega),\omega)$.

4.4.2 Explicit representations for the generalized gradients of the empirical risk function

The following proposition provides a representation for the generalized gradients of the approximations of the empirical risk function. Proposition 4.4.2 expands [63, Proposition 3.2] from shallow ANNs to deep ANNs.

Proposition 4.4.2. *Assume Setting 4.4.1 and let* $n \in \mathbb{N}_0$, $\omega \in \Omega$. *Then*

(i) *it holds for all* $r \in [1, \infty)$ *that* $(\mathbb{R}^{\mathfrak{d}} \ni \theta \mapsto \mathfrak{L}^n_r(\theta, \omega) \in \mathbb{R}) \in C^1(\mathbb{R}^{\mathfrak{d}}, \mathbb{R})$,

(ii) *it holds for all* $r \in [1, \infty)$, $k \in \{1, \dots, L\}$, $i \in \{1, \dots, \ell_k\}$, $j \in \{1, \dots, \ell_{k-1}\}$, $\theta = (\theta_1, \dots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ *that*

$$\begin{aligned}&\left(\frac{\partial}{\partial \theta_{(i-1)\ell_{k-1}+j+\sum_{h=1}^{k-1}\ell_h(\ell_{h-1}+1)}} \mathfrak{L}^n_r\right)(\theta, \omega) \\ &= \sum_{\substack{v_k, v_{k+1}, \dots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k, L] \colon v_w \le \ell_w}} \tfrac{2}{M_n} \sum_{m=1}^{M_n} \Big[\mathscr{R}_{r^{1/(\max\{k-1,1\})}}(\mathcal{N}^{\max\{k-1,1\},\theta}_{r,j}(X^{n,m}(\omega)))\mathbb{1}_{(1,L]}(k) \\ &\quad + X^{n,m}_j(\omega)\mathbb{1}_{\{1\}}(k)\Big]\Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathcal{N}^{L,\theta}_{r,v_L}(X^{n,m}(\omega)) - \xi_{v_L}\Big] \\ &\quad \cdot \Big[\textstyle\prod_{q=k+1}^{L}\big(\mathfrak{w}^{q,\theta}_{v_q, v_{q-1}}\big[(\mathscr{R}_{r^{1/(q-1)}})'(\mathcal{N}^{q-1,\theta}_{r,v_{q-1}}(X^{n,m}(\omega)))\big]\big)\Big],\end{aligned} \tag{4.206}$$

(iii) *it holds for all* $r \in [1, \infty)$, $k \in \{1, \dots, L\}$, $i \in \{1, \dots, \ell_k\}$, $\theta = (\theta_1, \dots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ *that*

$$\begin{aligned}&\left(\frac{\partial}{\partial \theta_{\ell_k \ell_{k-1}+i+\sum_{h=1}^{k-1}\ell_h(\ell_{h-1}+1)}} \mathfrak{L}^n_r\right)(\theta, \omega) \\ &= \sum_{\substack{v_k, v_{k+1}, \dots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k, L] \colon v_w \le \ell_w}} \tfrac{2}{M_n} \sum_{m=1}^{M_n} \Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathcal{N}^{L,\theta}_{r,v_L}(X^{n,m}(\omega)) - \xi_{v_L}\Big] \\ &\quad \cdot \Big[\textstyle\prod_{q=k+1}^{L}\big(\mathfrak{w}^{q,\theta}_{v_q, v_{q-1}}\big[(\mathscr{R}_{r^{1/(q-1)}})'(\mathcal{N}^{q-1,\theta}_{r,v_{q-1}}(X^{n,m}(\omega)))\big]\big)\Big],\end{aligned} \tag{4.207}$$

(iv) *it holds for all* $\theta = (\theta_1, \dots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ *that* $\limsup_{r \to \infty} \|(\nabla \mathfrak{L}^n_r)(\theta, \omega) - \mathfrak{G}^n(\theta, \omega)\| = 0$,

(v) *it holds for all* $k \in \{1, \dots, L\}$, $i \in \{1, \dots, \ell_k\}$, $j \in \{1, \dots, \ell_{k-1}\}$, $\theta = (\theta_1, \dots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ *that*

$$\begin{aligned}&\mathfrak{G}^n_{(i-1)\ell_{k-1}+j+\sum_{h=1}^{k-1}\ell_h(\ell_{h-1}+1)}(\theta, \omega) \\ &= \sum_{\substack{v_k, v_{k+1}, \dots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k, L] \colon v_w \le \ell_w}} \tfrac{2}{M_n} \sum_{m=1}^{M_n} \Big[\mathscr{R}_{\infty}(\mathcal{N}^{\max\{k-1,1\},\theta}_{\infty,j}(X^{n,m}(\omega)))\mathbb{1}_{(1,L]}(k) \\ &\quad + X^{n,m}_j(\omega)\mathbb{1}_{\{1\}}(k)\Big]\Big[\mathbb{1}_{\{i\}}(v_k)\Big]\Big[\mathcal{N}^{L,\theta}_{\infty,v_L}(X^{n,m}(\omega)) - \xi_{v_L}\Big] \\ &\quad \cdot \Big[\textstyle\prod_{q=k+1}^{L}\big(\mathfrak{w}^{q,\theta}_{v_q, v_{q-1}} \mathbb{1}_{\mathcal{X}^{q-1,\theta}_{v_{q-1}}}(X^{n,m}(\omega))\big)\Big],\end{aligned} \tag{4.208}$$

and

(vi) *it holds for all* $k \in \{1, \dots, L\}$, $i \in \{1, \dots, \ell_k\}$, $\theta = (\theta_1, \dots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ *that*

$$\begin{aligned}\mathfrak{G}^n_{\ell_k \ell_{k-1}+i+\sum_{h=1}^{k-1}\ell_h(\ell_{h-1}+1)}(\theta, \omega) &= \sum_{\substack{v_k, v_{k+1}, \dots, v_L \in \mathbb{N}, \\ \forall w \in \mathbb{N} \cap [k, L] \colon v_w \le \ell_w}} \tfrac{2}{M_n} \sum_{m=1}^{M_n} \Big[\mathbb{1}_{\{i\}}(v_k)\Big] \\ &\quad \cdot \Big[\mathcal{N}^{L,\theta}_{\infty,v_L}(X^{n,m}(\omega)) - \xi_{v_L}\Big]\Big[\textstyle\prod_{q=k+1}^{L}\big(\mathfrak{w}^{q,\theta}_{v_q, v_{q-1}} \mathbb{1}_{\mathcal{X}^{q-1,\theta}_{v_{q-1}}}(X^{n,m}(\omega))\big)\Big].\end{aligned} \tag{4.209}$$

Proof of Proposition 4.4.2. Note that items (i), (ii), (iii), (iv), (v) and (vi) of Theorem 4.1.9 (applied with $\mu \curvearrowleft (\mathcal{B}([a,b]^{\ell_0}) \ni A \mapsto \frac{1}{M_n}\sum_{m=1}^{M_n} \mathbb{1}_A(X^{n,m}(\omega)) \in [0,1])$, $f \curvearrowleft ([a,b]^{\ell_0} \ni x \mapsto \xi \in \mathbb{R}^{\ell_L})$ in the notation of Theorem 4.1.9) prove items (i), (ii), (iii), (iv), (v) and (vi). The proof of Proposition 4.4.2 is thus complete. □

4.4.3 Properties of expectations of the empirical risk functions

In the following proposition we establish that the expected value of the empirical risk function and the risk function agree. Proposition 4.4.3 is a generalization of [63, Proposition 3.3].

Proposition 4.4.3. *Assume Setting 4.4.1. Then it holds for all $n \in \mathbb{N}_0$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ that*

$$\mathbb{E}[\mathfrak{L}_\infty^n(\theta)] = \mathcal{L}(\theta). \tag{4.210}$$

Proof of Proposition 4.4.3. Observe that the fact that for all $n \in \mathbb{N}_0$, $\theta \in \mathbb{R}^{\mathfrak{d}}$, $\omega \in \Omega$ it holds that $\mathfrak{L}_\infty^n(\theta,\omega) = \frac{1}{M_n}\sum_{m=1}^{M_n} \|\mathcal{N}_\infty^{L,\theta}(X^{n,m}(\omega)) - \xi\|^2$, the fact that for all $n \in \mathbb{N}_0$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ it holds that $\mathcal{L}(\theta) = \mathbb{E}[\|\mathcal{N}_\infty^{L,\theta}(X^{0,0}) - \xi\|^2]$, and the assumption that $X^{n,m} \colon \Omega \to [a,b]^{\ell_0}$, $n, m \in \mathbb{N}_0$, are i.i.d. random variables imply that for all $n \in \mathbb{N}_0$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ we have that

$$\mathbb{E}[\mathfrak{L}_\infty^n(\theta)] = \frac{1}{M_n}\sum_{m=1}^{M_n} \mathbb{E}\big[\|\mathcal{N}_\infty^{L,\theta}(X^{n,m}) - \xi\|^2\big] = \mathbb{E}\big[\|\mathcal{N}_\infty^{L,\theta}(X^{0,0}) - \xi\|^2\big] = \mathcal{L}(\theta). \tag{4.211}$$

The proof of Proposition 4.4.3 is thus complete. □

The following lemma proves measurability and independence properties for certain stochastic processes. Lemma 4.4.4 extends the findings in [63, Lemma 3.4].

Lemma 4.4.4. *Assume Setting 4.4.1 and let $\mathbb{F}_n \subseteq \mathcal{F}$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}$ that $\mathbb{F}_0 = \sigma(\Theta_0)$ and $\mathbb{F}_n = \sigma(\Theta_0, (X^{\mathrm{n,m}})_{\mathrm{n,m} \in (\mathbb{N}\cap[0,n))\times\mathbb{N}_0})$. Then*

(i) it holds for all $n \in \mathbb{N}_0$ that $\mathbb{R}^{\mathfrak{d}} \times \Omega \ni (\theta,\omega) \mapsto \mathfrak{G}^n(\theta,\omega) \in \mathbb{R}^{\mathfrak{d}}$ is $(\mathcal{B}(\mathbb{R}^{\mathfrak{d}}) \otimes \mathbb{F}_{n+1})/\mathcal{B}(\mathbb{R}^{\mathfrak{d}})$-measurable,

(ii) it holds for all $n \in \mathbb{N}_0$ that Θ_n is $\mathbb{F}_n/\mathcal{B}(\mathbb{R}^{\mathfrak{d}})$-measurable, and

(iii) it holds for all $m, n \in \mathbb{N}_0$ that $\sigma(X^{n,m})$ and $\mathbb{F}_n$ are independent.

Proof of Lemma 4.4.4. Note that Lemma 4.1.10, items (ii) and (iii) of Proposition 4.4.2, and the assumption that for all $r \in [1,\infty)$ it holds that $\mathscr{R}_r \in C^1(\mathbb{R},\mathbb{R})$ ensure that for all $n \in \mathbb{N}_0$, $r \in [1,\infty)$, $\omega \in \Omega$ we have that $\mathbb{R}^{\mathfrak{d}} \ni \theta \mapsto (\nabla_\theta \mathfrak{L}_r^n)(\theta,\omega) \in \mathbb{R}^{\mathfrak{d}}$ is continuous. Furthermore, observe that items (ii) and (iii) of Proposition 4.4.2 and the assumption that for all $m, n \in \mathbb{N}_0$ it holds that $X^{n,m}$ is $\mathbb{F}_{n+1}/\mathcal{B}([a,b]^{\ell_0})$-measurable imply that for all $n \in \mathbb{N}_0$, $r \in [1,\infty)$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ it holds that $\Omega \ni \omega \mapsto (\nabla_\theta \mathfrak{L}_r^n(\theta,\omega)) \in \mathbb{R}^{\mathfrak{d}}$ is $\mathbb{F}_{n+1}/\mathcal{B}(\mathbb{R}^{\mathfrak{d}})$-measurable. Combining this and, e.g., Beck et al. [14, Lemma 2.4] (applied with $(E,\delta) \curvearrowleft (\mathbb{R}^{\mathfrak{d}}, \mathcal{B}(\mathbb{R}^{\mathfrak{d}}))$, $(\Omega,\mathcal{F}) \curvearrowleft (\Omega, \mathbb{F}_{n+1})$, $X \curvearrowleft (\mathbb{R}^{\mathfrak{d}} \times \Omega \ni (\theta,\omega) \mapsto (\nabla_\theta \mathfrak{L}_r^n)(\theta,\omega) \in \mathbb{R}^{\mathfrak{d}})$ in the notation of [14, Lemma 2.4]) demonstrates that for all $n \in \mathbb{N}_0$, $r \in [1,\infty)$ we have that $\mathbb{R}^{\mathfrak{d}} \times \Omega \ni (\theta,\omega) \mapsto (\nabla_\theta \mathfrak{L}_r^n)(\theta,\omega) \in \mathbb{R}^{\mathfrak{d}}$ is $(\mathcal{B}(\mathbb{R}^{\mathfrak{d}}) \otimes \mathbb{F}_{n+1})/\mathcal{B}(\mathbb{R}^{\mathfrak{d}})$-measurable. Item (iv) of Proposition 4.4.2 hence proves that for all $n \in \mathbb{N}_0$ it holds that $\mathbb{R}^{\mathfrak{d}} \times \Omega \ni (\theta,\omega) \mapsto \mathfrak{G}^n(\theta,\omega) \in \mathbb{R}^{\mathfrak{d}}$ is $(\mathcal{B}(\mathbb{R}^{\mathfrak{d}}) \otimes \mathbb{F}_{n+1})/\mathcal{B}(\mathbb{R}^{\mathfrak{d}})$-measurable. This establishes item (i).

Next we prove item (ii) by induction on $n \in \mathbb{N}_0$. The assumption that $\mathbb{F}_0 = \sigma(\Theta_0)$ implies that Θ_0 is $\mathbb{F}_0/\mathcal{B}(\mathbb{R}^{\mathfrak{d}})$-measurable. This establishes item (ii) in the base case $n = 0$. For the induction step let $n \in \mathbb{N}_0$ satisfy that Θ_n is $\mathbb{F}_n/\mathcal{B}(\mathbb{R}^{\mathfrak{d}})$-measurable. Note that the fact that $\mathbb{F}_n = \sigma(\Theta_0, (X^{\mathfrak{n},\mathfrak{m}})_{\mathfrak{n},\mathfrak{m}\in(\mathbb{N}\cap[0,n))\times\mathbb{N}_0})$ ensures that $\mathbb{F}_n \subseteq \mathbb{F}_{n+1}$. Therefore, we obtain that Θ_n is $\mathbb{F}_{n+1}/\mathcal{B}(\mathbb{R}^{\mathfrak{d}})$-measurable. Moreover, observe that the fact that $\mathbb{F}_n \subseteq \mathbb{F}_{n+1}$ and item (i) imply that $\mathfrak{G}_n$ is $\mathbb{F}_{n+1}/\mathcal{B}(\mathbb{R}^{\mathfrak{d}})$-measurable. Combining this, the fact that Θ_n is $\mathbb{F}_{n+1}/\mathcal{B}(\mathbb{R}^{\mathfrak{d}})$-measurable, and the assumption that $\Theta_{n+1} = \Theta_n - \gamma_n \mathfrak{G}^n(\Theta_n)$ proves that Θ_{n+1} is $\mathbb{F}_{n+1}/\mathcal{B}(\mathbb{R}^{\mathfrak{d}})$-measurable. Induction thus establishes item (ii).

In addition, note that the assumption that $X^{n,m}$, $n, m \in \mathbb{N}_0$, are independent, the assumption that Θ_0 and $(X^{n,m})_{(n,m)\in(\mathbb{N}_0)^2}$ are independent, and the fact that $\mathbb{F}_n = \sigma(\Theta_0, (X^{\mathfrak{n},\mathfrak{m}})_{\mathfrak{n},\mathfrak{m}\in(\mathbb{N}\cap[0,n))\times\mathbb{N}_0})$ establish item (iii). The proof of Lemma 4.4.4 is thus complete. □

Combining Proposition 4.4.3 and Lemma 4.4.4 leads to the following corollary. Corollary 4.4.5 is an extension of [63, Corollary 3.5].

Corollary 4.4.5. Assume Setting 4.4.1. Then it holds for all $n \in \mathbb{N}_0$ that

$$\mathbb{E}[\mathfrak{L}^n_\infty(\Theta_n)] = \mathbb{E}[\mathcal{L}(\Theta_n)]. \tag{4.212}$$

Proof of Corollary 4.4.5. Throughout this proof let $\mathbb{F}_n \subseteq \mathcal{F}$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}_0$ that $\mathbb{F}_0 = \sigma(\Theta_0)$ and $\mathbb{F}_n = \sigma(\Theta_0, (X^{\mathfrak{n},\mathfrak{m}})_{\mathfrak{n},\mathfrak{m}\in(\mathbb{N}\cap[0,n))\times\mathbb{N}_0})$ and let $\boldsymbol{L}^n\colon ([a,b]^{\ell_0})^{M_n} \times \mathbb{R}^{\mathfrak{d}} \to \mathbb{R}$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}_0$, $\theta \in \mathbb{R}^{\mathfrak{d}}$, $x_1, \ldots, x_{M_n} \in [a,b]^{\ell_0}$ that

$$\boldsymbol{L}^n(x_1, \ldots, x_{M_n}, \theta) = \frac{1}{M_n} \sum_{m=1}^{M_n} \|\mathcal{N}^{L,\theta}_\infty(x_m) - \xi\|^2. \tag{4.213}$$

Observe that (4.213) and the assumption that for all $n \in \mathbb{N}_0$, $\theta \in \mathbb{R}^{\mathfrak{d}}$, $\omega \in \Omega$ it holds that $\mathfrak{L}^n_\infty(\theta,\omega) = \frac{1}{M_n}\sum_{m=1}^{M_n}\|\mathcal{N}^{L,\theta}_\infty(X^{n,m}(\omega)) - \xi\|^2$ imply that for all $n \in \mathbb{N}_0$, $\theta \in \mathbb{R}^{\mathfrak{d}}$, $\omega \in \Omega$ we have that

$$\mathfrak{L}^n_\infty(\theta,\omega) = \boldsymbol{L}^n(X^{n,1}(\omega), \ldots, X^{n,M_n}(\omega), \theta). \tag{4.214}$$

Combining this with Proposition 4.4.3 demonstrates that for all $n \in \mathbb{N}_0$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ it holds that

$$\mathbb{E}[\boldsymbol{L}^n(X^{n,1}, \ldots, X^{n,M_n}, \theta)] = \mathbb{E}[\mathfrak{L}^n_\infty(\theta)] = \mathcal{L}(\theta). \tag{4.215}$$

Moreover, note that (4.214) assures that for all $n \in \mathbb{N}_0$ it holds that

$$\mathfrak{L}^n_\infty(\Theta_n) = \boldsymbol{L}^n(X^{n,1}, \ldots, X^{n,M_n}, \Theta_n). \tag{4.216}$$

Combining this, (4.215), Lemma 4.4.4, and, e.g., [66, Lemma 2.8] (applied with $(\Omega, \mathcal{F}, \mathbb{P}) \hookleftarrow (\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G} \hookleftarrow \mathbb{F}_n$, $(\mathbb{X}, \mathcal{X}) \hookleftarrow (([a,b]^{\ell_0})^{M_n}, \mathcal{B}(([a,b]^{\ell_0})^{M_n}))$, $(\mathbb{Y}, \mathcal{Y}) \hookleftarrow (\mathbb{R}^{\mathfrak{d}}, \mathcal{B}(\mathbb{R}^{\mathfrak{d}}))$, $X \hookleftarrow (\Omega \ni \omega \mapsto (X^{n,1}(\omega), \ldots, X^{n,M_n}(\omega)) \in ([a,b]^{\ell_0})^{M_n})$, $Y \hookleftarrow (\Omega \ni \omega \mapsto \Theta_n(\omega) \in \mathbb{R}^{\mathfrak{d}})$ in the notation of [66, Lemma 2.8]) proves that for all $n \in \mathbb{N}_0$ it holds that

$$\mathbb{E}[\boldsymbol{L}^n(X^{n,1}, \ldots, X^{n,M_n}, \Theta_n)] = \mathbb{E}[\mathcal{L}(\Theta_n)]. \tag{4.217}$$

This and (4.214) establish (4.212). The proof of Corollary 4.4.5 is thus complete. □

4.4.4 Upper estimates for the norm of the generalized gradients of the empirical risk function

The following lemma establishes an upper bound for the norm of the generalized gradients of the empirical risk function. Lemma 4.4.6 generalizes [63, Lemma 3.6].

Lemma 4.4.6. *Assume Setting 4.4.1 and let $n \in \mathbb{N}_0$, $\theta \in \mathbb{R}^{\mathfrak{d}}$, $\omega \in \Omega$. Then*

$$\|\mathfrak{G}^n(\theta,\omega)\|^2 \leq 4L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p + 1)\big](\|\theta\|^2 + 1)^{(L-1)}\mathfrak{L}^n_\infty(\theta,\omega). \tag{4.218}$$

Proof of Lemma 4.4.6. Observe that item (iv) of Theorem 4.1.11 (applied with $\mu \curvearrowleft (\mathcal{B}([a,b]^{\ell_0}) \ni A \mapsto \frac{1}{M_n}\sum_{m=1}^{M_n} \mathbb{1}_A(X^{n,m}(\omega)) \in [0,1])$, $f \curvearrowleft ([a,b]^{\ell_0} \ni x \mapsto \xi \in \mathbb{R}^{\ell_L})$ in the notation of Theorem 4.1.11) establishes (4.218). The proof of Lemma 4.4.6 is thus complete. ☐

This leads to the following uniform boundedness result of the norm of the generalized gradients of the empirical risk function which generalizes [63, Lemma 3.7].

Lemma 4.4.7. *Assume Setting 4.4.1 and let $K \subseteq \mathbb{R}^{\mathfrak{d}}$ be compact. Then*

$$\sup\nolimits_{n\in\mathbb{N}_0} \sup\nolimits_{\theta\in K} \sup\nolimits_{\omega\in\Omega} \|\mathfrak{G}^n(\theta,\omega)\| < \infty. \tag{4.219}$$

Proof of Lemma 4.4.7. Note that Lemma 4.1.10 (applied with $\mu \curvearrowleft (\mathcal{B}([a,b]^{\ell_0}) \ni A \mapsto \frac{1}{M_n}\sum_{m=1}^{M_n} \mathbb{1}_A(X^{n,m}(\omega) \in [0,1])$, $f \curvearrowleft ([a,b]^{\ell_0} \ni x \mapsto \xi \in \mathbb{R}^{\ell_L})$ in the notation of Lemma 4.1.10) ensures that there exists $\mathfrak{C} \in \mathbb{R}$ which satisfies for all $\theta \in K$ that $\sup_{x\in[a,b]^{\ell_0}} \|\mathcal{N}_\infty^{L,\theta}(x)\|^2 \leqslant \mathfrak{C}$. The fact that for all $n, m \in \mathbb{N}_0$, $\omega \in \Omega$ it holds that $X^{n,m}(\omega) \in [a,b]^{\ell_0}$ hence demonstrates that for all $n \in \mathbb{N}_0$, $\theta \in K$, $\omega \in \Omega$ we have that

$$\begin{aligned} \mathfrak{L}_\infty^n(\theta,\omega) &= \frac{1}{M_n}\sum_{m=1}^{M_n} \|\mathcal{N}_\infty^{L,\theta}(X^{n,m}(\omega)) - \xi\|^2 \\ &\leqslant \frac{2}{M_n}\sum_{m=1}^{M_n} \big[\|\mathcal{N}_\infty^{L,\theta}(X^{n,m}(\omega))\|^2 + \|\xi\|^2\big] \leqslant 2\mathfrak{C} + 2\|\xi\|^2. \end{aligned} \tag{4.220}$$

Combining this with Lemma 4.4.6 establishes (4.219). The proof of Lemma 4.4.7 is thus complete. ☐

4.4.5 Lyapunov type estimates for the dynamics of SGD processes

The following lemma presents a one-step Lyapunov estimate for the considered SGD method. Lemma 4.4.8 extends [63, Lemma 3.9] to the setting of deep ANNs .

Lemma 4.4.8. *Assume Setting 4.4.1 and let $n \in \mathbb{N}_0$, $\theta \in \mathbb{R}^{\mathfrak{d}}$, $\omega \in \Omega$. Then*

$$\begin{aligned} &V(\theta - \gamma_n \mathfrak{G}^n(\theta,\omega)) - V(\theta) \\ &= (\gamma_n)^2\|\mathfrak{G}^n(\theta,\omega)\|^2 + (\gamma_n)^2\Bigg[\sum_{k=1}^{L}\sum_{i=1}^{\ell_k}(k-1)|\mathfrak{G}^n_{\ell_k\ell_{k-1}+i+\sum_{h=1}^{k-1}\ell_h(\ell_{h-1}+1)}(\theta,\omega)|^2\Bigg] \\ &\quad - 4\gamma_n L \mathfrak{L}_\infty^n(\theta,\omega) \\ &\leqslant (\gamma_n)^2 L\|\mathfrak{G}^n(\theta,\omega)\|^2 - 4\gamma_n L\mathfrak{L}_\infty^n(\theta,\omega). \end{aligned} \tag{4.221}$$

Proof of Lemma 4.4.8. Observe that Corollary 4.3.2 (applied with $\mu \curvearrowleft (\mathcal{B}([a,b]^{\ell_0}) \ni A \mapsto \frac{1}{M_n}\sum_{m=1}^{M_n} \mathbb{1}_A(X^{n,m}(\omega)) \in [0,1])$, $f(0) \curvearrowleft \xi$, $\gamma \curvearrowleft \gamma_n$ in the notation of Corollary 4.3.2) establishes (4.221). The proof of Lemma 4.4.8 is thus complete. ☐

4.4.6 Upper estimates for compositions of Lyapunov functions and SGD processes

The following lemma provides an upper estimate for the composition of the Lyapunov function and the time-discrete SGD process. Lemma 4.4.9 is a generalization of [63, Lemma 3.10].

Lemma 4.4.9. *Assume Setting 4.4.1. Then it holds for all $n \in \mathbb{N}_0$ that*

$$\begin{aligned} &V(\Theta_{n+1}) - V(\Theta_n) \\ &\leq 4L\big((\gamma_n)^2 L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_n) + 4L^2\|\xi\|^2 + 1)^{(L-1)} - \gamma_n\big)\mathfrak{L}_\infty^n(\Theta_n). \end{aligned} \tag{4.222}$$

Proof of Lemma 4.4.9. Note that Lemma 4.4.6 and item (ii) of Proposition 4.2.1 (applied with $\mu \curvearrowleft (\mathcal{B}([a,b]^{\ell_0}) \ni A \mapsto \frac{1}{M_n}\sum_{m=1}^{M_n} \mathbb{1}_A(X^{n,m}(\omega)) \in [0,1])$, $f \curvearrowleft ([a,b]^{\ell_0} \ni x \mapsto \xi \in \mathbb{R}^{\ell_L})$ in the notation of Proposition 4.2.1) demonstrate that for all $n \in \mathbb{N}_0$ it holds that

$$\begin{aligned} \|\mathfrak{G}^n(\Theta_n)\|^2 &\leq 4L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](\|\Theta_n\|^2+1)^{(L-1)}\mathfrak{L}_\infty^n(\Theta_n) \\ &\leq 4L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_n) + 4L^2\|\xi\|^2+1)^{(L-1)}\mathfrak{L}_\infty^n(\Theta_n). \end{aligned} \tag{4.223}$$

Lemma 4.4.8 therefore implies that

$$\begin{aligned} &V(\Theta_{n+1}) - V(\Theta_n) \leq (\gamma_n)^2 L\|\mathfrak{G}^n(\Theta_n)\|^2 - 4\gamma_n L\mathfrak{L}_\infty^n(\Theta_n) \\ &\leq 4(\gamma_n)^2 L^2\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big] \\ &\quad\cdot (2V(\Theta_n) + 4L^2\|\xi\|^2+1)^{(L-1)}\mathfrak{L}_\infty^n(\Theta_n) - 4\gamma_n L\mathfrak{L}_\infty^n(\Theta_n) \\ &= 4L\big((\gamma_n)^2 L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big] \\ &\quad\cdot (2V(\Theta_n) + 4L^2\|\xi\|^2+1)^{(L-1)} - \gamma_n\big)\mathfrak{L}_\infty^n(\Theta_n). \end{aligned} \tag{4.224}$$

The proof of Lemma 4.4.9 is thus complete. □

Applying Lemma 4.4.9 inductively establishes the following lemma which generalizes [63, Corollary 3.11].

Corollary 4.4.10. Assume Setting 4.4.1 and assume

$$\mathbb{P}\big(\sup_{n\in\mathbb{N}_0}\gamma_n \leq \big[L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0)+4L^2\|\xi\|^2+1)^{(L-1)}\big]^{-1}\big) = 1. \tag{4.225}$$

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

$$\begin{aligned} \mathbb{P}\Bigg(&V(\Theta_{n+1}) - V(\Theta_n) \leq -4\gamma_n L\Big[1 - [\sup_{m\in\mathbb{N}_0}\gamma_m] \\ &\cdot L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0)+4L^2\|\xi\|^2+1)^{(L-1)}\Big]\mathfrak{L}_\infty^n(\Theta_n) \leq 0\Bigg) = 1. \end{aligned} \tag{4.226}$$

Proof of Corollary 4.4.10. Throughout this proof let $\mathfrak{g} \in \mathbb{R}$ satisfy that $\mathfrak{g} = \sup_{n\in\mathbb{N}_0}\gamma_n$. We prove (4.226) by induction on $n \in \mathbb{N}_0$. Observe that Lemma 4.4.9 and the assumption that $\gamma_0 \leq \mathfrak{g}$ ensure that it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} &V(\Theta_1) - V(\Theta_0) \\ &\leq 4L\Big[-\gamma_0 + \gamma_0^2 L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0)+4L^2\|\xi\|^2+1)^{(L-1)}\Big]\mathfrak{L}_\infty^0(\Theta_0) \\ &\leq 4L\Big(-\gamma_0 + \gamma_0\mathfrak{g}L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0)+4L^2\|\xi\|^2+1)^{(L-1)}\Big)\mathfrak{L}_\infty^0(\Theta_0) \\ &= -4\gamma_0 L\Big(1 - \mathfrak{g}L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0)+4L^2\|\xi\|^2+1)^{(L-1)}\Big)\mathfrak{L}_\infty^0(\Theta_0). \end{aligned} \tag{4.227}$$

This and (4.225) establish (4.226) in the base case $n = 0$. For the induction step let $n \in \mathbb{N}$ satisfy that for all $m \in \{0, 1, \dots, n-1\}$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} &V(\Theta_{m+1}) - V(\Theta_m) \\ &\leq -4\gamma_m L\Big(1 - \mathfrak{g}L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0)+4L^2\|\xi\|^2+1)^{(L-1)}\Big)\mathfrak{L}_\infty^m(\Theta_m) \leq 0. \end{aligned} \tag{4.228}$$

Note that (4.228) implies that it holds $\mathbb{P}$-a.s. that $V(\Theta_n) \leqslant V(\Theta_{n-1}) \leqslant \ldots \leqslant V(\Theta_0)$. Combining Lemma 4.4.9 with (4.225) and the assumption that $\gamma_n \leqslant \mathfrak{g}$ hence shows that it holds $\mathbb{P}$-a.s. that

$$\begin{aligned}
&V(\Theta_{n+1}) - V(\Theta_n) \\
&\leqslant 4L\Big(-\gamma_n + \gamma_n^2 L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|\xi\|^2 + 1)^{(L-1)}\Big)\mathfrak{L}_\infty^n(\Theta_n) \\
&\leqslant 4L\Big(-\gamma_n + \gamma_n \mathfrak{g} L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|\xi\|^2 + 1)^{(L-1)}\Big)\mathfrak{L}_\infty^n(\Theta_n) \\
&= -4\gamma_n L\Big(1 - \mathfrak{g} L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|\xi\|^2 + 1)^{(L-1)}\Big)\mathfrak{L}_\infty^n(\Theta_n) \leqslant 0.
\end{aligned} \tag{4.229}$$

Induction thus establishes (4.226). The proof of Corollary 4.4.10 is thus complete. □

4.4.7 Convergence analysis for SGD processes

The following theorem proves that in the training of DNNs it holds that the sequence of risks of any time-discrete SGD processes converges to zero provided that the target function is constant and the SGD steps are sufficiently small but not L^1-summable. For this reason, Theorem 4.4.11 extends [63, Theorem 3.12] from shallow ANNs to DNNs with an arbitrary number of hidden layers.

Theorem 4.4.11. *Assume Setting 4.4.1, assume $\sum_{n=0}^{\infty}\gamma_n = \infty$, and let $\delta \in (0,1)$ satisfy*

$$\inf\nolimits_{n\in\mathbb{N}} \mathbb{P}\Big(\gamma_n L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|\xi\|^2 + 1)^{(L-1)} \leqslant \delta\Big) = 1. \tag{4.230}$$

Then

(i) there exists $\mathfrak{C} \in \mathbb{R}$ such that $\mathbb{P}(\sup_{n\in\mathbb{N}_0}\|\Theta_n\| \leqslant \mathfrak{C}) = 1$,

(ii) it holds that $\mathbb{P}(\limsup_{n\to\infty}\mathcal{L}(\Theta_n) = 0) = 1$, and

(iii) it holds that $\limsup_{n\to\infty}\mathbb{E}[\mathcal{L}(\Theta_n)] = 0$.

Proof of Theorem 4.4.11. Throughout this proof let $\mathfrak{g} \in [0,\infty]$ satisfy $\mathfrak{g} = \sup_{n\in\mathbb{N}_0}\gamma_n$. Observe that (4.230) and the assumption that $\sum_{n=0}^{\infty}\gamma_n = \infty$ ensure that $\mathfrak{g} \in (0,\infty)$. This and (4.230) imply that there exists $\mathfrak{C} \in [1,\infty)$ which satisfies

$$\mathbb{P}(V(\Theta_0) \leqslant \mathfrak{C}) = 1. \tag{4.231}$$

Note that (4.231) and Corollary 4.4.10 demonstrates that

$$\mathbb{P}\big(\sup\nolimits_{n\in\mathbb{N}_0} V(\Theta_n) \leqslant \mathfrak{C}\big) = 1. \tag{4.232}$$

Item (ii) of Proposition 4.2.1 (applied with $\mu \curvearrowleft (\mathcal{B}([a,b]^{\ell_0}) \ni A \mapsto \frac{1}{M_n}\sum_{m=1}^{M_n}\mathbb{1}_A(X^{n,m}(\omega)) \in [0,1])$, $f \curvearrowleft ([a,b]^{\ell_0} \ni x \mapsto \xi \in \mathbb{R}^{\ell_L})$ in the notation of Proposition 4.2.1) therefore shows that

$$\mathbb{P}\big(\sup\nolimits_{n\in\mathbb{N}_0}\|\Theta_n\| \leqslant \mathfrak{C}\big) = 1. \tag{4.233}$$

This establishes item (i). Observe that Corollary 4.4.10 and (4.230) ensure that for all $n \in \mathbb{N}_0$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned}
&-(V(\Theta_n) - V(\Theta_{n+1})) \\
&\leqslant -4\gamma_n L\big(1 - \mathfrak{g} L\mathbf{a}^2\big[\textstyle\prod_{p=0}^{L}(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|\xi\|^2 + 1)^{(L-1)}\big)\mathfrak{L}_\infty^n(\Theta_n).
\end{aligned} \tag{4.234}$$

This and (4.230) prove that for all $n \in \mathbb{N}_0$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} &\gamma_n \mathfrak{L}^n_\infty(\Theta_n) \\ &\leqslant \frac{V(\Theta_n) - V(\Theta_{n+1})}{4L[1 - \mathfrak{g}L\mathbf{a}^2[\prod_{p=0}^{L}(\ell_p+1)](2V(\Theta_0) + 4L^2\|\xi\|^2 + 1)^{(L-1)}]} \\ &\leqslant \frac{V(\Theta_n) - V(\Theta_{n+1})}{4L(1-\delta)}. \end{aligned} \tag{4.235}$$

Furthermore, note that item (ii) of Proposition 4.2.1 (applied with $\mu \curvearrowleft (\mathcal{B}([a,b]^{\ell_0}) \ni A \mapsto \frac{1}{M_n}\sum_{m=1}^{M_n} \mathbb{1}_A(X^{n,m}(\omega)) \in [0,1])$, $f \curvearrowleft ([a,b]^{\ell_0} \ni x \mapsto \xi \in \mathbb{R}^{\ell_L})$ in the notation of Proposition 4.2.1) ensures that for all $n \in \mathbb{N}_0$ it holds $\mathbb{P}$-a.s. that $V(\Theta_n) \geqslant \frac{1}{2}\|\Theta_n\|^2 - 2L^2\|f(0)\|^2 \geqslant -2L^2\|f(0)\|^2$. Combining this with (4.232) and (4.235) implies that for all $N \in \mathbb{N}_0$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} &\sum_{n=0}^{N-1} \gamma_n \mathfrak{L}^n_\infty(\Theta_n) \leqslant \frac{\sum_{n=0}^{N-1}(V(\Theta_n) - V(\Theta_{n+1}))}{4L(1-\delta)} = \frac{V(\Theta_0) - V(\Theta_N)}{4L(1-\delta)} \\ &\leqslant \frac{V(\Theta_0) + 2L^2\|f(0)\|^2}{4L(1-\delta)} \leqslant \frac{\mathfrak{C} + 2L^2\|f(0)\|^2}{4L(1-\delta)} < \infty. \end{aligned} \tag{4.236}$$

Hence, we obtain that

$$\textstyle\sum_{n=0}^{\infty} \gamma_n \mathbb{E}[\mathfrak{L}^n_\infty(\Theta_n)] = \lim_{N\to\infty}\big[\sum_{n=0}^{N-1}\gamma_n \mathfrak{L}^n_\infty(\Theta_n)\big] \leqslant \frac{\mathfrak{C}+2L^2\|f(0)\|^2}{4L(1-\delta)} < \infty. \tag{4.237}$$

Moreover, observe that Corollary 4.4.5 proves that for all $n \in \mathbb{N}_0$ it holds that $\mathbb{E}[\mathfrak{L}^n_\infty(\Theta_n)] = \mathbb{E}[\mathcal{L}(\Theta_n)]$. This and (4.237) show that $\sum_{n=0}^{\infty}\gamma_n\mathbb{E}[\mathcal{L}(\Theta_n)] < \infty$. The monotone convergence theorem and the fact that for all $n \in \mathbb{N}_0$, $\omega \in \Omega$ it holds that $\mathcal{L}(\Theta_n(\omega)) \geqslant 0$ therefore ensure that

$$\textstyle\mathbb{E}\big[\sum_{n=0}^{\infty}\gamma_n\mathcal{L}(\Theta_n)\big] = \sum_{n=0}^{\infty}\mathbb{E}[\gamma_n\mathcal{L}(\Theta_n)] < \infty. \tag{4.238}$$

Hence, we obtain that

$$\textstyle\mathbb{P}\big(\sum_{n=0}^{\infty}\gamma_n\mathcal{L}(\Theta_n) < \infty\big) = 1. \tag{4.239}$$

Next let $A \subseteq \Omega$ satisfy

$$\textstyle A = \big\{\omega \in \Omega \colon \big[\big(\sum_{n=0}^{\infty}\gamma_n\mathcal{L}(\Theta_n(\omega)) < \infty\big) \wedge \big(\sup_{n\in\mathbb{N}_0}\|\Theta_n(\omega)\| \leqslant \mathfrak{C}\big)\big]\big\}. \tag{4.240}$$

Note that (4.233), (4.239) and (4.240) ensure that $A \in \mathcal{F}$ and $\mathbb{P}(A) = 1$. In the following let $\omega \in A$ be arbitrary. Observe that (4.240) ensures that

$$\sup_{n\in\mathbb{N}_0}\|\Theta_n(\omega)\| \leqslant \mathfrak{C}. \tag{4.241}$$

In addition, note that the assumption that $\sum_{n=0}^{\infty}\gamma_n = \infty$ and the fact that $\sum_{n=0}^{\infty}\gamma_n\mathcal{L}(\Theta_n(\omega)) < \infty$ show that $\liminf_{n\to\infty}\mathcal{L}(\Theta_n(\omega)) = 0$. In the following we prove item (ii) by contradiction. For this assume that

$$\limsup_{n\to\infty}\mathcal{L}(\Theta_n(\omega)) > 0. \tag{4.242}$$

This implies that there exists $\varepsilon \in (0,\infty)$ which satisfies

$$0 = \liminf_{n\to\infty}\mathcal{L}(\Theta_n(\omega)) < \varepsilon < 2\varepsilon < \limsup_{n\to\infty}\mathcal{L}(\Theta_n(\omega)). \tag{4.243}$$

Observe that (4.243) assures that there exist $(m_k, n_k) \in \mathbb{N}^2$, $k \in \mathbb{N}$, which satisfy that for all $k \in \mathbb{N}$ it holds that $m_k < n_k < m_{k+1}$, $\mathcal{L}(\Theta_{m_k}(\omega)) > 2\varepsilon$, and $\mathcal{L}(\Theta_{n_k}(\omega)) < \varepsilon <$

$\min_{j\in\mathbb{N}\cap[m_k,n_k)} \mathcal{L}(\Theta_j(\omega))$. Combining this with the fact that $\sum_{n=0}^{\infty} \gamma_n \mathcal{L}(\Theta_n(\omega)) < \infty$ shows that

$$\sum_{k=1}^{\infty}\sum_{j=m_k}^{n_k-1} \gamma_j \leqslant \frac{1}{\varepsilon}\left[\sum_{k=1}^{\infty}\sum_{j=m_k}^{n_k-1} (\gamma_j \mathcal{L}(\Theta_j(\omega)))\right] \leqslant \frac{1}{\varepsilon}\left[\sum_{k=0}^{\infty}(\gamma_k \mathcal{L}(\Theta_k(\omega)))\right] < \infty. \tag{4.244}$$

Furthermore, note that (4.241) and Lemma 4.4.7 ensure that there exists $\mathfrak{D} \in \mathbb{R}$ which satisfies for all $n \in \mathbb{N}_0$ that $\|\mathfrak{G}^n(\Theta_n(\omega),\omega)\| \leqslant \mathfrak{D}$. This, (4.244), and the fact that for all $n \in \mathbb{N}_0$, $\omega \in \Omega$ it holds that $\Theta_{n+1}(\omega) - \Theta_n(\omega) = -\gamma_n \mathfrak{G}^n(\Theta_n(\omega),\omega)$ demonstrate that

$$\begin{aligned}\sum_{k=1}^{\infty}\|\Theta_{n_k}(\omega) - \Theta_{m_k}(\omega)\| &\leqslant \sum_{k=1}^{\infty}\sum_{j=m_k}^{n_k-1} \|\Theta_{j+1}(\omega) - \Theta_j(\omega)\| \\ &= \sum_{k=1}^{\infty}\sum_{j=m_k}^{n_k-1} \gamma_j \|\mathfrak{G}^j(\Theta_j(\omega),\omega)\| \leqslant \mathfrak{D}\left[\sum_{k=1}^{\infty}\sum_{j=m_k}^{n_k-1}\gamma_j\right] < \infty.\end{aligned} \tag{4.245}$$

Moreover, observe that Lemma 4.1.10 (applied with $\mu \curvearrowleft (\mathcal{B}([a,b]^{\ell_0}) \ni A \mapsto \frac{1}{M_n}\sum_{m=1}^{M_n} \cdot \mathbb{1}_A(X^{n,m}(\omega)) \in [0,1])$, $f \curvearrowleft ([a,b]^{\ell_0} \ni x \mapsto \xi \in \mathbb{R}^{\ell_L})$ in the notation of Lemma 4.1.10) and (4.241) prove that there exists $\mathscr{L} \in \mathbb{R}$ which satisfies for all $m,n \in \mathbb{N}_0$ that $|\mathcal{L}(\Theta_m(\omega)) - \mathcal{L}(\Theta_n(\omega))| \leqslant \mathscr{L}\|\Theta_m(\omega) - \Theta_n(\omega)\|$. Combining this with (4.245) proves that

$$\limsup_{k\to\infty}|\mathcal{L}(\Theta_{n_k}(\omega)) - \mathcal{L}(\Theta_{m_k}(\omega))| \leqslant \limsup_{k\to\infty}\big(\mathscr{L}\|\Theta_{n_k}(\omega) - \Theta_{m_k}(\omega)\|\big) = 0. \tag{4.246}$$

This and the fact that for all $k \in \mathbb{N}_0$ it holds that $\mathcal{L}(\Theta_{n_k}(\omega)) < \varepsilon < 2\varepsilon \leqslant \mathcal{L}(\Theta_{m_k})$ demonstrate that

$$\begin{aligned}0 < \varepsilon &\leqslant \inf_{k\in\mathbb{N}}|\mathcal{L}(\Theta_{n_k}(\omega)) - \mathcal{L}(\Theta_{m_k}(\omega))| \\ &\leqslant \limsup_{k\to\infty}|\mathcal{L}(\Theta_{n_k}(\omega)) - \mathcal{L}(\Theta_{m_k}(\omega))| = 0.\end{aligned} \tag{4.247}$$

This contradiction proves that $\limsup_{n\to\infty} \mathcal{L}(\Theta_n(\omega)) = 0$. Combining this with the fact that $\mathbb{P}(A) = 1$ establishes item (ii). Note that item (i) and the fact that $\mathcal{L}$ is continuous show that there exists $\mathfrak{C} \in \mathbb{R}$ which satisfies that

$$\mathbb{P}(\sup_{n\in\mathbb{N}_0} \mathcal{L}(\Theta_n) \leqslant \mathfrak{C}) = 1. \tag{4.248}$$

Observe that item (ii), (4.248), and the dominated convergence theorem establish item (iii). The proof of Theorem 4.4.11 is thus complete. □

The following corollary, Corollary 4.4.12, is a special case of Theorem 4.4.11 and generalizes [63, Corollary 3.13].

Corollary 4.4.12. Assume Setting 4.4.1, assume $\sum_{n=0}^{\infty} \gamma_n = \infty$, and assume for all $n \in \mathbb{N}_0$ that

$$\mathbb{P}\big((4L\mathfrak{d}\max\{\mathbf{a}, \|\xi\|\})^{2L}\gamma_n \leqslant (\|\Theta_0\| + 1)^{-2L}\big) = 1. \tag{4.249}$$

Then

(i) there exists $\mathfrak{C} \in \mathbb{R}$ such that $\mathbb{P}(\sup_{n\in\mathbb{N}_0}\|\Theta_n\| \leqslant \mathfrak{C}) = 1$,

(ii) it holds that $\mathbb{P}(\limsup_{n\to\infty} \mathcal{L}(\Theta_n) = 0) = 1$, and

(iii) it holds that $\limsup_{n\to\infty} \mathbb{E}[\mathcal{L}(\Theta_n)] = 0$.

Proof of Corollary 4.4.12. Throughout this proof let $\mathbf{B} \in \mathbb{R}$ satisfy $\mathbf{B} = \max\{\mathbf{a}, \|\xi\|\}$. Note that item (ii) of Proposition 4.2.1 (applied with $\mu \curvearrowleft (\mathcal{B}([a,b]^{\ell_0}) \ni A \mapsto \frac{1}{M_n}\sum_{m=1}^{M_n} \cdot \mathbb{1}_A(X^{n,m}(\omega)) \in [0,1])$, $f \curvearrowleft ([a,b]^{\ell_0} \ni x \mapsto \xi \in \mathbb{R}^{\ell_L})$ in the notation of Proposition 4.2.1) and the fact that for all $x, y \in \mathbb{R}$, $M \in \mathbb{N}$ it holds that $(x+y)^M \leqslant (2^{M+1}-1)(x^M + y^M)$ ensure that it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} &(2V(\Theta_0) + 4L^2\|\xi\|^2 + 1)^{(L-1)} \leqslant \big(2(2L\|\Theta_0\|^2 + L\|\xi\|^2) + 4L^2\|\xi\|^2 + 1\big)^{(L-1)} \\ &\leqslant (4L\|\Theta_0\|^2 + 6L^2\|\xi\|^2 + 1)^{(L-1)} \\ &\leqslant (2^L - 1)\big((4L\|\Theta_0\|^2)^{(L-1)} + (6L^2\|\xi\|^2 + 1)^{(L-1)}\big). \end{aligned} \tag{4.250}$$

Therefore, we obtain that it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} &(2V(\Theta_0) + 4L^2\|\xi\|^2 + 1)^{(L-1)} \leqslant (2^L - 1)\Big[(4L)^{(L-1)}\|\Theta_0\|^{2(L-1)} + (7\mathbf{B}^2L^2)^{(L-1)}\Big] \\ &\leqslant (2^L-1)(7L^2)^{(L-1)}\mathbf{B}^{2(L-1)}(\|\Theta_0\|^{2(L-1)} + 1) \leqslant 14^L L^{2(L-1)}\mathbf{B}^{2(L-1)}(\|\Theta_0\| + 1)^{2L}. \end{aligned} \tag{4.251}$$

This and (4.249) show that for all $n \in \mathbb{N}_0$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} &\gamma_n L\mathbf{a}^2\big[\textstyle\prod_{p=0}^L(\ell_p+1)\big](2V(\Theta_0) + 4L^2\|\xi\|^2 + 1)^{(L-1)} \\ &\leqslant \gamma_n L^{2L-1}\mathbf{B}^{2L}14^L\big[\textstyle\prod_{p=0}^L(\ell_p+1)\big](\|\Theta_0\|+1)^{2L} \\ &\leqslant \gamma_n 14^L(L\mathbf{B})^{2L}\big[\textstyle\prod_{p=0}^L(\ell_p+1)\big](\|\Theta_0\|+1)^{2L} \\ &\leqslant \gamma_n 14^L(L\mathbf{B})^{2L}\mathfrak{d}^{L+1}(\|\Theta_0\|+1)^{2L} \\ &\leqslant \gamma_n 14^L(L\mathbf{B}\mathfrak{d})^{2L}(\|\Theta_0\|+1)^{2L} \leqslant 14^L 4^{-2L} = 7^L 8^{-L} < 1. \end{aligned} \tag{4.252}$$

Combining this with Theorem 4.4.11 establishes items (i), (ii) and (iii). The proof of Corollary 4.4.12 is thus complete. □

4.5 Numerical simulations

In this section we illustrate the theoretical findings of Section 4.4 by two numerical examples. First, we study a shallow ANN with one hidden layer in Subsection 4.5.1. Then we consider a deep ANN with two hidden layers in Subsection 4.5.2.

4.5.1 Numerical simulation of an SGD processes for certain ANNs with one hidden layer

In this subsection we present numerical simulations for a certain SGD process in the training of some shallow ANNs with 1 neuron on the input layer, with 7 neurons on the hidden layer, and with 1 neuron on the output layer (see Figure 4.2, Table 4.1, and Figure 4.3). More formally, assume Setting 4.4.1, let $e_1, e_2, \ldots, e_{\mathfrak{d}} \in \mathbb{R}^{\mathfrak{d}}$ satisfy $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, \ldots, 0)$, $\ldots$, $e_{\mathfrak{d}} = (0, \ldots, 0, 1)$, assume

$$L = 2, \qquad \ell_0 = 1, \qquad \ell_1 = 7, \qquad \ell_2 = 1, \qquad \xi = 1, \qquad a = 0, \qquad \text{and} \qquad b = 1, \tag{4.253}$$

assume for all $k \in \{1, 2, \ldots, \mathfrak{d}\}$, $x \in \mathbb{R}$ that $\mathbb{P}(X^{0,0} < x) = \max\{\min\{x, 1\}, 0\}$ and

$$\mathbb{P}(\langle e_k, (\ell_1)^{1/2}\Theta_0\rangle < x) = \textstyle\int_{-\infty}^x (2\pi)^{-1/2}\exp\big(-\frac{y^2}{2}\big)\,\mathrm{d}y, \tag{4.254}$$

and assume for all $n \in \mathbb{N}_0$ that $M_n = 100$ and $\gamma_n = \frac{1}{2000}$. Observe that (4.253) ensures that the number of ANN parameters $\mathfrak{d}$ satisfies

$$\mathfrak{d} = \textstyle\sum_{k=1}^L \ell_k(\ell_{k-1}+1) = 2\ell_1 + \ell_1 + 1 = 3\ell_1 + 1 = 22. \tag{4.255}$$

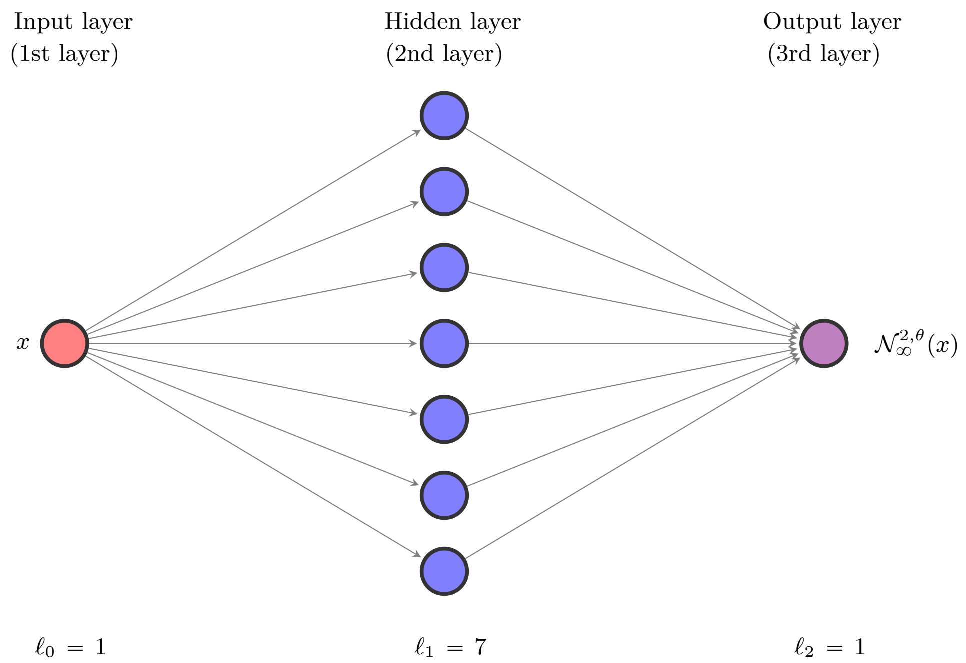

Figure 4.2: Graphical illustration of the ANN architecture used in Subsection 4.5.1 ($\ell_0 = 1$ neuron on the input layer, $\ell_1 = 7$ neurons on the hidden layer, and $\ell_2 = 1$ neuron on the output layer). In this situation we have for every ANN parameter vector $\theta \in \mathbb{R}^{\mathfrak{d}} = \mathbb{R}^{22}$ that the realization function $\mathbb{R} \ni x \mapsto \mathcal{N}_\infty^{2,\theta}(x) \in \mathbb{R}$ of the considered ANN maps the 1-dimensional input $x \in [0, 1]$ to the 1-dimensional output $\mathcal{N}_\infty^{2,\theta}(x) \in \mathbb{R}$.

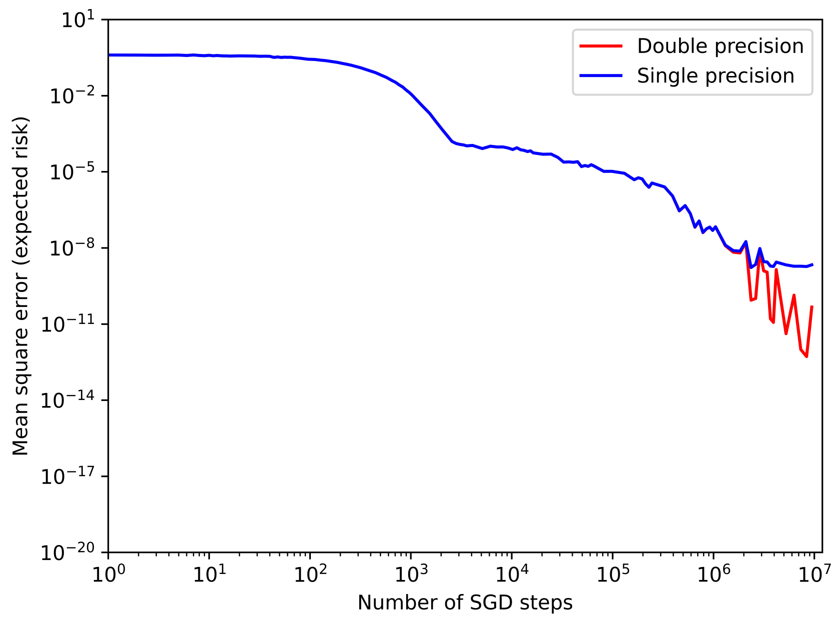

Figure 4.3: Plot of the estimated mean square error (expected risk) against the number of SGD steps for Subsection 4.5.1

Furthermore, note that (4.253) assures that the considered ANNs consist of 14 weight parameters and 8 bias parameters. We also refer to Figure 4.2 for a graphical illustration of the ANN architecture described in (4.253). Moreover, observe that (4.254) ensures that for all $n, m \in \mathbb{N}_0$ we have that $X^{n,m}$ is continuous uniformly distributed on $(0, 1)$. In addition, note that (4.254) assures that $(\ell_1)^{1/2}\Theta_0$ is a 22-dimensional standard normal random vector. In Table 4.1 we approximately specify the mean square error (expected risk)

$$\mathbb{E}[\mathcal{L}(\Theta_n)] \tag{4.256}$$

against the number n of SGD steps for several $n \in \{1, 2, \ldots, 10^6\}$ and in Figure 4.3 we approximately plot the values of Table 4.1. In Table 4.1 and Figure 4.3 we approximate the expectations in (4.256) by means of Monte Carlo averages with 10^6 samples. The Python source code used to create Table 4.1 and Figure 4.3 can be found in Source Code 4.1. Moreover, in Source Code 4.2 we present a simplified variant of the Python code in Source Code 4.1.

Table 4.1: Estimated mean square error (expected risk) against the number of SGD steps for Subsection 4.5.1

Number of SGD steps	Estimated mean square error (expected risk)	
	Single precision	Double precision
1	0.400 147 825 479 507	0.400 147 757 334 473
2	0.396 873 384 714 127	0.396 873 312 108 460
3	0.392 277 061 939 240	0.392 276 965 724 788
4	0.394 556 581 974 030	0.394 556 505 069 431
5	0.397 466 093 301 773	0.397 465 997 116 809
6	0.382 976 621 389 389	0.382 976 568 877 238
7	0.400 645 166 635 513	0.400 645 150 542 984
8	0.384 297 370 910 645	0.384 297 314 109 511
9	0.375 544 518 232 346	0.375 544 491 625 370
10	0.389 181 315 898 895	0.389 181 306 471 692
11	0.373 209 863 901 138	0.373 209 880 330 711
12	0.382 972 270 250 320	0.382 972 322 131 645
13	0.373 731 762 170 792	0.373 731 763 163 821
14	0.368 176 460 266 113	0.368 176 516 276 078
15	0.368 835 985 660 553	0.368 836 011 472 468
16	0.363 389 700 651 169	0.363 389 745 330 901
20	0.370 313 495 397 568	0.370 313 523 500 136
24	0.367 217 361 927 032	0.367 217 434 342 550
28	0.364 903 062 582 016	0.364 903 210 618 829
32	0.355 952 948 331 833	0.355 953 070 249 735
36	0.358 898 729 085 922	0.358 898 892 476 224
40	0.355 168 431 997 299	0.355 168 573 533 324
44	0.321 540 951 728 821	0.321 541 034 787 801
48	0.335 629 969 835 281	0.335 630 057 706 390
52	0.319 862 246 513 367	0.319 862 350 559 697
56	0.328 886 300 325 394	0.328 886 373 083 635
60	0.323 966 592 550 278	0.323 966 670 151 027
64	0.326 734 960 079 193	0.326 734 973 376 403
80	0.298 752 039 670 944	0.298 752 152 929 733
96	0.272 277 683 019 638	0.272 277 733 625 831
112	0.266 797 602 176 666	0.266 797 774 851 729
128	0.250 351 667 404 175	0.250 351 894 209 231
144	0.239 569 738 507 271	0.239 570 030 816 244
160	0.224 916 756 153 107	0.224 917 150 023 337
176	0.212 934 911 251 068	0.212 935 429 076 293
192	0.201 422 691 345 215	0.201 423 289 459 759
208	0.187 521 025 538 445	0.187 521 578 610 189
224	0.177 595 525 979 996	0.177 596 129 156 231
240	0.167 538 091 540 337	0.167 538 661 745 551
256	0.158 610 120 415 688	0.158 610 717 976 725
320	0.125 099 778 175 354	0.125 100 202 566 862
384	0.098 203 480 243 683	0.098 203 721 155 615
448	0.080 468 311 905 861	0.080 468 544 308 340
512	0.063 573 107 123 375	0.063 573 213 667 754

Table 4.1: Estimated mean square error (expected risk) against the number of SGD steps for Subsection 4.5.1

Number of SGD steps	Estimated mean square error (expected risk)	
	Single precision	Double precision
576	0.051 985 166 966 915	0.051 985 308 415 053
640	0.040 908 116 847 277	0.040 908 288 056 718
704	0.033 746 954 053 640	0.033 747 045 870 249
768	0.026 494 434 103 370	0.026 494 502 661 920
832	0.021 851 047 873 497	0.021 851 090 644 621
896	0.016 827 533 021 569	0.016 827 623 175 335
960	0.013 607 602 566 481	0.013 607 615 542 691
1024	0.010 773 501 358 926	0.010 773 505 154 788
1280	0.004 222 723 655 403	0.004 222 709 641 419
1536	0.002 011 111 238 971	0.002 011 089 506 127
1792	0.000 904 137 094 039	0.000 904 125 545 469
2048	0.000 461 409 537 820	0.000 461 409 349 747
2304	0.000 260 941 335 000	0.000 260 947 146 381
2560	0.000 156 190 508 278	0.000 156 196 426 245
2816	0.000 129 971 638 671	0.000 129 974 761 110
3072	0.000 119 722 702 948	0.000 119 724 364 107
3328	0.000 114 423 724 881	0.000 114 423 680 885
3584	0.000 104 955 994 175	0.000 104 955 114 117
3840	0.000 107 468 149 508	0.000 107 467 034 215
4096	0.000 109 119 770 059	0.000 109 117 431 248
5120	0.000 082 289 298 007	0.000 082 288 763 445
6144	0.000 101 874 691 609	0.000 101 874 056 815
7168	0.000 094 819 704 827	0.000 094 819 460 726
8192	0.000 095 279 065 135	0.000 095 279 867 034
9216	0.000 086 695 865 321	0.000 086 697 109 961
10240	0.000 075 270 487 287	0.000 075 272 454 079
11264	0.000 088 507 214 969	0.000 088 508 677 075
12288	0.000 073 549 970 693	0.000 073 551 115 951
13312	0.000 069 636 291 300	0.000 069 636 594 517
14336	0.000 062 663 675 635	0.000 062 665 247 234
15360	0.000 066 875 043 558	0.000 066 874 971 056
16384	0.000 055 498 523 579	0.000 055 497 834 259
20480	0.000 049 061 156 460	0.000 049 058 356 528
24576	0.000 049 805 254 093	0.000 049 802 507 159
28672	0.000 036 901 041 312	0.000 036 899 520 277
32768	0.000 024 025 355 742	0.000 024 024 804 251
36864	0.000 024 539 234 801	0.000 024 536 956 892
40960	0.000 023 680 300 728	0.000 023 680 059 877
45056	0.000 025 016 855 943	0.000 025 013 418 711
49152	0.000 016 035 986 846	0.000 016 033 751 002
53248	0.000 017 649 770 598	0.000 017 648 362 293
57344	0.000 016 380 216 039	0.000 016 378 699 475
61440	0.000 018 981 387 257	0.000 018 978 539 977
65536	0.000 016 854 279 238	0.000 016 853 033 162

Table 4.1: Estimated mean square error (expected risk)
against the number of SGD steps for Subsection 4.5.1

Number of SGD steps	Estimated mean square error (expected risk)	
	Single precision	Double precision
81920	0.000 010 368 178 664	0.000 010 369 734 776
98304	0.000 010 430 661 860	0.000 010 432 632 213
114688	0.000 009 490 538 105	0.000 009 493 783 692
131072	0.000 008 730 737 136	0.000 008 732 925 002
147456	0.000 006 400 678 103	0.000 006 398 105 292
163840	0.000 004 872 399 131	0.000 004 867 179 719
180224	0.000 005 805 000 455	0.000 005 801 188 995
196608	0.000 005 238 976 428	0.000 005 228 108 959
212992	0.000 003 307 908 855	0.000 003 295 626 299
229376	0.000 002 462 194 061	0.000 002 440 680 873
245760	0.000 003 611 708 735	0.000 003 593 632 356
262144	0.000 003 336 253 030	0.000 003 317 915 866
327680	0.000 002 536 854 481	0.000 002 520 310 529
393216	0.000 001 119 518 060	0.000 001 108 944 636
458752	0.000 000 290 656 800	0.000 000 286 179 921
524288	0.000 000 465 935 557	0.000 000 462 486 579
589824	0.000 000 226 278 445	0.000 000 224 031 931
655360	0.000 000 066 586 750	0.000 000 065 128 884
720896	0.000 000 115 260 363	0.000 000 113 909 821
786432	0.000 000 040 924 679	0.000 000 039 977 428
851968	0.000 000 056 510 213	0.000 000 055 672 129
917504	0.000 000 066 116 016	0.000 000 065 633 661
983040	0.000 000 048 902 429	0.000 000 048 764 219
1048576	0.000 000 067 959 760	0.000 000 067 938 356
1310720	0.000 000 012 907 666	0.000 000 012 231 249
1572864	0.000 000 007 840 144	0.000 000 006 764 327
1835008	0.000 000 007 447 369	0.000 000 006 311 093
2097152	0.000 000 017 747 798	0.000 000 016 540 680
2359296	0.000 000 001 681 142	0.000 000 000 086 757
2621440	0.000 000 002 234 637	0.000 000 000 101 382
2883584	0.000 000 009 444 941	0.000 000 007 922 566
3145728	0.000 000 002 858 044	0.000 000 001 248 282
3407872	0.000 000 002 784 514	0.000 000 001 107 729
3670016	0.000 000 001 924 778	0.000 000 000 016 104
3932160	0.000 000 001 853 827	0.000 000 000 011 564
4194304	0.000 000 002 756 206	0.000 000 001 396 554
5242880	0.000 000 002 140 371	0.000 000 000 004 159
6291456	0.000 000 001 891 979	0.000 000 000 137 176
7340032	0.000 000 001 896 622	0.000 000 000 000 992
8388608	0.000 000 001 844 519	0.000 000 000 000 522
9437184	0.000 000 002 155 239	0.000 000 000 046 701

```python
import numpy as np
import torch
import copy
import matplotlib.pyplot as plt

# use GPU if available
dev = torch.device("cuda") if torch.cuda.is_available() else torch.device("cpu")

# set precision to double
torch.set_default_dtype(torch.float64)

# configure the training parameters
steps = 10000000              # n
batch_size = 100              # M_n, number of sample points
gamma = 1/2000                # step size (learning rate)

# get sample data
torch.manual_seed(0)
sample_data_double = (torch.rand(batch_size*(steps+1), 1)).to(dev)
sample_data_single = sample_data_double.type(torch.float32)

# set parameters
l_0, l_1, l_2, = 1, 7, 1         # dimensions of the layers
a, b = 0, 1                      # the domain will be [a,b]
xi = 1                           # constant target function


# define loss function
def loss(N, x):
    y = N(x)
    return (y-xi).square().mean()


# function setting the starting model parameters as normally distributed
def init_weights(n):
    if isinstance(n, torch.nn.Linear):
        torch.nn.init.normal_(n.weight, 0, 1/np.sqrt(l_1))
        torch.nn.init.normal_(n.bias, 0, 1/np.sqrt(l_1))


# set up NN model
N_double = torch.nn.Sequential(
    torch.nn.Linear(l_0, l_1),
    torch.nn.ReLU(),
    torch.nn.Linear(l_1, l_2)
).to(dev)

# set up normally distributed starting model parameters
torch.manual_seed(0)
N_double.apply(init_weights)
N_single = copy.deepcopy(N_double)
N_single.type(torch.float32)

# define observation points
obs = np.empty(shape=129)
index = 0
for i in range(0, 23, 2):
    for k in range(2**i, min(2**(i+2), steps)):
```

```
        if k % max(1, 2**(i-2)) == False:
            obs[index] = k
            index = index + 1

# set up table
losses_table = np.empty((len(obs), 3), dtype=object)

# approximation in double precision
# optimizer function
optimizer_double = torch.optim.SGD(N_double.parameters(), lr=gamma)

# train the network
for step in range(steps+1):
    # generate uniformly distributed samples from [0,1]
    x = sample_data_double[step*batch_size:(step+1)*batch_size]
    # set the gradients back to zero
    optimizer_double.zero_grad()
    # compute the loss
    L = loss(N_double, x)
    # compute the gradients of L w.r.t. the model parameters
    L.backward()
    # update the model parameters
    optimizer_double.step()
    # save results in the table
    if step in obs:
        losses_table[np.where(obs == step)[0], [0, 2]] = step, L.data

# approximation in single precision
# optimizer function
optimizer_single = torch.optim.SGD(N_single.parameters(), lr=gamma)

# train the network
for step in range(steps+1):
    # generate uniformly distributed samples from [0,1]
    x = sample_data_single[step*batch_size:(step+1)*batch_size]
    # set the gradients back to zero
    optimizer_single.zero_grad()
    # compute the loss
    L = loss(N_single, x)
    # compute the gradients of L w.r.t. the model parameters
    L.backward()
    # update the model parameters
    optimizer_single.step()
    # print and save results in the table
    if step in obs:
        losses_table[np.where(obs == step)[0], [0, 1]] = step, L.data

# save table
np.savetxt('mean_square_error_example1.csv', losses_table, delimiter=',',
           fmt='%s, %.64f, %.64f')

# plot mean square error
plt.loglog(losses_table[:, 0], losses_table[:, 2], "r", label='Double precision')
plt.loglog(losses_table[:, 0], losses_table[:, 1], "b", label='Single precision')
plt.ylim([0.00000000000000000001, 10])
plt.xlim([1, 12000000])
plt.xlabel('Number of SGD steps')
```

```python
plt.ylabel('Mean square error (expected risk)')
plt.legend()
plt.savefig('mean_square_error_example1_loglog_graph', dpi=1200)
```

Source Code 4.1: PYTHON code used to create Table 4.1 and Figure 4.3 in Subsection 4.5.1 (filename: sgd1.py).

```python
import numpy as np
import torch
import matplotlib.pyplot as plt

# use GPU if available
dev = torch.device("cuda") if torch.cuda.is_available() else torch.↩
    device("cpu")

# set precision to double
torch.set_default_dtype(torch.float64)
torch.manual_seed(0)

# configure the training parameters
steps = 10000000             # n
batch_size = 100             # M_n, number of sample points
gamma = 1/2000               # step size (learning rate)

# set parameters
l_0, l_1, l_2, = 1, 7, 1     # dimensions of the layers
a, b = 0, 1                  # the domain will be [a,b]
xi = 1                       # constant target function


# define loss function
def loss(N, x):
    y = N(x)
    return (y-xi).square().mean()


# function setting the starting model parameters as normally ↩
    distributed
def init_weights(n):
    if isinstance(n, torch.nn.Linear):
        torch.nn.init.normal_(n.weight, 0, 1/np.sqrt(l_1))
        torch.nn.init.normal_(n.bias, 0, 1/np.sqrt(l_1))


# set up NN model
N = torch.nn.Sequential(
    torch.nn.Linear(l_0, l_1),
    torch.nn.ReLU(),
    torch.nn.Linear(l_1, l_2)
).to(dev)

# set the starting model parameters as normally distributed
N.apply(init_weights)

# optimizer function
optimizer = torch.optim.SGD(N.parameters(), lr=gamma)

```

```
# define observation points
obs = np.empty(shape=129)
index = 0
for i in range(0, 23, 2):
    for k in range(2**i, min(2**(i+2), steps)):
        if k % max(1, 2**(i-2)) == False:
            obs[index] = k
            index = index + 1

# train the network
loss_table = np.empty((len(obs), 2), dtype=object)
for step in range(steps+1):
    # generate uniformly distributed samples from [0,1]
    x = (torch.rand(batch_size, 1)).to(dev)
    # set the gradients back to zero
    optimizer.zero_grad()
    # compute the loss
    L = loss(N, x)
    # compute the gradients of L w.r.t. the model parameters
    L.backward()
    # update the model parameters
    optimizer.step()
    # save results in the table
    if step in obs:
        loss_table[np.where(obs == step)[0], :] = step, L.data

# plot mean square error
plt.loglog(loss_table[:, 0], loss_table[:, 1], color= "r")
plt.ylim([0.00000000000000000001, 10])
plt.xlim([1, 12000000])
plt.xlabel('Number of SGD steps')
plt.ylabel('Mean square error (expected risk)')
plt.show()
```

Source Code 4.2: Simplified variant of the PYTHONscript in Source Code 4.1 above (filename: `sgd1_short.py`).

4.5.2 Numerical simulation of an SGD process for certain ANNs with two hidden layers

In this subsection we present numerical simulations for a certain SGD process in the training of some deep ANNs with two hidden layers with 1 neuron on the input layer, with 3 neurons on the first hidden layer, with 7 neurons on the second hidden layer, and with 1 neuron on the output layer (see Figure 4.4, Table 4.2, and Figure 4.5). Assume Setting 4.4.1, let $e_1, e_2, \dots, e_{\mathfrak{d}} \in \mathbb{R}^{\mathfrak{d}}$ satisfy $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, \dots, 0)$, …, $e_{\mathfrak{d}} = (0, \dots, 0, 1)$, assume

$$L = 3, \quad \ell_0 = 1, \quad \ell_1 = 3, \quad \ell_2 = 7, \quad \ell_3 = 1, \quad \xi = 1, \quad a = 0, \quad \text{and} \quad b = 1, \tag{4.257}$$

assume for all $k \in \{1, 2, \dots, \mathfrak{d}\}$, $x \in \mathbb{R}$ that $\mathbb{P}(X^{0,0} < x) = \max\{\min\{x, 1\}, 0\}$ and

$$\mathbb{P}(\langle e_k, (\ell_1)^{1/2}\Theta_0 \rangle < x) = \textstyle\int_{-\infty}^{x} (2\pi)^{-1/2} \exp\!\left(-\frac{y^2}{2}\right) \mathrm{d}y, \tag{4.258}$$

and assume for all $n \in \mathbb{N}_0$ that $M_n = 100$ and $\gamma_n = \frac{1}{2000}$. Observe that (4.257) ensures that the number of deep ANN parameters $\mathfrak{d}$ satisfies

$$\mathfrak{d} = \textstyle\sum_{k=1}^{L} \ell_k(\ell_{k-1} + 1) = 2\ell_1 + \ell_2(\ell_1 + 1) + \ell_2 + 1 = 42. \tag{4.259}$$

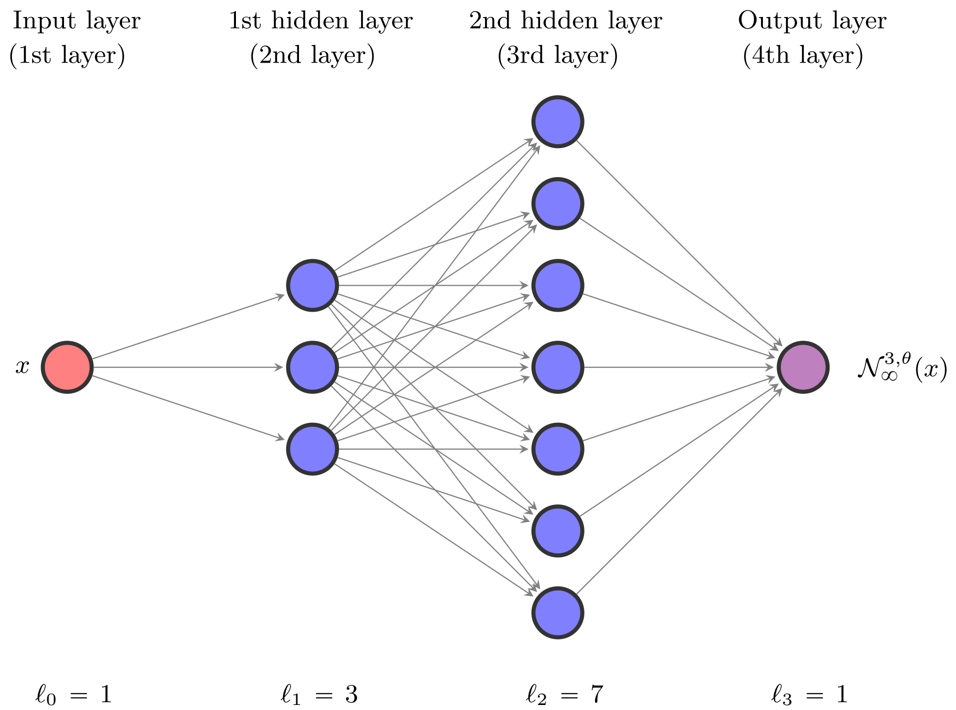

Figure 4.4: Graphical illustration of the ANN architecture used in Subsection 4.5.2 ($\ell_0 = 1$ neuron on the input layer, $\ell_1 = 3$ neurons on the first hidden layer, $\ell_2 = 7$ neurons on the second hidden layer, and $\ell_3 = 1$ neuron on the output layer). In this situation we have for every ANN parameter vector $\theta \in \mathbb{R}^{\mathfrak{d}} = \mathbb{R}^{42}$ that the realization function $\mathbb{R} \ni x \mapsto \mathcal{N}_{\infty}^{3,\theta}(x) \in \mathbb{R}$ of the considered ANN maps the 1-dimensional input $x \in [0,1]$ to the 1-dimensional output $\mathcal{N}_{\infty}^{3,\theta}(x) \in \mathbb{R}$.

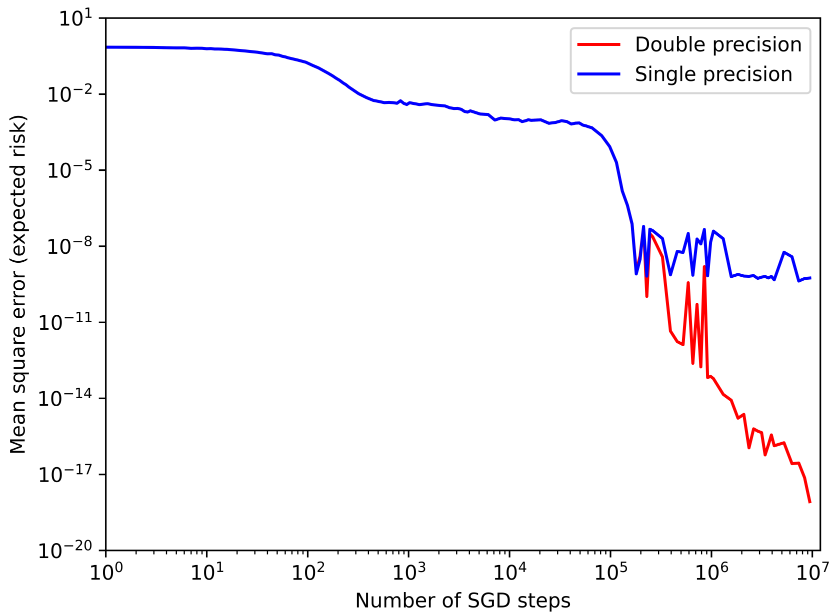

Figure 4.5: Plot of the estimated mean square error (expected risk) against the number of SGD steps for Subsection 4.5.2

Furthermore, note that (4.257) assures that the considered deep ANNs consist of 31 weight parameters and 11 bias parameters. We also refer to Figure 4.4 for a graphical illustration of the deep ANN architecture described in (4.257). Moreover, observe that (4.258) ensures that for all $n, m \in \mathbb{N}_0$ we have that $X^{n,m}$ is continuous uniformly distributed on $(0, 1)$. In addition, note that (4.258) assures that $(\ell_1)^{1/2}\Theta_0$ is a 42-dimensional standard normal random vector. In Table 4.2 we approximately specify the mean square error (expected risk)

$$\mathbb{E}[\mathcal{L}(\Theta_n)] \tag{4.260}$$

against the number n of SGD steps for several $n \in \{1, 2, \dots, 10^6\}$ and in Figure 4.5 we approximately plot the values of Table 4.2. In Table 4.2 and Figure 4.5 we approximate the expectations in (4.260) by means of Monte Carlo averages with 10^6 samples. The Python source code used to create Table 4.2 and Figure 4.5 can be found in Source Code 4.3. In addition, in Source Code 4.4 we present a simplified variant of the Python code in Source Code 4.3.

Table 4.2: Estimated mean square error (expected risk) against the number of SGD steps for Subsection 4.5.2

Number of SGD steps	Estimated mean square error (expected risk)	
	Single precision	Double precision
1	0.706 932 902 336 120 605	0.706 932 967 138 238 499
2	0.700 160 920 619 964 600	0.700 161 001 923 618 631
3	0.692 209 482 192 993 164	0.692 209 648 713 430 048
4	0.674 339 354 038 238 525	0.674 339 512 871 169 444
5	0.663 867 592 811 584 473	0.663 867 713 542 396 265
6	0.665 305 376 052 856 445	0.665 305 491 643 243 152
7	0.637 214 303 016 662 598	0.637 214 477 437 771 265
8	0.644 249 260 425 567 627	0.644 249 489 759 242 344
9	0.640 169 501 304 626 465	0.640 169 765 706 354 976
10	0.614 596 188 068 389 893	0.614 596 437 233 441 484
11	0.622 903 585 433 959 961	0.622 903 790 396 874 713
12	0.599 834 263 324 737 549	0.599 834 490 971 253 409
13	0.599 541 306 495 666 504	0.599 541 385 384 985 026
14	0.594 309 568 405 151 367	0.594 309 729 521 948 493
15	0.584 528 028 964 996 338	0.584 528 235 794 900 985
16	0.581 495 761 871 337 891	0.581 495 960 783 550 370
20	0.536 507 546 901 702 881	0.536 507 626 272 968 663
24	0.502 628 862 857 818 604	0.502 628 907 151 508 431
28	0.470 990 091 562 271 118	0.470 990 082 727 713 422
32	0.446 294 248 104 095 459	0.446 294 283 281 347 737
36	0.413 737 267 255 783 081	0.413 737 193 101 316 181
40	0.387 975 394 725 799 561	0.387 975 368 523 888 475
44	0.391 652 226 448 059 082	0.391 652 243 089 415 986
48	0.352 867 037 057 876 587	0.352 867 130 722 372 913
52	0.343 804 329 633 712 769	0.343 804 420 667 347 455
56	0.310 209 423 303 604 126	0.310 209 522 295 310 847
60	0.293 376 773 595 809 937	0.293 376 803 131 859 376
64	0.269 867 300 987 243 652	0.269 867 384 751 054 873
80	0.219 015 121 459 960 938	0.219 015 261 939 558 031
96	0.180 515 885 353 088 379	0.180 515 989 031 416 441
112	0.135 529 562 830 924 988	0.135 529 872 406 490 698
128	0.108 072 988 688 945 770	0.108 073 233 786 722 886
144	0.083 469 435 572 624 207	0.083 469 614 461 651 020
160	0.066 193 692 386 150 360	0.066 193 793 764 961 095
176	0.052 916 873 246 431 351	0.052 916 943 858 673 952
192	0.041 961 964 219 808 578	0.041 962 060 046 887 420
208	0.034 614 630 043 506 622	0.034 614 665 473 569 797
224	0.027 720 309 793 949 127	0.027 720 308 228 354 095
240	0.023 386 158 049 106 598	0.023 386 173 634 538 612
256	0.018 949 629 738 926 888	0.018 949 635 241 030 473
320	0.010 387 994 349 002 838	0.010 387 979 366 239 279
384	0.007 233 595 941 215 754	0.007 233 591 252 614 298
448	0.005 628 840 532 153 845	0.005 628 851 876 781 546
512	0.005 078 634 712 845 087	0.005 078 631 535 377 190

Table 4.2: Estimated mean square error (expected risk) against the number of SGD steps for Subsection 4.5.2

Number of SGD steps	Estimated mean square error (expected risk)	
	Single precision	Double precision
576	0.004 586 841 445 416 212	0.004 586 845 028 329 780
640	0.004 703 775 513 917 208	0.004 703 779 346 907 735
704	0.004 578 266 292 810 440	0.004 578 280 407 570 739
768	0.004 344 602 581 113 577	0.004 344 615 838 743 911
832	0.005 446 200 259 029 865	0.005 446 222 167 284 568
896	0.004 311 462 398 618 460	0.004 311 483 074 770 449
960	0.003 878 945 717 588 067	0.003 878 958 647 074 864
1024	0.004 567 413 590 848 446	0.004 567 433 820 685 929
1280	0.003 858 407 726 511 359	0.003 858 417 882 718 607
1536	0.004 175 057 634 711 266	0.004 175 069 395 585 085
1792	0.003 732 035 402 208 567	0.003 732 030 766 466 828
2048	0.003 560 277 167 707 682	0.003 560 283 940 305 226
2304	0.003 390 471 450 984 478	0.003 390 475 492 865 211
2560	0.002 882 433 589 547 873	0.002 882 426 419 749 709
2816	0.002 699 112 985 283 136	0.002 699 110 136 409 389
3072	0.002 723 007 462 918 758	0.002 722 985 090 608 629
3328	0.002 500 885 399 058 461	0.002 500 865 152 138 916
3584	0.002 069 632 522 761 822	0.002 069 605 185 433 262
3840	0.001 949 147 321 283 817	0.001 949 128 522 125 018
4096	0.002 179 505 769 163 370	0.002 179 503 693 993 366
5120	0.001 634 338 521 398 604	0.001 634 332 821 930 987
6144	0.001 564 364 531 077 445	0.001 564 358 853 229 929
7168	0.000 946 448 592 003 435	0.000 946 448 757 445 718
8192	0.001 129 143 405 705 690	0.001 129 149 484 013 035
9216	0.001 083 385 315 723 717	0.001 083 391 904 055 565
10240	0.001 037 654 001 265 764	0.001 037 657 068 415 507
11264	0.000 960 678 677 074 611	0.000 960 677 619 513 325
12288	0.000 981 489 778 496 325	0.000 981 488 230 215 470
13312	0.000 824 468 035 716 563	0.000 824 465 739 739 446
14336	0.000 878 931 663 464 755	0.000 878 932 036 012 423
15360	0.000 969 799 584 709 108	0.000 969 792 660 626 467
16384	0.000 923 463 318 031 281	0.000 923 458 249 482 142
20480	0.000 961 496 261 879 802	0.000 961 492 195 195 671
24576	0.000 704 382 197 000 086	0.000 704 390 226 468 234
28672	0.000 762 685 143 854 469	0.000 762 688 985 258 432
32768	0.000 886 771 129 444 242	0.000 886 774 141 138 815
36864	0.000 828 889 256 808 907	0.000 828 894 363 205 109
40960	0.000 659 720 099 065 453	0.000 659 727 355 584 151
45056	0.000 712 938 548 531 383	0.000 712 942 853 875 637
49152	0.000 725 765 014 067 292	0.000 725 755 426 905 822
53248	0.000 598 815 735 429 525	0.000 598 795 392 153 763
57344	0.000 556 507 555 302 233	0.000 556 480 800 988 887
61440	0.000 505 671 196 151 525	0.000 505 635 702 427 932
65536	0.000 469 137 914 478 779	0.000 469 114 685 532 213

Table 4.2: Estimated mean square error (expected risk) against the number of SGD steps for Subsection 4.5.2

Number of SGD steps	Estimated mean square error (expected risk)	
	Single precision	Double precision
81920	0.000 229 416 749 789 380	0.000 229 405 646 326 853
98304	0.000 084 995 779 616 293	0.000 084 993 291 888 553
114688	0.000 020 156 248 865 533	0.000 020 153 375 721 047
131072	0.000 001 501 737 756 371	0.000 001 501 306 134 988
147456	0.000 000 407 918 861 356	0.000 000 404 662 873 524
163840	0.000 000 076 140 302 951	0.000 000 073 643 535 799
180224	0.000 000 000 803 792 644	0.000 000 001 018 668 246
196608	0.000 000 003 045 388 830	0.000 000 004 077 964 365
212992	0.000 000 060 278 104 286	0.000 000 058 752 246 754
229376	0.000 000 000 660 325 405	0.000 000 000 104 608 825
245760	0.000 000 046 855 735 292	0.000 000 031 151 713 762
262144	0.000 000 041 359 754 732	0.000 000 023 552 372 543
327680	0.000 000 020 014 702 784	0.000 000 003 767 749 569
393216	0.000 000 000 740 123 351	0.000 000 000 004 497 087
458752	0.000 000 006 188 518 586	0.000 000 000 001 728 945
524288	0.000 000 005 677 824 877	0.000 000 000 001 306 597
589824	0.000 000 031 732 060 535	0.000 000 000 362 378 984
655360	0.000 000 000 716 110 560	0.000 000 000 000 239 891
720896	0.000 000 019 258 559 192	0.000 000 000 050 518 082
786432	0.000 000 012 250 540 138	0.000 000 000 000 173 275
851968	0.000 000 045 770 057 966	0.000 000 001 553 045 828
917504	0.000 000 000 677 049 472	0.000 000 000 000 066 103
983040	0.000 000 014 426 149 164	0.000 000 000 000 073 339
1048576	0.000 000 039 385 387 396	0.000 000 000 000 059 871
1310720	0.000 000 019 737 989 021	0.000 000 000 000 014 316
1572864	0.000 000 000 630 332 564	0.000 000 000 000 008 420
1835008	0.000 000 000 771 909 037	0.000 000 000 000 001 648
2097152	0.000 000 000 663 941 346	0.000 000 000 000 002 328
2359296	0.000 000 000 649 475 085	0.000 000 000 000 000 111
2621440	0.000 000 000 691 767 865	0.000 000 000 000 000 626
2883584	0.000 000 000 538 107 503	0.000 000 000 000 000 506
3145728	0.000 000 000 602 551 176	0.000 000 000 000 000 444
3407872	0.000 000 000 639 608 255	0.000 000 000 000 000 058
3670016	0.000 000 000 559 835 511	0.000 000 000 000 000 144
3932160	0.000 000 000 643 322 007	0.000 000 000 000 000 360
4194304	0.000 000 000 470 332 384	0.000 000 000 000 000 134
5242880	0.000 000 005 787 236 024	0.000 000 000 000 000 177
6291456	0.000 000 003 829 629 946	0.000 000 000 000 000 026
7340032	0.000 000 000 421 432 528	0.000 000 000 000 000 028
8388608	0.000 000 000 530 893 718	0.000 000 000 000 000 007
9437184	0.000 000 000 555 500 201	0.000 000 000 000 000 001

```
import numpy as np
import torch
import copy
import matplotlib.pyplot as plt

# use GPU if available
dev = torch.device("cuda") if torch.cuda.is_available() else torch.device("cpu")

# set precision to double
torch.set_default_dtype(torch.float64)

# configure the training parameters
steps = 10000000            # n
batch_size = 100            # M_n, number of sample points
gamma = 1/2000              # step size (learning rate)

# get sample data
torch.manual_seed(0)
sample_data_double = (torch.rand(batch_size*(steps+1), 1)).to(dev)
sample_data_single = sample_data_double.type(torch.float32)

# set parameters
l_0, l_1, l_2, l_3 = 1, 3, 7, 1  # dimensions of the layers
a, b = 0, 1                      # the domain will be [a,b]
xi = 1                           # constant target function


# define loss function
def loss(N, x):
    y = N(x)
    return (y-xi).square().mean()


# function setting the starting model parameters as normally distributed
def init_weights(n):
    if isinstance(n, torch.nn.Linear):
        torch.nn.init.normal_(n.weight, 0, 1/np.sqrt(l_1))
        torch.nn.init.normal_(n.bias, 0, 1/np.sqrt(l_1))


# set up NN model
N_double = torch.nn.Sequential(
    torch.nn.Linear(l_0, l_1),
    torch.nn.ReLU(),
    torch.nn.Linear(l_1, l_2),
    torch.nn.ReLU(),
    torch.nn.Linear(l_2, l_3)
).to(dev)

# set up normally distributed starting model parameters
torch.manual_seed(0)
N_double.apply(init_weights)
N_single = copy.deepcopy(N_double)
N_single.type(torch.float32)

# define observation points
obs = np.empty(shape=129)
index = 0
```

```
for i in range(0, 23, 2):
    for k in range(2**i, min(2**(i+2), steps)):
        if k % max(1, 2**(i-2)) == False:
            obs[index] = k
            index = index + 1

# set up table
losses_table = np.empty((len(obs), 3), dtype=object)

# approximation in double precision
# optimizer function
optimizer_double = torch.optim.SGD(N_double.parameters(), lr=gamma)

# train the network
for step in range(steps+1):
    # generate uniformly distributed samples from [0,1]
    x = sample_data_double[step*batch_size:(step+1)*batch_size]
    # set the gradients back to zero
    optimizer_double.zero_grad()
    # compute the loss
    L = loss(N_double, x)
    # compute the gradients of L w.r.t. the model parameters
    L.backward()
    # update the model parameters
    optimizer_double.step()
    # save results in the table
    if step in obs:
        losses_table[np.where(obs == step)[0], [0, 2]] = step, L.data

# approximation in single precision
# optimizer function
optimizer_single = torch.optim.SGD(N_single.parameters(), lr=gamma)

# train the network
for step in range(steps+1):
    # generate uniformly distributed samples from [0,1]
    x = sample_data_single[step*batch_size:(step+1)*batch_size]
    # set the gradients back to zero
    optimizer_single.zero_grad()
    # compute the loss
    L = loss(N_single, x)
    # compute the gradients of L w.r.t. the model parameters
    L.backward()
    # update the model parameters
    optimizer_single.step()
    # save results in the table
    if step in obs:
        losses_table[np.where(obs == step)[0], [0, 1]] = step, L.data

# save table
np.savetxt('mean_square_error_example2.csv', losses_table, delimiter=',',
          fmt='%s, %.64f, %.64f')

# plot mean square error
plt.loglog(losses_table[:, 0], losses_table[:, 2], "r", label='Double precision')
plt.loglog(losses_table[:, 0], losses_table[:, 1], "b", label='Single precision')
plt.ylim([0.00000000000000000001, 10])
```

```python
plt.xlim([1, 12000000])
plt.xlabel('Number of SGD steps')
plt.ylabel('Mean square error (expected risk)')
plt.legend()
plt.savefig('mean_square_error_example2_loglog_graph', dpi=1200)
```

Source Code 4.3: PYTHON code used to create Table 4.2 and Figure 4.5 in Subsection 4.5.2 (filename: sgd2.py).

```python
import numpy as np
import torch
import matplotlib.pyplot as plt

# use GPU if available
dev = torch.device("cuda") if torch.cuda.is_available() else torch.device("cpu")

# set precision to double
torch.set_default_dtype(torch.float64)
torch.manual_seed(0)

# configure the training parameters
steps = 10000000                # n
batch_size = 100                # M_n, number of sample points
gamma = 1/2000                  # step size (learning rate)

# set parameters
l_0, l_1, l_2, l_3 = 1, 3, 7, 1  # dimensions of the layers
a, b = 0, 1                     # the domain will be [a,b]
xi = 1                          # constant target function


# define loss function
def loss(N, x):
    y = N(x)
    return (y-xi).square().mean()


# function setting the starting model parameters as normally distributed
def init_weights(n):
    if isinstance(n, torch.nn.Linear):
        torch.nn.init.normal_(n.weight, 0, 1/np.sqrt(l_1))
        torch.nn.init.normal_(n.bias, 0, 1/np.sqrt(l_1))


# set up NN model
N = torch.nn.Sequential(
    torch.nn.Linear(l_0, l_1),
    torch.nn.ReLU(),
    torch.nn.Linear(l_1, l_2),
    torch.nn.ReLU(),
    torch.nn.Linear(l_2, l_3)
).to(dev)

# set the starting model parameters as normally distributed
N.apply(init_weights)
```

```python

# optimizer function
optimizer = torch.optim.SGD(N.parameters(), lr=gamma)

# define observation points
obs = np.empty(shape=129)
index = 0
for i in range(0, 23, 2):
    for k in range(2**i, min(2**(i+2), steps)):
        if k % max(1, 2**(i-2)) == False:
            obs[index] = k
            index = index + 1

# train the network
loss_table = np.empty((len(obs), 2), dtype=object)
for step in range(steps+1):
    # generate uniformly distributed samples from [0,1]
    x = (torch.rand(batch_size, 1)).to(dev)
    # set the gradients back to zero
    optimizer.zero_grad()
    # compute the loss
    L = loss(N, x)
    # compute the gradients of L w.r.t. the model parameters
    L.backward()
    # update the model parameters
    optimizer.step()
    # save results in the table
    if step in obs:
        loss_table[np.where(obs == step)[0], :] = step, L.data

# plot mean square error
plt.loglog(loss_table[:, 0], loss_table[:, 1], color= "r")
plt.ylim([0.00000000000000000001, 10])
plt.xlim([1, 12000000])
plt.xlabel('Number of SGD steps')
plt.ylabel('Mean square error (expected risk)')
plt.show()
```

Source Code 4.4: Simplified variant of the PYTHONscript in Source Code 4.3 above (filename: `sgd2_short.py`).

Bibliography

[1] P.-A. Absil, R. Mahony, and B. Andrews. “Convergence of the iterates of descent methods for analytic cost functions”. In: *SIAM J. Optim.* 16.2 (2005), pp. 531–547. ISSN: 1052-6234. DOI: 10.1137/040605266. URL: https://doi.org/10.1137/040605266.

[2] Ömer Deniz Akyildiz and Sotirios Sabanis. “Nonasymptotic analysis of Stochastic Gradient Hamiltonian Monte Carlo under local conditions for nonconvex optimization”. In: *arXiv:2002.05465* (2021), 26 pages. arXiv: 2002.05465 [math.OC].

[3] Zeyuan Allen-Zhu, Yuanzhi Li, and Zhao Song. “A Convergence Theory for Deep Learning via Over-Parameterization”. In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, Sept. 2019, pp. 242–252. URL: https://proceedings.mlr.press/v97/allen-zhu19a.html.

[4] Sanjeev Arora et al. “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 Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. Long Beach, California, USA: PMLR, June 2019, pp. 322–332. URL: http://proceedings.mlr.press/v97/arora19a.html.

[5] Hedy Attouch and Jerome Bolte. “On the convergence of the proximal algorithm for nonsmooth functions involving analytic features”. In: *Mathematical Programming* 116.1 (May 2007), pp. 5–16. DOI: 10.1007/s10107-007-0133-5. URL: https://hal.archives-ouvertes.fr/hal-00803898.

[6] Francis Bach and Eric Moulines. “Non-strongly-convex smooth stochastic approximation with convergence rate O(1/n)”. In: *Advances in Neural Information Processing Systems*. Ed. by C. J. C. Burges et al. Vol. 26. Curran Associates, Inc., 2013. URL: https://proceedings.neurips.cc/paper/2013/file/7fe1f8abaad094e0b5cb1b01d712f708-Paper.pdf.

[7] Sören Bartels. *Numerical Approximation of Partial Differential Equations*. eng. 1st ed. Vol. 64. Texts in Applied Mathematics. Cham: Springer Nature, 2016. ISBN: 3319323547.

[8] Erhan Bayraktar et al. “Valuation of mortality risk via the instantaneous Sharpe ratio: applications to life annuities”. In: *J. Econom. Dynam. Control* 33.3 (2009), pp. 676–691. ISSN: 0165-1889,1879-1743. DOI: 10.1016/j.jedc.2008.09.004. URL: https://doi.org/10.1016/j.jedc.2008.09.004.

[9] Christian Beck, Weinan E, and Arnulf Jentzen. “Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations”. In: *J. Nonlinear Sci.* 29.4 (2019), pp. 1563–1619. ISSN: 0938-8974,1432-1467. DOI: 10.1007/s00332-018-9525-3. URL: https://doi.org/10.1007/s00332-018-9525-3.

[10] Christian Beck, Martin Hutzenthaler, and Arnulf Jentzen. "On nonlinear Feynman-Kac formulas for viscosity solutions of semilinear parabolic partial differential equations". In: *Stoch. Dyn.* 21.8 (2021), Paper No. 2150048, 68. ISSN: 0219-4937,1793-6799. DOI: 10.1142/S0219493721500489. URL: https://doi.org/10.1142/S0219493721500489.

[11] Christian Beck, Arnulf Jentzen, and Benno Kuckuck. "Full error analysis for the training of deep neural networks". In: *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* 25.2 (2022), Paper No. 2150020, 76. ISSN: 0219-0257. DOI: 10.1142/S021902572150020X. URL: https://doi.org/10.1142/S021902572150020X.

[12] Christian Beck et al. "Deep splitting method for parabolic PDEs". In: *SIAM J. Sci. Comput.* 43.5 (2021), A3135–A3154. ISSN: 1064-8275,1095-7197. DOI: 10.1137/19M1297919. URL: https://doi.org/10.1137/19M1297919.

[13] Christian Beck et al. "On existence and uniqueness properties for solutions of stochastic fixed point equations". In: *Discrete Contin. Dyn. Syst. Ser. B* 26.9 (2021), pp. 4927–4962. ISSN: 1531-3492. DOI: 10.3934/dcdsb.2020320. URL: https://doi.org/10.3934/dcdsb.2020320.

[14] Christian Beck et al. "Solving the Kolmogorov PDE by means of deep learning". In: *J. Sci. Comput.* 88.3 (2021), Paper No. 73, 28. ISSN: 0885-7474. DOI: 10.1007/s10915-021-01590-0. URL: https://doi.org/10.1007/s10915-021-01590-0.

[15] Sebastian Becker, Patrick Cheridito, and Arnulf Jentzen. "Deep optimal stopping". In: *J. Mach. Learn. Res.* 20 (2019), Paper No. 74, 25. ISSN: 1532-4435,1533-7928.

[16] Sebastian Becker et al. "Numerical simulations for full history recursive multilevel Picard approximations for systems of high-dimensional partial differential equations". In: *Commun. Comput. Phys.* 28.5 (2020), pp. 2109–2138. ISSN: 1815-2406,1991-7120. DOI: 10.4208/cicp.oa-2020-0130. URL: https://doi.org/10.4208/cicp.oa-2020-0130.

[17] Richard Bellman. *Dynamic programming.* Princeton Landmarks in Mathematics. Reprint of the 1957 edition, With a new introduction by Stuart Dreyfus. Princeton University Press, Princeton, NJ, 2010, pp. xxx+340. ISBN: 978-0-691-14668-3.

[18] Dimitri P. Bertsekas and John N. Tsitsiklis. "Gradient convergence in gradient methods with errors". In: *SIAM J. Optim.* 10.3 (2000), pp. 627–642. ISSN: 1052-6234. DOI: 10.1137/S1052623497331063. URL: https://doi.org/10.1137/S1052623497331063.

[19] Bruno Bouchard et al. "Numerical approximation of BSDEs using local polynomial drivers and branching processes". In: *Monte Carlo Methods Appl.* 23.4 (2017), pp. 241–263. ISSN: 0929-9629,1569-3961. DOI: 10.1515/mcma-2017-0116. URL: https://doi.org/10.1515/mcma-2017-0116.

[20] James B. Carrell. *Groups, Matrices, and Vector Spaces : A Group Theoretic Approach to Linear Algebra.* eng. New York, NY, 2017.

[21] Quentin Chan-Wai-Nam, Joseph Mikael, and Xavier Warin. "Machine learning for semi linear PDEs". In: *J. Sci. Comput.* 79.3 (2019), pp. 1667–1712. ISSN: 0885-7474,1573-7691. DOI: 10.1007/s10915-019-00908-3. URL: https://doi.org/10.1007/s10915-019-00908-3.

[22] Zhengdao Chen et al. "A Dynamical Central Limit Theorem for Shallow Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 22217–22230. URL: https://proceedings.neurips.cc/paper/2020/file/fc5b3186f1cf0daece964f78259b7ba0-Paper.pdf.

[23] Patrick Cheridito, Arnulf Jentzen, and Florian Rossmannek. "Non-convergence of stochastic gradient descent in the training of deep neural networks". In: *J. Complexity* 64 (2021), Paper No. 101540, 10. ISSN: 0885-064X. DOI: 10.1016/j.jco.2020.101540. URL: https://doi.org/10.1016/j.jco.2020.101540.

[24] Patrick Cheridito et al. "A proof of convergence for gradient descent in the training of artificial neural networks for constant target functions". In: *J. Complexity* 72 (2022), Paper No. 101646, 26. ISSN: 0885-064X. DOI: 10.1016/j.jco.2022.101646. URL: https://doi.org/10.1016/j.jco.2022.101646.

[25] Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. "User's guide to viscosity solutions of second order partial differential equations". In: *Bull. Amer. Math. Soc. (N.S.)* 27.1 (1992), pp. 1–67. ISSN: 0273-0979,1088-9485. DOI: 10.1090/S0273-0979-1992-00266-5. URL: https://doi.org/10.1090/S0273-0979-1992-00266-5.

[26] Michael G. Crandall and Pierre-Louis Lions. "Condition d'unicité pour les solutions généralisées des équations de Hamilton-Jacobi du premier ordre". In: *C. R. Acad. Sci. Paris Sér. I Math.* 292.3 (1981), pp. 183–186. ISSN: 0249-6291.

[27] Stéphane Crépey et al. "Counterparty risk and funding: the four wings of the TVA". In: *Int. J. Theor. Appl. Finance* 16.2 (2013), pp. 1350006, 31. ISSN: 0219-0249,1793-6322. DOI: 10.1142/S0219024913500064. URL: https://doi.org/10.1142/S0219024913500064.

[28] Philip J. Davis. *Methods of numerical integration.* 2. ed., 1. [Dr.] Computer science and applied mathematics. Orlando, Fla. [u.a: Acad. Pr., 1984. ISBN: 0122063600.

[29] Steffen Dereich and Sebastian Kassing. "Convergence of stochastic gradient descent schemes for Lojasiewicz-landscapes". In: *arXiv:2102.09385* (2021), 24 pages. arXiv: 2102.09385 [cs.LG].

[30] Steffen Dereich and Thomas Müller-Gronbach. "General multilevel adaptations for stochastic approximation algorithms of Robbins-Monro and Polyak-Ruppert type". In: *Numer. Math.* 142.2 (2019), pp. 279–328. ISSN: 0029-599X. DOI: 10.1007/s00211-019-01024-y. URL: https://doi.org/10.1007/s00211-019-01024-y.

[31] Simon Du et al. "Gradient Descent Finds Global Minima of Deep Neural Networks". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. Long Beach, California, USA: PMLR, June 2019, pp. 1675–1685. URL: http://proceedings.mlr.press/v97/du19c.html.

[32] Simon Du et al. "Gradient Descent Finds Global Minima of Deep Neural Networks". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, Sept. 2019, pp. 1675–1685. URL: https://proceedings.mlr.press/v97/du19c.html.

[33] Weinan E, Jiequn Han, and Arnulf Jentzen. “Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations”. In: *Commun. Math. Stat.* 5.4 (2017), pp. 349–380. ISSN: 2194-6701,2194-671X. DOI: 10.1007/s40304-017-0117-6. URL: https://doi.org/10.1007/s40304-017-0117-6.

[34] Weinan E, Chao Ma, and Lei Wu. “A comparative analysis of optimization and generalization properties of two-layer neural network and random feature models under gradient descent dynamics”. In: *Sci. China Math.* 63.7 (2020), pp. 1235–1258. ISSN: 1674-7283. DOI: 10.1007/s11425-019-1628-5. URL: https://doi.org/10.1007/s11425-019-1628-5.

[35] Weinan E and Bing Yu. “The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems”. In: *Commun. Math. Stat.* 6.1 (2018), pp. 1–12. ISSN: 2194-6701,2194-671X. DOI: 10.1007/s40304-018-0127-z. URL: https://doi.org/10.1007/s40304-018-0127-z.

[36] Weinan E et al. “Multilevel Picard iterations for solving smooth semilinear parabolic heat equations”. In: *Partial Differ. Equ. Appl.* 2.6 (2021), Paper No. 80, 31. ISSN: 2662-2963,2662-2971. DOI: 10.1007/s42985-021-00089-5. URL: https://doi.org/10.1007/s42985-021-00089-5.

[37] Weinan E et al. “On multilevel Picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations”. In: *J. Sci. Comput.* 79.3 (2019), pp. 1534–1571. ISSN: 0885-7474,1573-7691. DOI: 10.1007/s10915-018-00903-0. URL: https://doi.org/10.1007/s10915-018-00903-0.

[38] Simon Eberle et al. “Existence, uniqueness, and convergence rates for gradient flows in the training of artificial neural networks with ReLU activation”. In: *Electron. Res. Arch.* 31.5 (2023), pp. 2519–2554. DOI: 10.3934/era.2023128. URL: https://doi.org/10.3934/era.2023128.

[39] Ibrahim Ekren and Johannes Muhle-Karbe. “Portfolio choice with small temporary and transient price impact”. In: *Math. Finance* 29.4 (2019), pp. 1066–1115. ISSN: 0960-1627,1467-9965. DOI: 10.1111/mafi.12204. URL: https://doi.org/10.1111/mafi.12204.

[40] K. D. Elworthy and X.-M. Li. “Formulae for the derivatives of heat semigroups”. In: *J. Funct. Anal.* 125.1 (1994), pp. 252–286. ISSN: 0022-1236. DOI: 10.1006/jfan.1994.1124. URL: https://doi.org/10.1006/jfan.1994.1124.

[41] Benjamin Fehrman, Benjamin Gess, and Arnulf Jentzen. “Convergence rates for the stochastic gradient descent method for non-convex objective functions”. In: *J. Mach. Learn. Res.* 21 (2020), Paper No. 136, 48. ISSN: 1532-4435.

[42] R. P. Feynman. “Space-time approach to non-relativistic quantum mechanics”. In: *Rev. Modern Physics* 20 (1948), pp. 367–387. DOI: 10.1103/revmodphys.20.367. URL: https://doi.org/10.1103/revmodphys.20.367.

[43] Emmanuel Gobet, Jean-Philippe Lemor, and Xavier Warin. “A regression-based Monte Carlo method to solve backward stochastic differential equations”. In: *Ann. Appl. Probab.* 15.3 (2005), pp. 2172–2202. ISSN: 1050-5164,2168-8737. DOI: 10.1214/105051605000000412. URL: https://doi.org/10.1214/105051605000000412.

[44] Christian GroSSmann, Hans-Görg Roos, and Martin Stynes. *Numerical treatment of partial differential equations*. eng. Universitext. Berlin [u.a: Springer, 2007. ISBN: 9783540715825.

[45] Martin Hairer, Martin Hutzenthaler, and Arnulf Jentzen. "Loss of regularity for Kolmogorov equations". In: *Ann. Probab.* 43.2 (2015), pp. 468–527. ISSN: 0091-1798. DOI: 10.1214/13-AOP838. URL: https://doi.org/10.1214/13-AOP838.

[46] Jiequn Han, Arnulf Jentzen, and Weinan E. "Solving high-dimensional partial differential equations using deep learning". In: *Proc. Natl. Acad. Sci. USA* 115.34 (2018), pp. 8505–8510. ISSN: 0027-8424,1091-6490. DOI: 10.1073/pnas.1718942115. URL: https://doi.org/10.1073/pnas.1718942115.

[47] Jiequn Han and Jihao Long. "Convergence of the deep BSDE method for coupled FBSDEs". In: *Probab. Uncertain. Quant. Risk* 5 (2020), Paper No. 5, 33. ISSN: 2095-9672,2367-0126. DOI: 10.1186/s41546-020-00047-w. URL: https://doi.org/10.1186/s41546-020-00047-w.

[48] Pierre Henry-Labordère, Xiaolu Tan, and Nizar Touzi. "A numerical algorithm for a class of BSDEs via the branching process". In: *Stochastic Process. Appl.* 124.2 (2014), pp. 1112–1140. ISSN: 0304-4149,1879-209X. DOI: 10.1016/j.spa.2013.10.005. URL: https://doi.org/10.1016/j.spa.2013.10.005.

[49] Pierre Henry-Labordère et al. "Branching diffusion representation of semilinear PDEs and Monte Carlo approximation". In: *Ann. Inst. Henri Poincaré Probab. Stat.* 55.1 (2019), pp. 184–210. ISSN: 0246-0203,1778-7017. DOI: 10.1214/17-aihp880. URL: https://doi.org/10.1214/17-aihp880.

[50] Anselm Hudde, Martin Hutzenthaler, and Sara Mazzonetto. "A stochastic Gronwall inequality and applications to moments, strong completeness, strong local Lipschitz continuity, and perturbations". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 57.2 (2021). DOI: 10.1214/20-aihp1064. URL: https://doi.org/10.1214%2F20-aihp1064.

[51] Côme Huré, Huyên Pham, and Xavier Warin. "Deep backward schemes for high-dimensional nonlinear PDEs". In: *Math. Comp.* 89.324 (2020), pp. 1547–1579. ISSN: 0025-5718,1088-6842. DOI: 10.1090/mcom/3514. URL: https://doi.org/10.1090/mcom/3514.

[52] Martin Hutzenthaler, Arnulf Jentzen, and Thomas Kruse. "Overcoming the curse of dimensionality in the numerical approximation of parabolic partial differential equations with gradient-dependent nonlinearities". In: *Found. Comput. Math.* 22.4 (2022), pp. 905–966. ISSN: 1615-3375,1615-3383. DOI: 10.1007/s10208-021-09514-y. URL: https://doi.org/10.1007/s10208-021-09514-y.

[53] Martin Hutzenthaler, Arnulf Jentzen, and Philippe von Wurstemberger. "Overcoming the curse of dimensionality in the approximative pricing of financial derivatives with default risks". In: *Electron. J. Probab.* 25 (2020), Paper No. 101, 73. ISSN: 1083-6489. DOI: 10.1214/20-ejp423. URL: https://doi.org/10.1214/20-ejp423.

[54] Martin Hutzenthaler and Thomas Kruse. "Multilevel Picard approximations of high-dimensional semilinear parabolic differential equations with gradient-dependent nonlinearities". In: *SIAM J. Numer. Anal.* 58.2 (2020), pp. 929–961. ISSN: 0036-1429,1095-7170. DOI: 10.1137/17M1157015. URL: https://doi.org/10.1137/17M1157015.

[55] Martin Hutzenthaler and Thomas Kruse. “Multilevel Picard approximations of high-dimensional semilinear parabolic differential equations with gradient-dependent nonlinearities”. In: *SIAM J. Numer. Anal.* 58.2 (2020), pp. 929–961. ISSN: 0036-1429,1095-7170. DOI: 10.1137/17M1157015. URL: https://doi.org/10.1137/17M1157015.

[56] Martin Hutzenthaler et al. “A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations”. In: *Partial Differ. Equ. Appl.* 1.2 (2020), Paper No. 10, 34. ISSN: 2662-2963,2662-2971. DOI: 10.1007/s42985-019-0006-9. URL: https://doi.org/10.1007/s42985-019-0006-9.

[57] Martin Hutzenthaler et al. *Convergence proof for stochastic gradient descent in the training of deep neural networks with ReLU activation for constant target functions.* 2023. arXiv: 2112.07369 [cs.LG].

[58] Martin Hutzenthaler et al. “Overcoming the curse of dimensionality in the numerical approximation of semilinear parabolic partial differential equations”. In: *Proc. A.* 476.2244 (2020), pp. 20190630, 25. ISSN: 1364-5021,1471-2946. DOI: 10.1098/rspa.2019.0630. URL: https://doi.org/10.1098/rspa.2019.0630.

[59] Martin Hutzenthaler et al. “Overcoming the curse of dimensionality in the numerical approximation of semilinear parabolic partial differential equations”. In: *Proc. A.* 476.2244 (2020), pp. 20190630, 25. ISSN: 1364-5021,1471-2946. DOI: 10.1098/rspa.2019.0630. URL: https://doi.org/10.1098/rspa.2019.0630.

[60] Peter Imkeller, Gonçalo dos Reis, and William Salkeld. “Differentiability of SDEs with drifts of super-linear growth”. In: *Electronic Journal of Probability* 24.none (2019), pp. 1–43. DOI: 10.1214/18-EJP261. URL: https://doi.org/10.1214/18-EJP261.

[61] Kiyosi Itô. “Stochastic integral”. In: *Proc. Imp. Acad. Tokyo* 20 (1944), pp. 519–524. ISSN: 0369-9846. URL: http://projecteuclid.org/euclid.pja/1195572786.

[62] Arnulf Jentzen and Timo Kröger. “Convergence rates for gradient descent in the training of overparameterized artificial neural networks with biases”. In: *arXiv:2102.11840* (2021), 38 pages. arXiv: 2102.11840 [cs.LG].

[63] Arnulf Jentzen and Adrian Riekert. “A proof of convergence for stochastic gradient descent in the training of artificial neural networks with ReLU activation for constant target functions”. In: *Z. Angew. Math. Phys.* 73.5 (2022), Paper No. 188. ISSN: 0044-2275. DOI: 10.1007/s00033-022-01716-w. URL: https://doi.org/10.1007/s00033-022-01716-w.

[64] Arnulf Jentzen and Adrian Riekert. “A proof of convergence for the gradient descent optimization method with random initializations in the training of neural networks with ReLU activation for piecewise linear target functions”. In: *J. Mach. Learn. Res.* 23 (2022), Paper No. [260], 50. ISSN: 1532-4435. DOI: 10.1086/287492. URL: https://doi.org/10.1086/287492.

[65] Arnulf Jentzen and Philippe von Wurstemberger. “Lower error bounds for the stochastic gradient descent optimization algorithm: sharp convergence rates for slowly and fast decaying learning rates”. In: *J. Complexity* 57 (2020), pp. 101438, 16. ISSN: 0885-064X. DOI: 10.1016/j.jco.2019.101438. URL: https://doi.org/10.1016/j.jco.2019.101438.

[66] Arnulf Jentzen et al. "Strong error analysis for stochastic gradient descent optimization algorithms". In: *IMA J. Numer. Anal.* 41.1 (2021), pp. 455–492. ISSN: 0272-4979. DOI: 10.1093/imanum/drz055. URL: https://doi.org/10.1093/imanum/drz055.

[67] M. Kac. "On distributions of certain Wiener functionals". In: *Trans. Amer. Math. Soc.* 65 (1949), pp. 1–13. ISSN: 0002-9947,1088-6850. DOI: 10.2307/1990512. URL: https://doi.org/10.2307/1990512.

[68] Olav Kallenberg. *Foundations of Modern Probability*. eng. 3rd ed. 2021. Probability Theory and Stochastic Modelling 99. Cham: Springer International Publishing, 2021. ISBN: 9783030618711.

[69] Ioannis Karatzas. *Brownian motion and stochastic calculus*. eng. 2. ed., corr. softcover ed. Graduate texts in mathematics 113. New York, NY: Springer, 1998. ISBN: 0387976558.

[70] Belhal Karimi et al. "Non-asymptotic Analysis of Biased Stochastic Approximation Scheme". In: *Proceedings of the Thirty-Second Conference on Learning Theory*. Ed. by Alina Beygelzimer and Daniel Hsu. Vol. 99. Proceedings of Machine Learning Research. PMLR, 25–28 Jun 2019, pp. 1944–1974. URL: https://proceedings.mlr.press/v99/karimi19a.html.

[71] Hamed Karimi, Julie Nutini, and Mark Schmidt. "Linear Convergence of Gradient and Proximal-Gradient Methods Under the Polyak-ojasiewicz Condition". In: *arXiv:1608.04636* (2020), 25 pages. arXiv: 1608.04636 [cs.LG].

[72] Hiroshi Kunita and Shinzo Watanabe. "On square integrable martingales". In: *Nagoya Math. J.* 30 (1967), pp. 209–245. ISSN: 0027-7630,2152-6842. URL: http://projecteuclid.org/euclid.nmj/1118796812.

[73] Serge Lang. *Fundamentals of differential geometry*. eng. Corr. 2. printing. Graduate texts in mathematics 191. New York: Springer, 2001. ISBN: 038798593X.

[74] Yunwen Lei et al. "Stochastic gradient descent for nonconvex learning without bounded gradient assumptions". In: *IEEE Trans. Neural Netw. Learn. Syst.* 31.10 (2020), pp. 4394–4400. ISSN: 2162-237X.

[75] Zichao Long, Yiping Lu, and Bin Dong. "PDE-Net 2.0: learning PDEs from data with a numeric-symbolic hybrid deep network". In: *J. Comput. Phys.* 399 (2019), pp. 108925, 17. ISSN: 0021-9991,1090-2716. DOI: 10.1016/j.jcp.2019.108925. URL: https://doi.org/10.1016/j.jcp.2019.108925.

[76] Attila Lovas et al. "Taming neural networks with TUSLA: Non-convex learning via adaptive stochastic gradient Langevin algorithms". In: *arXiv:2006.14514* (2021), 37 pages. arXiv: 2006.14514 [cs.LG].

[77] Lu Lu et al. "Dying ReLU and initialization: theory and numerical examples". In: *Commun. Comput. Phys.* 28.5 (2020), pp. 1671–1706. ISSN: 1815-2406. DOI: 10.4208/cicp.oa-2020-0165. URL: https://doi.org/10.4208/cicp.oa-2020-0165.

[78] Kjetil O. Lye, Siddhartha Mishra, and Deep Ray. "Deep learning observables in computational fluid dynamics". In: *J. Comput. Phys.* 410 (2020), pp. 109339, 26. ISSN: 0021-9991,1090-2716. DOI: 10.1016/j.jcp.2020.109339. URL: https://doi.org/10.1016/j.jcp.2020.109339.

[79] Ludovic Moreau, Johannes Muhle-Karbe, and H. Mete Soner. "Trading with small price impact". In: *Math. Finance* 27.2 (2017), pp. 350–400. ISSN: 0960-1627,1467-9965. DOI: 10.1111/mafi.12098. URL: https://doi.org/10.1111/mafi.12098.

[80] Eric Moulines and Francis Bach. “Non-Asymptotic Analysis of Stochastic Approximation Algorithms for Machine Learning”. In: *Advances in Neural Information Processing Systems*. Ed. by J. Shawe-Taylor et al. Vol. 24. Curran Associates, Inc., 2011. URL: https://proceedings.neurips.cc/paper/2011/file/40008b9a5380fcacce3976bf7c08af5b-Paper.pdf.

[81] Yu. E. Nesterov. “A method for solving the convex programming problem with convergence rate $O(1/k^2)$”. In: *Dokl. Akad. Nauk SSSR* 269.3 (1983), pp. 543–547. ISSN: 0002-3264.

[82] Yurii Nesterov. *Introductory lectures on convex optimization*. Vol. 87. Applied Optimization. A basic course. Kluwer Academic Publishers, Boston, MA, 2004, pp. xviii+236. ISBN: 1-4020-7553-7. DOI: 10.1007/978-1-4419-8853-9. URL: https://doi.org/10.1007/978-1-4419-8853-9.

[83] Erich Novak and Henryk Woniakowski. *Tractability of multivariate problems. Vol. 1: Linear information*. Vol. 6. EMS Tracts in Mathematics. European Mathematical Society (EMS), Zürich, 2008, pp. xii+384. ISBN: 978-3-03719-026-5. DOI: 10.4171/026. URL: https://doi.org/10.4171/026.

[84] Erich Novak and Henryk Woniakowski. *Tractability of multivariate problems. Volume II: Standard information for functionals*. Vol. 12. EMS Tracts in Mathematics. European Mathematical Society (EMS), Zürich, 2010, pp. xviii+657. ISBN: 978-3-03719-084-5. DOI: 10.4171/084. URL: https://doi.org/10.4171/084.

[85] Erich Novak and Henryk Woniakowski. *Tractability of multivariate problems. Volume III: Standard information for operators*. Vol. 18. EMS Tracts in Mathematics. European Mathematical Society (EMS), Zürich, 2012, pp. xviii+586. ISBN: 978-3-03719-116-3. DOI: 10.4171/116. URL: https://doi.org/10.4171/116.

[86] David Nualart. *The Malliavin Calculus and Related Topics*. Probability and its Applications (New York). New York: Springer-Verlag, 1995, pp. xi+266. ISBN: 0-387-94432-X.

[87] Vivak Patel. “Stopping Criteria for, and Strong Convergence of, Stochastic Gradient Descent on Bottou-Curtis-Nocedal Functions”. In: *arXiv:2004.00475* (2021), 44 pages. arXiv: 2004.00475 [math.OC].

[88] Maziar Raissi. “Deep hidden physics models: deep learning of nonlinear partial differential equations”. In: *J. Mach. Learn. Res.* 19 (2018), Paper No. 25, 24. ISSN: 1532-4435,1533-7928.

[89] Alexander Rakhlin, Ohad Shamir, and Karthik Sridharan. “Making Gradient Descent Optimal for Strongly Convex Stochastic Optimization”. In: *Proceedings of the 29th International Coference on International Conference on Machine Learning*. ICML’12. Edinburgh, Scotland: Omnipress, 2012, pp. 1571–1578. ISBN: 9781450312851.

[90] Herbert Robbins. “A Remark on Stirling’s Formula”. In: *The American Mathematical Monthly* 62.1 (1955), pp. 26–29. ISSN: 00029890, 19300972. URL: http://www.jstor.org/stable/2308012 (visited on 09/08/2023).

[91] Walter Rudin. *Principles of Mathematical Analysis*. Intern. ed., 3. ed. McGraw-Hill international editions : Mathematics series. McGraw-Hill, 1976, pp. x+342. ISBN: 9780070856134.

[92] Walter Rudin. *Real and complex analysis*. eng. 2. ed. McGraw-Hill series in higher mathematics. New York: McGraw-Hill, 1974. ISBN: 0070542333.

[93] Karthik A. Sankararaman et al. "The Impact of Neural Network Overparameterization on Gradient Confusion and Stochastic Gradient Descent". In: *arXiv:1904.06963* (2020), 28 pages. arXiv: 1904.06963 [cs.LG].

[94] Mark Schmidt and Nicolas Le Roux. "Fast Convergence of Stochastic Gradient Descent under a Strong Growth Condition". In: *arXiv:1308.6370* (2013), 5 pages. arXiv: 1308.6370 [math.OC].

[95] Ohad Shamir. "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 Alina Beygelzimer and Daniel Hsu. Vol. 99. Proceedings of Machine Learning Research. PMLR, 25–28 Jun 2019, pp. 2691–2713. URL: https://proceedings.mlr.press/v99/shamir19a.html.

[96] Ralph E. Showalter. *Hilbert space methods for partial differential equations*. Paperback ed. San Francisco: PITMAN ADVANCED PUBL. PROGRAM, 1979.

[97] Justin Sirignano and Konstantinos Spiliopoulos. "DGM: a deep learning algorithm for solving partial differential equations". In: *J. Comput. Phys.* 375 (2018), pp. 1339–1364. ISSN: 0021-9991,1090-2716. DOI: 10.1016/j.jcp.2018.08.029. URL: https://doi.org/10.1016/j.jcp.2018.08.029.

[98] H. Windcliff et al. "Hedging with a correlated asset: solution of a nonlinear pricing PDE". In: *J. Comput. Appl. Math.* 200.1 (2007), pp. 86–115. ISSN: 0377-0427,1879-1778. DOI: 10.1016/j.cam.2005.12.008. URL: https://doi.org/10.1016/j.cam.2005.12.008.

[99] Stephan Wojtowytsch. "Stochastic gradient descent with noise of machine learning type. Part I: Discrete time analysis". In: *arXiv:2105.01650* (2021), 34 pages. arXiv: 2105.01650 [stat.ML].

[100] Yangyang Xu and Wotao Yin. "A Block Coordinate Descent Method for Regularized Multiconvex Optimization with Applications to Nonnegative Tensor Factorization and Completion". In: *SIAM Journal on Imaging Sciences* 6.3 (2013), pp. 1758–1789. DOI: 10.1137/120887795. eprint: https://doi.org/10.1137/120887795. URL: https://doi.org/10.1137/120887795.

[101] Guodong Zhang, James Martens, and Roger B Grosse. "Fast Convergence of Natural Gradient Descent for Over-Parameterized Neural Networks". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8082–8093. URL: http://papers.nips.cc/paper/9020-fast-convergence-of-natural-gradient-descent-for-over-parameterized-neural-networks.pdf.

[102] Difan Zou et al. "Gradient descent optimizes over-parameterized deep ReLU networks". In: *Mach. Learn.* 109.3 (2020), pp. 467–492. ISSN: 0885-6125. DOI: 10.1007/s10994-019-05839-6. URL: https://doi.org/10.1007/s10994-019-05839-6.